Sincerest thanks to my wonderful wife, Linda, and our children, David and Cathy, for their patience and support. Also, the editorial help of our three Cavalier King Charles puppies, Winston, Humphrey, and Punkin, was invaluable.

Gary S. Firestein

Sincere thanks for the kind mentoring from Edward D. Harris, Jr., as well as for the support of my wife, Lenore, and my children, Graham and Laura.

Ralph C. Budd

To my three boys: my dear husband, Frank Cockerill, and our two wonderful sons, Richard and Matthew, for being my constant source of inspiration, love, and pride. And to my parents, Huda and Ezzat, for their love and tireless support.

Sherine E. Gabriel

To my wife, Karin, for her patience, understanding, and love and to our wonderful girls, Megan and Rebecca, who continue to enlighten me.

Iain B. McInnes

Sincere thanks to my wife, Deb, for her patience and love and to our wonderful children, Kim and Andy, Jennie and Dan, and Scott and Melissa. I also want to thank the members of my division who continue to support me in all my efforts.

James R. O’Dell
Edward D. “Ted” Harris, Jr., was one of the four founding editors of the *Textbook of Rheumatology*. In the late 1970s, Bill Kelley sensed the need for a text that reflected the growth of rheumatology into a mature discipline. He met with Ted, who quickly agreed, and they identified Shaun Ruddy and Clem Sledge as co-editors. A prime concern was that the new book should be grounded in the abundant information in basic science that supported our subspecialty. The standards they set were responsible for the high quality of the finished Textbook. Ted’s choice of the iconic profile of Renoir, who suffered from rheumatoid arthritis, has graced the cover of nine editions of the book and served to connect us to the humanitarian aspect of our discipline.

Ted was a graduate of Dartmouth College and its medical school and of Harvard Medical School. Following his residency at Massachusetts General Hospital he moved to the National Institutes of Health (NIH), where he engaged in research on collagen. In his spare moments he also formed a jazz ensemble, with himself playing bass. Upon Ted’s return to Mass General he entered a rheumatology fellowship and joined the laboratory of Dr. Stephen Krane, where Ted applied his knowledge of collagen to the inflammatory synovium of rheumatoid arthritis.

In 1970 Ted was recruited back to Dartmouth, where he built a robust connective tissue disease unit and received one of the NIH’s first arthritis center awards. Along with long-time colleague Dr. Constance Brinckerhoff, Ted’s group defined the role of collagenase and metalloproteinases in the rheumatoid synovium. In 1979 Ted was sole author of the seminal monograph *Rheumatoid Arthritis*, which detailed the complex interactions of the immune system with connective tissue in rheumatoid arthritis. In 1983 Rutgers Medical School recruited Ted to become Chair of Medicine, and four years later he assumed the Chair of Medicine position at Stanford, a position he held until 1995. During Ted’s career he authored well over 100 peer-reviewed publications and 70 reviews, chapters, editorials, and books.

Ted served as President of the American College of Rheumatology (ACR) and, during his tenure, skillfully helped arrange an amicable separation of the ACR and the Arthritis Foundation so that each organization could better pursue its mission. He was named a fellow of the British Royal College of Physicians in 2002 and received the Presidential Gold Medal from ACR in 2007.

Ted had a remarkably perceptive intellect and a razor wit. A former English major, his writing was crisp and vigorous. His love of language elevated and animated text. Colleagues knew that an “EDH note” could be mellifluous, mirthful, and merciless all at once. As academic secretary to Stanford, Ted’s amusing touches to the minutes of the Stanford Senate were legendary. He might squeeze in a quote from Dr. Seuss’s *Horton Hatches the Egg*, add footnotes on faculty members’ attire, or slip in sly editorial comments such as “wisely interrupted” or “introduced with appreciated brevity.” As a result, Ted’s words resonated and got results.

The English degree came in handy when, in 1997, Ted was named executive secretary of Alpha Omega Alpha (AOA) and editor of *The Pharos*, the society’s nontechnical compendium of essays, poetry, art, and articles on medical history, ethics, and health policy. Ted breathed new life and style into the journal during his 13-year tenure as editor. Ted also created a 532-page anthology called *Creative Healers: A Collection of Essays, Reviews, and Poems from The Pharos, 1938-1998*, published by AOA in 2004. Reviewers on Amazon.com have mentioned the editor’s keen eye for engaging writing, calling the volume’s contents “moving” and “a tribute to the range of interests percolating around in active intellects.”

Ted Harris mentored a generation of rheumatologists and taught us all by his example of dynamic creative thought and a deep humanitarian spirit. All of us involved with Kelley’s *Textbook of Rheumatology* feel a profound sadness with the loss of Ted, but even here Ted would provide the appropriate perspective, with a passage he wrote in a *Pharos* editorial: “Melancholy, that gray veil that takes color out of life, can, at the same time, add to the brilliance and value of life, if we feel what it is asking of us. Melancholy and sadness, similar to love, can make those compartment walls in our minds permeable, enabling us to express empathy that is truly felt within.”

Ted Harris was a consummate scholar and a great humanitarian, with a facile mind that spanned a wide array of interests from science to the arts. He was in essence a civilized man, something that has always been distinguished by its rarity.
CONTRIBUTORS

Steven B. Abramson, MD
Professor of Medicine and Pathology
Department of Medicine, Division of Rheumatology
NYU School of Medicine
New York, New York
Neutrophils; Eosinophils; Pathogenesis of Osteoarthritis

Kai-Nan An, PhD
Professor of Biomedical Engineering
Mayo Clinic College of Medicine;
Program Co-Director, Biomechanics and Motion Analysis Lab
Department of Orthopedic Surgery
Mayo Clinic
Rochester, Minnesota
Biomechanics

Felipe Andrade, MD, PhD
Assistant Professor of Medicine
Department of Medicine, Division of Rheumatology
Johns Hopkins University School of Medicine;
Center for Innovative Medicine
Johns Hopkins Medicine
Baltimore, Maryland
Autoantibodies in Rheumatoid Arthritis

John P. Atkinson, MD
Samuel B. Grant Professor of Medicine and Professor of Molecular Microbiology and Immunology
Washington University in St. Louis School of Medicine;
Physician, Barnes-Jewish Hospital
St. Louis, Missouri
Complement System

Dominique Baeten, MD, PhD
Associate Professor of Rheumatology
Department of Clinical Immunology and Rheumatology
University of Amsterdam Faculty of Medicine
Academic Medical Center
Amsterdam, The Netherlands
Ankylosing Spondylitis

Robert P. Baughman, MD
Professor of Medicine
Department of Internal Medicine
University of Cincinnati College of Medicine
Cincinnati, Ohio
Sarcoidosis

Dorcas E. Beaton, BScOT, PhD
Associate Professor
Graduate Department of Rehabilitation Science and Department of Occupational Science and Occupational Therapy
Faculty of Medicine;
Clinician-Investigator
Institute of Health Policy, Management and Evaluation;
Scientist, Health Measurement Institute for Work and Health
University of Toronto;
Director, Mobility Program Clinical Research Unit, Li Ka Shing Knowledge Institute
St. Michael’s Hospital
Toronto, Ontario, Canada
Assessment of Health Outcomes

Robert Bennett, MD
Professor of Medicine
Oregon Health & Science University School of Medicine
Portland, Oregon
Overlap Syndromes

Susanne M. Benseler, MD
Associate Professor of Pediatrics
Department of Pediatrics, Division of Rheumatology
Faculty of Medicine;
Clinician-Investigator
Institute of Health Policy, Evaluation and Management University of Toronto;
Associate Scientist
Research Institute
The Hospital for Sick Children
Toronto, Ontario, Canada
Pediatric Systemic Lupus Erythematosus, Dermatomyositis, Scleroderma, and Vasculitis

George Bertsias, MD, PhD
Fellow, Internal Medicine
Research Associate in Rheumatology, Clinical Immunology, and Allergy
University of Crete Faculty of Medicine
Heraklion, Crete, Greece
Treatment of Systemic Lupus Erythematosus

Nina Bhardwaj, MD, PhD
Professor of Medicine, Dermatology, and Pathology
Department of Medicine
NYU School of Medicine
New York, New York
Dendritic Cells
Contributors

Johannes W.J. Bijlsma, MD, PhD
Professor and Chair, Department of Rheumatology and Clinical Immunology
University of Utrecht Faculty of Medicine
Utrecht, The Netherlands
Glucocorticoid Therapy

Linda K. Bockenstedt, MD
Harold W. Jockers Professor of Medicine
Department of Internal Medicine, Section of Rheumatology
Yale University School of Medicine
New Haven, Connecticut
Lyme Disease

Maarten Boers, MD, PhD, MSc
Professor of Clinical Epidemiology
Department of Clinical Epidemiology and Biostatistics
VU University Amsterdam Faculty of Medicine
Amsterdam, The Netherlands
Assessment of Health Outcomes

Francesco Boin, MD
Assistant Professor of Medicine
Department of Medicine, Division of Rheumatology
Johns Hopkins University School of Medicine
Baltimore, Maryland
Clinical Features and Treatment of Scleroderma

Dimitrios T. Boumpas, MD, FACP
Professor of Internal Medicine
Professor of Rheumatology, Clinical Immunology, and Allergy
University of Crete Faculty of Medicine
Heraklion, Crete, Greece
Treatment of Systemic Lupus Erythematosus

†Barry Bresnihan, MD
Professor of Rheumatology
University College Dublin School of Medicine and Medical Science
National University of Ireland;
Consultant Rheumatologist
St. Vincent’s University Hospital;
Principal Investigator
Conway Institute of of Biomolecular and Biomedical Research
Dublin, Ireland
Synovium

Doreen B. Brettler, MD
Professor of Medicine
University of Massachusetts Medical School;
Director, New England Hemophilia Center
UMass Memorial Medical Center
Worcester, Massachusetts
Hemophilic Arthropathy

Christopher D. Buckley, DPhil, FRCP
Arthritis Research UK Professor of Rheumatology
College of Medical and Dental Sciences, School of Immunity and Infection;
Head, Rheumatology Research Group
Institute for Biomedical Research
University of Birmingham;
Birmingham, United Kingdom
Fibroblasts and Fibroblast-like Synoviocytes

Ralph C. Budd, MD
Professor of Medicine
Director, Vermont Center for Immunology and Infectious Diseases
The University of Vermont College of Medicine
Burlington, Vermont
T Lymphocytes

Christopher M. Burns, MD
Assistant Professor of Medicine
Department of Medicine, Section of Rheumatology
Geisel School of Medicine at Dartmouth;
Staff Rheumatologist
Dartmouth-Hitchcock Medical Center
Lebanon, New Hampshire
Clinical Features and Treatment of Gout

Amy C. Cannella, MD
Assistant Professor of Internal Medicine
Department of Internal Medicine, Division of Rheumatology
University of Nebraska College of Medicine
Omaha, Nebraska
Traditional DMARDs: Methotrexate, Leflunomide, Sulfasalazine, Hydroxychloroquine, and Combination Therapies

Eliza F. Chakravarty, MD, MS
Associate Member
Arthritis & Clinical Immunology Program
Oklahoma Medical Research Foundation
Oklahoma City, Oklahoma
Pregnancy in the Rheumatic Diseases; Musculoskeletal Syndromes in Malignancy

Christopher Chang, MD, PhD
Professor of Pediatrics
Chief, Division of Allergy, Asthma, and Immunology
Department of Pediatrics
Thomas Jefferson University
Philadelphia, Pennsylvania;
Associate Clinical Professor of Medicine
Department of Internal Medicine, Division of Rheumatology, Allergy, Clinical Immunology
UC Davis School of Medicine
Davis, California
Osteonecrosis

†Deceased.
Contributors

Joseph S. Cheng, MD, MS
Associate Professor of Neurological Surgery, Orthopedic Surgery, and Rehabilitation
Vanderbilt University School of Medicine;
Director, Neurosurgery Spine Program
Vanderbilt University Medical Center
Nashville, Tennessee

Neck Pain

Christopher P. Chiodo, MD
Chief, Foot and Ankle Service
Department of Orthopedic Surgery
Brigham and Women’s Hospital;
Instructor in Orthopaedic Surgery
Harvard Medical School
Boston, Massachusetts

Foot and Ankle Pain

Leslie G. Cleland, MBBS, MD
Clinical Professor
Department of Medicine
University of Adelaide School of Medicine, Faculty of Health Sciences
Head, Rheumatology Unit
Royal Adelaide Hospital
Adelaide, South Australia, Australia

Nutrition and Rheumatic Diseases

Megan E. Clowse, MD, MPH
Assistant Professor
Department of Medicine, Division of Rheumatology and Immunology
Duke University School of Medicine
Durham, North Carolina

Pregnancy in the Rheumatic Diseases

Paul P. Cook, MD
Professor of Medicine
Department of Medicine, Division of Infectious Diseases
Brody School of Medicine at East Carolina University
Greenville, North Carolina

Bacterial Arthritis

Joseph E. Craft, MD
Paul B. Beeson Professor of Medicine and Professor of Immunobiology
Director, Investigative Medicine Program
Yale University School of Medicine;
Chief of Rheumatology
Yale–New Haven Hospital
New Haven, Connecticut

Antinuclear Antibodies

Leslie J. Crofford, MD
Gloria W. Singletary Professor
Chief, Division of Rheumatology
Department of Internal Medicine
University of Kentucky School of Medicine
Director, Center for the Advancement of Women’s Health
UK HealthCare
Lexington, Kentucky

Prostanoid Biology and Its Therapeutic Targeting

Bruce N. Cronstein, MD
Paul R. Esserman Professor of Medicine
NYU School of Medicine
New York, New York

Acute Phase Reactants and the Concept of Inflammation

Mary K. Crow, MD
Joseph P. Routh Professor of Rheumatic Diseases in Medicine
Weill Cornell Medical College;
Benjamin M. Rosen Chair in Immunology and Inflammation Research
Divisions of Rheumatology and Research Hospital for Special Surgery
New York, New York

Etiology and Pathogenesis of Systemic Lupus Erythematosus

Gaye Cunnane, MB, PhD, FRCPI
Clinical Professor
Trinity College Dublin;
Consultant Rheumatologist
St. James’s Hospital
Dublin, Ireland

Relapsing Polychondritis; Hemochromatosis

John J. Cush, MD
Professor of Medicine and Rheumatology
Baylor University Medical Center–Dallas;
Director, Clinical Rheumatology
Baylor Research Institute–Rheumatology
Dallas, Texas

Polyarticular Arthritis

Maurizio Cutolo, MD
Professor of Rheumatology
University of Genova;
Director, Research Laboratories and Academic Unit of Clinical Rheumatology
Medical School University of Genova
Genova, Italy

Endocrine Diseases and the Musculoskeletal System

Maria Dall’Era, MD
Associate Professor of Medicine
Division of Rheumatology
University of California, San Francisco
San Francisco, California

Clinical Features of Systemic Lupus Erythematosus

Kathryn H. Dao, MD, FACP, FACR
Associate Director, Clinical Rheumatology
Department of Rheumatology
Baylor Research Institute
Dallas, Texas

Polyarticular Arthritis
Erika Darrah, PhD
Postdoctoral Fellow
Division of Rheumatology
Department of Medicine
Johns Hopkins University
Baltimore, Maryland
  Autoantibodies in Rheumatoid Arthritis

John M. Davis III, MD
Assistant Professor of Medicine
Department of Medicine, Division of Rheumatology
Mayo Clinic College of Medicine;
Consultant in Rheumatology
Mayo Clinic
Rochester, Minnesota
  History and Physical Examination of the Musculoskeletal System

Jeroen DeGroot, PhD
Research Manager
Pharmacokinetics & Human Studies
TNO Quality of Life
Business Unit Biomedical Research
Zeist, The Netherlands
  Biologic Markers

Clint Devin, MD
Assistant Professor of Neurological Surgery, Orthopaedic Surgery, and Rehabilitation
Vanderbilt University School of Medicine
Nashville, Tennessee
  Neck Pain

Betty Diamond, MD
Investigator
Center for Autoimmune and Musculoskeletal Diseases
The Feinstein Institute for Medical Research
Manhasset, New York
  B Cells

Federico Díaz-González, MD
Professor of Medicine
Department of Internal Medicine
University of La Laguna;
Staff, Rheumatology Service
University Hospital of the Canary Islands
San Cristobal de La Laguna, Tenerife, Spain
  Platelets

Paul E. Di Cesare, MD, FACS
Professor and Michael W. Chapman Chair, Department of Orthopaedic Surgery
UC Davis School of Medicine
Davis, California
  Pathogenesis of Osteoarthritis

Rajiv Dixit, MD
Clinical Professor of Medicine
Department of Medicine
UCSF School of Medicine
San Francisco;
Director, Northern California Arthritis Center
Walnut Creek, California
  Low Back Pain

Joost P.H. Drenth, MD, PhD
Professor of Molecular Gastroenterology and Hepatology
Department of Gastroenterology and Hepatology
Radboud University Nijmegen Faculty of Medical Sciences
Nijmegen, The Netherlands
  Familial Autoinflammatory Syndromes

Michael L. Dustin, PhD
Muriel G. and George W. Singer Professor of Molecular Immunology
Program in Molecular Pathogenesis
Helen L. and Martin S. Kimmel Center for Biology and Medicine, Skirball Institute of Biomolecular Medicine
New York, New York
  Adaptive Immunity and Organization of Lymphoid Tissues

Hani S. El-Gabalawy, MD, FRCPC
Endowed Research Chair in Rheumatology
Professor of Medicine and Immunology
University of Manitoba Faculty of Medicine;
Rheumatologist, Winnipeg Health Sciences Centre
Winnipeg, Manitoba, Canada
  Synovial Fluid Analyses, Synovial Biopsy, and Synovial Pathology

Keith B. Elkon, MD
Professor of Medicine and Immunology
University of Washington School of Medicine
Seattle, Washington
  Cell Survival and Death in Rheumatic Diseases

Doruk Erkan, MD
Associate Professor of Medicine
Weill Cornell Medical College;
Associate Physician-Scientist and Associate Attending Physician
Barbara Volcker Center for Women and Rheumatic Diseases
Hospital for Special Surgery
New York, New York
  Antiphospholipid Syndrome

Antonios Fanouriakis, MD
Professor of Rheumatology, Clinical Immunology, and Allergy
University of Crete Faculty of Medicine
Heraklion, Greece
  Treatment of Systemic Lupus Erythematosus
Max Field, MD, FRCP
Associate Academic
Division of Immunology, Institute of Infection, Immunology and Immunity
College of Medical, Veterinary and Life Sciences
University of Glasgow
Glasgow, United Kingdom
Acute Monoarthritis

Andrew Filer, PhD, MRCP
Senior Lecturer
College of Medical and Dental Sciences, School of Immunity and Infection;
Rheumatology Research Group
Institute for Biomedical Research
University of Birmingham;
Honorary Consultant Rheumatologist
University Hospitals NHS Foundation Trust Birmingham
Birmingham, United Kingdom
Fibroblasts and Fibroblast-like Synoviocytes

Gary S. Firestein, MD
Professor of Medicine
Dean and Associate Vice Chancellor of Translational Medicine
UC San Diego Health Sciences
La Jolla, California
Synovium; Etiology and Pathogenesis of Rheumatoid Arthritis; Clinical Features of Rheumatoid Arthritis

Oliver Fitzgerald, MD, FRCPI, FRCP(UK)
Newman Clinical Research Professor
University College Dublin School of Medicine and Medical Science
National University of Ireland;
Fellow, Conway Institute of Biomolecular and Biomedical Research;
Consultant Rheumatologist
St. Vincent's University Hospital
Dublin, Ireland
Psoriatic Arthritis

John P. Flaherty, MD
Professor in Medicine–Infectious Diseases
Associate Chief and Director of Clinical Services
Division of Infectious Diseases
Department of Medicine
Northwestern University Feinberg School of Medicine;
Chicago, Illinois
Mycobacterial Infections of Bones and Joints; Fungal Infections of Bones and Joints

Adrienne M. Flanagan, MD, PhD
Professor
Institute of Orthopaedics and Musculoskeletal Science
University College London
London;
Royal National Orthopaedic Hospital
Stanmore;
Department of Histopathology
University College Hospital
London, United Kingdom
Synovium

Karen A. Fortner, PhD
Assistant Professor
Immunobiology Program
Department of Medicine
The University of Vermont College of Medicine
Burlington, Vermont
T Lymphocytes

Sherine E. Gabriel, MD, MSc
William J. and Charles H. Mayo Professor
Professor of Medicine and Epidemiology
Mayo Clinic College of Medicine
Rochester, Minnesota
Cardiovascular Risk in Rheumatic Disease

J.S. Hill Gaston, MA, BM, PhD, FRCP, FMedSci
Professor of Rheumatology
Department of Medicine
University of Cambridge;
Addenbrooke's Hospital
Cambridge, United Kingdom
Reactive Arthritis and Undifferentiated Spondyloarthritis

Steffen Gay, MD
Center for Experimental Rheumatology
Zurich Center for Integrative Human Physiology, Life Science Zurich Graduate School/University Hospital
Zurich
Zurich, Switzerland
Epigenetics

M. Eric Gershwin, MD
Distinguished Professor of Medicine
Chief, Division of Rheumatology, Allergy and Clinical Immunology
Department of Medicine
UC Davis School of Medicine
Davis, California
Osteonecrosis

Allan Gibofsky, MD, JD, FACP, FCLM
Professor of Medicine and Public Health
Weill Cornell Medical College;
Attending Rheumatologist
Hospital for Special Surgery
New York, New York
Poststreptococcal Arthritis and Rheumatic Fever

Mark H. Ginsberg, MD
Professor of Medicine
Department of Medicine, Section of Rheumatology
UC San Diego School of Medicine
La Jolla, California
Platelets
Contributors

Mary B. Goldring, PhD
Professor of Cell and Developmental Biology
Weill Cornell Medical College;
Senior Scientist, Research Division
Hospital for Special Surgery
New York, New York

*Biology of the Normal Joint; Cartilage and Chondrocytes*

Steven R. Goldring, MD
Professor of Medicine
Weill Cornell Medical College;
Chief Scientific Officer
Hospital for Special Surgery
New York, New York

*Biology of the Normal Joint*

Yvonne M. Golightly, PT, PhD
Postdoctoral Research Associate
UNC Thurston Arthritis Research Center;
Department of Epidemiology
UNC Gillings School of Global Public Health
Chapel Hill, North Carolina

*Principles of Epidemiology in Rheumatic Disease*

Stuart Goodman, MD, PhD
Professor of Orthopaedics
Stanford University School of Medicine
Stanford, California

*Hip and Knee Pain*

Siamon Gordon, MB, ChB, PhD
Professor Emeritus
Sir William Dunn School of Pathology
University of Oxford
Oxford, United Kingdom

*Mononuclear Phagocytes in Rheumatic Diseases*

Walter Grassi, MD
Professor of Rheumatology
Clinica Reumatologica
Università Politecnica delle Marche
Jesi, Ancona, Italy

*Imaging Modalities in Rheumatic Diseases*

Adam Greenspan, MD, FACR
Professor Emeritus of Radiology
Department of Radiology, Section of Musculoskeletal Imaging
UC Davis School of Medicine
Davis, California

*Osteonecrosis*

Peter K. Gregersen, MD
Director, Robert S. Boas Center for Genomics and Human Genetics
The Feinstein Institute for Medical Research;
Professor of Molecular Medicine
Hofstra University School of Medicine
Manhasset, New York

*Genetics of Rheumatic Diseases*

Christine Grimaldi, PhD
Senior Principal Scientist
Nonclinical Drug Safety, US
Boehringer Ingelheim Pharmaceuticals, Inc.
Ridgefield, Connecticut

*B Cells*

Rula A. Hajj-Ali, MD
Assistant Professor of Medicine
Cleveland Clinic Lerner College of Medicine of Case Western University;
Staff Physician, Center for Vasculitis Care and Research
Cleveland Clinic
Cleveland, Ohio

*Primary Angiitis of the Central Nervous System*

Lorraine Harper, PhD, MRCP
Professor of Nephrology
College of Medical and Dental Sciences, School of Immunity and Infection
University of Birmingham
Birmingham, United Kingdom

*Antineutrophil Cytoplasm Antibody–Associated Vasculitis*

†Edward D. Harris, Jr., MD, MACR
George DeForest Barnett Professor of Medicine, Emeritus
Stanford University School of Medicine;
Academic Secretary to Stanford University, Emeritus
Stanford University
Stanford, California

*Clinical Features of Rheumatoid Arthritis*

Dominik R. Haudenschild, PhD
Assistant Professor in Residence
Department of Orthopaedic Surgery
UC Davis School of Medicine
Davis, California

*Pathogenesis of Osteoarthritis*

David B. Hellmann, MD
Aliki Perrotti Professor of Medicine
Department of Medicine
Johns Hopkins University School of Medicine;
Vice Dean and Chairman, Department of Medicine
Johns Hopkins Bayview
Baltimore, Maryland

*Giant Cell Arteritis, Polymyalgia Rheumatica, and Takayasu’s Arteritis*

Rikard Holmdahl, MD, PhD
Professor of Medical Biochemistry and Biophysics
Karolinska Institute
Stockholm, Sweden

*Experimental Models for Rheumatoid Arthritis*

†Deceased.
Joyce J. Hsu, MD, MS
Clinical Assistant Professor of Pediatrics
Department of Pediatrics, Division of Pediatric Rheumatology
Stanford University
Palo Alto, California
  Treatment of Juvenile Idiopathic Arthritis

James I. Huddleston, MD
Assistant Professor
Department of Orthopaedic Surgery
Stanford University School of Medicine
Stanford, California
  Hip and Knee Pain

Thomas W.J. Huizinga, MD, PhD
Chairman, Department of Rheumatology
Leiden University Faculty of Medicine
Leiden, The Netherlands
  Early Synovitis and Early Undifferentiated Arthritis

Gene G. Hunder, MD, MS
Professor of Medicine
Department of Medicine, Division of Rheumatology
Mayo Clinic College of Medicine;
Emeritus Consultant in Rheumatology
Mayo Clinic
Rochester, Minnesota
  History and Physical Examination of the Musculoskeletal System

Emily W. Hung, MD
Internal Medicine/Rheumatology
University of Texas Medical School at Houston
Houston, Texas
  Rheumatic Manifestations of Human Immunodeficiency Virus Infection

Robert D. Inman, MD
Professor of Medicine and Immunology
University of Toronto Faculty of Medicine;
Director, Arthritis Centre of Excellence
Toronto Western Hospital
Toronto, Ontario, Canada
  Pathogenesis of Ankylosing Spondylitis and Reactive Arthritis

Maura Daly Iversen, DPT, ScD, MPH
Professor and Chair, Department of Physical Therapy
Northeastern University Bouvé College of Health Sciences;
Senior Lecturer and Behavioral Scientist
Brigham and Women’s Hospital/Harvard Medical School
Boston, Massachusetts
  Introduction to Physical Medicine, Physical Therapy, and Rehabilitation

Johannes W.G. Jacobs, MD, PhD
Associate Professor
Department of Rheumatology and Clinical Immunology
University of Utrecht Faculty of Medicine;
Rheumatologist and Senior Researcher
University Medical Center Utrecht
Utrecht, The Netherlands
  Glucocorticoid Therapy

Joanne M. Jordan, MD, MPH
Herman and Louise Smith Distinguished Professor of Medicine, Professor of Orthopaedics, and Adjunct Professor of Epidemiology
Chief, Division of Rheumatology, Allergy, and Immunology
University of North Carolina School of Medicine;
Director, UNC Thurston Arthritis Research Center
Chapel Hill, North Carolina
  Principles of Epidemiology in Rheumatic Disease; Clinical Features of Osteoarthritis

Joseph L. Jorizzo, MD
Professor and Former (Founding) Chair, Department of Dermatology
Wake Forest University School of Medicine
Winston-Salem, North Carolina
  Behçet’s Disease

Kenton R. Kaufman, PhD, PE
Professor of Biomedical Engineering
Mayo Clinic College of Medicine;
Program Co-Director, Biomechanics and Motion Analysis Lab
Department of Orthopedic Surgery
Mayo Clinic
Rochester, Minnesota
  Biomechanics

William S. Kaufman, MD
Resident Physician
Department of Dermatology
Wake Forest Baptist Medical Center
Winston-Salem, North Carolina
  Behçet’s Disease

Arthur Kavanaugh, MD
Professor of Medicine
Department of Medicine, Division of Rheumatology, Allergy, and Immunology
UC San Diego School of Medicine;
Director, UCSD Center for Innovative Therapy
La Jolla, California
  Anticytokine Therapies

Robert T. Keenan, MD, MPH
Assistant Professor of Medicine
Division of Rheumatology and Immunology
Duke University School of Medicine
Durham, North Carolina
  Etiology and Pathogenesis of Hyperuricemia and Gout
Shaukat Khan, PhD  
Postdoctoral Fellow  
NYU Cancer Institute  
NYU Langone Medical Center  
New York, New York  
Dendritic Cells

Alisa E. Koch, MD  
Frederick G.L. Huetwell and William D. Robinson, MD,  
Professor of Rheumatology  
University of Michigan Medical School  
Ann Arbor, Michigan  
Cell Recruitment and Angiogenesis

Dwight H. Kono, MD  
Professor of Immunology  
Department of Immunology and Microbial Science  
The Scripps Research Institute Kellogg School of Science and Technology  
La Jolla, California  
Autoimmunity

Deborah Krakow, MD  
Professor  
Department of Orthopaedic Surgery and Department of Human Genetics  
David Geffen School of Medicine at UCLA  
Los Angeles, California  
Heritable Diseases of Connective Tissue

Robert G.W. Lambert, MB, FRCR, FRCPC  
Professor of Radiology  
Department of Radiology and Diagnostic Imaging  
University of Alberta Faculty of Medicine and Dentistry  
Edmonton, Alberta, Canada  
Imaging Modalities in Rheumatic Diseases

Robert B.M. Landewé, MD  
Professor of Rheumatology  
University of Amsterdam Faculty of Medicine  
Academic Medical Center  
Amsterdam;  
Consultant, Atrium Medical Center  
Heerlen, The Netherlands  
Clinical Trial Design and Analysis

Nancy E. Lane, MD  
Endowed Professor of Medicine and Rheumatology  
Director, Musculoskeletal and Aging Research Center  
UC Davis School of Medicine  
Davis, California  
Metabolic Bone Disease

Carol A. Langford, MD, MHS  
Director, Center for Vasculitis Care and Research  
Department of Rheumatic and Immunologic Diseases  
Cleveland Clinic  
Cleveland, Ohio  
Primary Angiitis of the Central Nervous System

Daniel M. Laskin, DDS, MS, DSc(Hon)  
Professor and Chairman Emeritus, Department of Oral and Maxillofacial Surgery  
Virginia Commonwealth University Schools of Dentistry and Medicine  
Richmond, Virginia  
Temporomandibular Joint Pain

Ronald M. Laxer, MDCM, FRCPC  
Professor of Pediatrics and Medicine  
University of Toronto Faculty of Medicine;  
Rheumatologist  
The Hospital for Sick Children  
Toronto, Ontario, Canada  
Pediatric Systemic Lupus Erythematosus, Dermatomyositis, Scleroderma, and Vasculitis

David M. Lee, MD  
Head, ATI Translational Research  
Autoimmunity, Transplantation and Inflammation  
Novartis Institutes for BioMedical Research  
Novartis Pharma, AG  
Basel, Switzerland  
Mast Cells

Lela A. Lee, MD  
Professor of Dermatology and Medicine  
University of Colorado School of Medicine;  
Director of Dermatology  
Denver Health Medical Center  
Denver, Colorado  
The Skin and Rheumatic Diseases

Tzielan Chang Lee, MD  
Clinical Associate Professor of Pediatrics  
Department of Pediatrics, Division of Pediatric Rheumatology  
Stanford University  
Palo Alto, California  
Treatment of Juvenile Idiopathic Arthritis

Michael D. Lockshin, MD  
Professor of Medicine and Obstetrics-Gynecology  
Weill Cornell Medical College;  
Director and Attending Physician  
Barbara Volcker Center for Women and Rheumatic Diseases  
Hospital for Special Surgery  
New York, New York  
Antiphospholipid Syndrome

Rik Lories, MD, PhD  
Professor  
Department of Musculoskeletal Sciences, Division of Rheumatology;  
Faculty of Medicine;  
Head, Homeostasis, Regeneration, and Ageing Laboratory for Skeletal Development and Joint Disorders  
Katholieke Universiteit Leuven  
Leuven, Belgium  
Pathogenesis of Ankylosing Spondylitis and Reactive Arthritis
<table>
<thead>
<tr>
<th>Contributors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carlos J. Lozada, MD</td>
<td>Professor</td>
</tr>
<tr>
<td>Department of Medicine, Division of Rheumatology</td>
<td>University of Miami Miller School of Medicine</td>
</tr>
<tr>
<td>Miami, Florida</td>
<td><strong>Treatment of Osteoarthritis</strong></td>
</tr>
<tr>
<td>Ingrid E. Lundberg, MD, PhD</td>
<td>Professor of Medicine</td>
</tr>
<tr>
<td>Department of Medicine, Head, Rheumatology Unit</td>
<td>Karolinska Institute/Karolinska University Hospital, Stockholm, Sweden</td>
</tr>
<tr>
<td><strong>Inflammatory Diseases of Muscle and Other Myopathies</strong></td>
<td></td>
</tr>
<tr>
<td>Raashid Luqmani, BMedSci, BM, BS, MRCP, FRCPed, FRCP, DM</td>
<td>Professor of Rheumatology</td>
</tr>
<tr>
<td>Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Science</td>
<td>University of Oxford, Oxford, United Kingdom</td>
</tr>
<tr>
<td><strong>Polyarteritis Nodosa and Related Disorders</strong></td>
<td></td>
</tr>
<tr>
<td>Frank P. Luyten, MD, PhD</td>
<td>Chairman, Department of Musculoskeletal Sciences, Head, Division of Rheumatology, University Hospitals, Leuven, Belgium</td>
</tr>
<tr>
<td><strong>Regenerative Medicine and Tissue Engineering</strong></td>
<td></td>
</tr>
<tr>
<td>Reuven Mader, MD</td>
<td>Head, Rheumatic Diseases Unit</td>
</tr>
<tr>
<td>Department of Rheumatology</td>
<td>Afula; Associate Clinical Professor of Medicine, B. Rappaport Faculty of Medicine, Technion Institute of Technology, Haifa, Israel</td>
</tr>
<tr>
<td><strong>Proliferative Bone Diseases</strong></td>
<td></td>
</tr>
<tr>
<td>Walter P. Maksymowycz, MD, PhD, FRCP, FACP, FRCP(UK)</td>
<td>Professor of Medicine</td>
</tr>
<tr>
<td>Department of Medicine, Division of Rheumatology</td>
<td>University of Alberta Faculty of Medicine and Dentistry, Edmonton, Alberta, Canada</td>
</tr>
<tr>
<td><strong>Ankylosing Spondylitis</strong></td>
<td></td>
</tr>
<tr>
<td>Brian Mandell, MD</td>
<td>Professor and Chairman, Department of Medicine, Senior Staff, Department of Rheumatic and Immunologic Diseases, Center for Vasculitis Care and Research, Cleveland Clinic, Cleveland, Ohio</td>
</tr>
<tr>
<td><strong>Rheumatologic Manifestations of Hemoglobinopathies</strong></td>
<td></td>
</tr>
<tr>
<td>Scott David Martin, MD</td>
<td>Associate Professor of Orthopedics, Attending Staff Physician, Brigham and Women’s Hospital, Boston, Massachusetts</td>
</tr>
<tr>
<td><strong>Shoulder Pain</strong></td>
<td></td>
</tr>
<tr>
<td>Eric L. Matteson, MD, MPH</td>
<td>Professor of Medicine</td>
</tr>
<tr>
<td>Mayo Clinic College of Medicine, Consultant in Rheumatology, Mayo Clinic, Rochester, Minnesota</td>
<td></td>
</tr>
<tr>
<td><strong>Cancer Risk in Rheumatic Diseases</strong></td>
<td></td>
</tr>
<tr>
<td>Matthew J. McGirt, MD</td>
<td>Assistant Professor of Neurological Surgery, Orthopedic Surgery, and Rehabilitation, Vanderbilt University School of Medicine, Nashville, Tennessee</td>
</tr>
<tr>
<td><strong>Neck Pain</strong></td>
<td></td>
</tr>
<tr>
<td>Iain B. McInnes, PhD, FRCP, FRSE</td>
<td>Muirhead Professor of Medicine, Director, Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom</td>
</tr>
<tr>
<td><strong>Cytokines</strong></td>
<td></td>
</tr>
<tr>
<td>Elizabeth Kaufman McNamara, MD</td>
<td>Dermatologist, Roanoke, Virginia</td>
</tr>
<tr>
<td><strong>Behçet’s Disease</strong></td>
<td></td>
</tr>
<tr>
<td>Ted R. Mikuls, MD, MSPH</td>
<td>Professor of Medicine</td>
</tr>
<tr>
<td>Department of Internal Medicine, Division of Rheumatology, University of Nebraska College of Medicine, UNMC Physician, University of Nebraska Medical Center and Omaha VA Medical Center, Omaha, Nebraska</td>
<td></td>
</tr>
<tr>
<td><strong>Antihyperuricemic Agents</strong></td>
<td></td>
</tr>
<tr>
<td>Mark S. Miller, PhD</td>
<td>Research Associate, Department of Molecular Physiology and Biophysics, The University of Vermont College of Medicine, Burlington, Vermont</td>
</tr>
<tr>
<td><strong>Muscle: Anatomy, Physiology, and Biochemistry</strong></td>
<td></td>
</tr>
</tbody>
</table>
Kevin G. Moder, MD
Associate Professor of Medicine
Department of Medicine, Division of Rheumatology
Mayo Clinic College of Medicine;
Consultant in Rheumatology
Mayo Clinic
Rochester, Minnesota

Kevin G. Moder, MD
Associate Professor of Medicine
Department of Medicine, Division of Rheumatology
Mayo Clinic College of Medicine;
Consultant in Rheumatology
Mayo Clinic
Rochester, Minnesota

History and Physical Examination of the
Musculoskeletal System

Kanneboyina Nagaraju, DVM, PhD
Professor of Integrative Systems Biology and Pediatrics
George Washington University School of Medicine and Health Sciences;
Director, Murine Drug Testing Facility
Center for Genetic Medicine Research
Children’s Research Institute
Children’s National Medical Center
Washington, DC
Inflammatory Diseases of Muscle and Other Myopathies

Stanley J. Naides, MD
Medical Director, Immunology Research and Development
Quest Diagnostics Nichols Institute
San Juan Capistrano, California
Viral Arthritis

Amanda E. Nelson, MD, MSCR
Assistant Professor
Department of Medicine, Division of Rheumatology, Allergy and Immunology
University of North Carolina School of Medicine
Chapel Hill, North Carolina
Clinical Features of Osteoarthritis

Peter A. Nigrovic, MD
Assistant Professor of Medicine
Division of Rheumatology, Immunology, and Allergy
Department of Medicine at Brigham and Women’s Hospital and Harvard Medical School;
Director, Center for Adults with Pediatric Rheumatic Illness (CAPRI)
Brigham and Women’s Hospital;
Division of Immunology
Boston Children’s Hospital
Boston, Massachusetts
Mast Cells

Johannes Nowatzky, MD
Assistant Professor of Medicine
Department of Medicine, Division of Rheumatology
NYU School of Medicine
New York, New York
Etiology and Pathogenesis of Hyperuricemia and Gout

James R. O’Dell, MD
Bruce Professor of Medicine
Vice Chairman, Department of Internal Medicine
University of Nebraska College of Medicine;
Chief, Division of Rheumatology and Immunology
University of Nebraska Medical Center;
Omaha VA
Omaha, Nebraska
Traditional DMARDs: Methotrexate, Leflunomide, Sulfasalazine, Hydroxychloroquine, and Combination Therapies; Treatment of Rheumatoid Arthritis

Yasunori Okada, MD, PhD
Professor of Pathology
Keio University School of Medicine
Tokyo, Japan
Proteinases and Matrix Degradation

Nataly Manjarrez Orduño, PhD
Center for Autoimmune and Musculoskeletal Diseases
The Feinstein Institute for Medical Research
Manhasset, New York
B Cells

Caroline Ospelt, MD
Center for Experimental Rheumatology
Zurich Center for Integrative Human Physiology, Life Science Zurich Graduate School/University Hospital
Zurich
Zurich, Switzerland
Epigenetics

Mikkel Østergaard, MD, PhD, DMSc
Professor of Rheumatology
Department of Orthopaedics and Internal Medicine, Division of Rheumatology
Copenhagen University Faculty of Health and Medical Sciences/Glostrup Hospital
Copenhagen, Denmark
Imaging Modalities in Rheumatic Diseases

Bradley M. Palmer, PhD
Research Assistant Professor
Department of Molecular Physiology and Biophysics
The University of Vermont College of Medicine
Burlington, Vermont
Muscle: Anatomy, Physiology, and Biochemistry

Richard S. Panush, MD, MACP, MACR
Professor of Medicine
Department of Medicine, Division of Rheumatology
Keck School of Medicine of USC
Los Angeles, California
Occupational and Recreational Musculoskeletal Disorders
Stanford L. Peng, MD, PhD  
Assistant Clinical Professor  
Department of Medicine, Division of Rheumatology  
University of Washington School of Medicine;  
Head, Rheumatology Clinical Research Unit  
Benaroya Research Institute at Virginia Mason;  
Physician, Section of Rheumatology  
Virginia Mason Medical Center  
Seattle, Washington  
Antinuclear Antibodies

Michael H. Pillinger, MD  
Associate Professor of Medicine and Pharmacology  
Department of Medicine, Division of Rheumatology  
NYU School of Medicine;  
Section Chief, Rheumatology  
Department of Medicine  
VA New York Harbor Healthcare System, Manhattan Campus  
New York, New York  
Neutrophils; Eosinophils; Etiology and Pathogenesis of Hyperuricemia and Gout

Gregory R. Polston, MD  
Associate Professor of Clinical Anesthesiology  
Department of Anesthesiology  
UC San Diego School of Medicine;  
Associate Medical Director, Center for Pain Medicine  
UC San Diego Medical Center  
La Jolla, California  
Analgesic Agents in Rheumatic Disease

Steven A. Porcelli, MD  
Murray and Evelyne Weinstock Professor of Microbiology and Immunology  
Department of Microbiology and Immunology and Department of Medicine  
Albert Einstein College of Medicine  
Bronx, New York  
Innate Immunity

Mark D. Price, MD, PhD  
Assistant Professor of Orthopedics and Rehabilitation  
University of Massachusetts Medical School;  
Orthopedic Surgeon, Sports Medicine Center  
UMass Memorial Medical Center  
Worcester, Massachusetts  
Foot and Ankle Pain

Johannes J. Rasker, MD, PhD  
Professor Emeritus of Rheumatology  
Department of Psychology and Communication of Health and Risk  
University of Twente Faculty of Behavioural Sciences  
Enschede, The Netherlands  
Fibromyalgia

John D. Reveille, MD  
Professor of Internal Medicine  
Director, Division of Rheumatology and Clinical Immunogenetics  
University of Texas Medical School at Houston;  
Memorial Hermann-Texas Medical Center  
Houston, Texas  
Rheumatic Manifestations of Human Immunodeficiency Virus Infection

W. Neal Roberts, Jr., MD  
Charles W. Thomas Professor and Rheumatology Fellowship Program Director  
Department of Internal Medicine, Division of Rheumatology, Allergy, and Immunology  
Virginia Commonwealth University School of Medicine, Medical College of Virginia Campus  
Richmond, Virginia  
Psychosocial Management of Rheumatic Diseases

Monika Ronneberger, DrMed, DiplBiol  
Medizinische Klinik 3 mit Poliklinik  
University of Erlangen–Nuremberg  
Erlangen, Germany  
Enteropathic Arthritis

Antony Rosen, ChB, BSc, MB  
Mary Betty Stevens Professor of Medicine and Professor of Pathology  
Director, Division of Rheumatology  
Department of Medicine  
Johns Hopkins University School of Medicine  
Baltimore, Maryland  
Autoantibodies in Rheumatoid Arthritis

James T. Rosenbaum, MD  
Professor of Ophthalmology, Medicine and Cell Biology and Edward E Rosenbaum Professor of Inflammation Research  
Department of Ophthalmology  
Oregon Health & Science University School of Medicine  
Portland, Oregon  
The Eye and Rheumatic Diseases

Andrew E. Rosenberg, MD  
Professor of Pathology  
Director, Anatomic Pathology  
Director, Bone and Soft Tissue Pathology  
University of Miami Miller School of Medicine  
Miami, Florida  
Tumors and Tumor-like Lesions of Joints and Related Structures

Eric M. Ruderman, MD  
Professor of Medicine  
Department of Medicine, Division of Rheumatology  
Northwestern University Feinberg School of Medicine;  
Clinical Practice Director, Rheumatology Clinic  
Northwestern Memorial Hospital  
Chicago, Illinois  
Mycobacterial Infections of Bones and Joints; Fungal Infections of Bones and Joints
Contributors

Merja Ruutu, MD
Postdoctoral Fellow
Diamantina Institute for Cancer, Immunology, and Metabolic Medicine
University of Queensland
Princess Alexandra Hospital
Queensland, Australia
Dendritic Cells

Jane E. Salmon, MD
Professor of Medicine
Weill Cornell Medical College;
Co-Director, Mary Kirkland Center for Lupus Research;
Attending Physician, Hospital for Special Surgery
New York, New York
Antiphospholipid Syndrome

Jonathan Samuels, MD
Instructor in Medicine–Rheumatology
NYU University School of Medicine;
Director, Clinical Immunology Laboratory
NYU Langone Medical Center
New York, New York
Pathogenesis of Osteoarthritis

Christy I. Sandborg, MD
Professor of Pediatrics
Department of Pediatrics, Division of Pediatric Rheumatology
Stanford University
Palo Alto, California
Treatment of Juvenile Idiopathic Arthritis

Caroline O.S. Savage, PhD, FRCP, FMedSci
Professor of Nephrology
College of Medical and Dental Sciences, School of Immunity and Infection
University of Birmingham
Birmingham, United Kingdom
Antineutrophil Cytoplasm Antibody–Associated Vasculitis

Amit Saxena, MD
Clinical Instructor in Medicine–Rheumatology
NYU School of Medicine
New York, New York
Acute Phase Reactants and the Concept of Inflammation

Jose U. Scher, MD
Instructor in Medicine–Rheumatology
NYU School of Medicine;
Director, Microbiome Center for Rheumatology and Autoimmunity;
Staff Physician
NYU Langone Medical Center Hospital for Joint Diseases
New York, New York
Neutrophils; Eosinophils

Georg Schett, MD
Professor of Medicine
Chief of Rheumatology
Department of Internal Medicine 3
Institute for Clinical Immunology
University of Erlangen–Nuremberg;
Erlangen, Germany
Biology, Physiology, and Morphology of Bone

David C. Seldin, MD, PhD
Professor of Medicine
Boston University School of Medicine;
Chief, Section of Hematology-Oncology
Boston Medical Center;
Director, Amyloidosis Treatment and Research Program
Boston University School of Medicine/Boston Medical Center
Boston, Massachusetts
Amyloidosis

Anna Simon, MD, PhD
Clinical Investigator
Department of Medicine, Division of General Internal Medicine
Radboud University Nijmegen Faculty of Medical Sciences
Nijmegen, The Netherlands
Familial Autoinflammatory Syndromes

Dawd S. Siraj, MD, MPH&TM
Clinical Associate Professor of Medicine
Brody School of Medicine at East Carolina University;
Director, ECU Physicians International Travel Clinic, Section of Infectious Diseases
Greenville, North Carolina
Bacterial Arthritis

Martha Skinner, MD
Professor of Medicine
Director, Special Programs
Amyloidosis Treatment and Research Program
Boston University School of Medicine
Boston, Massachusetts
Amyloidosis

E. William St. Clair, MD
Professor of Medicine and Immunology
Duke University School of Medicine;
Chief, Division of Rheumatology and Immunology
Duke University Medical Center
Durham, North Carolina
Sjögren’s Syndrome

Lisa K. Stamp, MBChB, PhD, FRACP
Associate Professor
Department of Medicine
Christchurch School of Medicine and Health Sciences
University of Otago Faculty of Medicine
Christchurch, New Zealand
Nutrition and Rheumatic Diseases
John H. Stone, MD, MPH  
Associate Professor of Medicine  
Department of Medicine, Division of Rheumatology  
Harvard Medical School;  
Director, Clinical Rheumatology  
Massachusetts General Hospital  
Boston, Massachusetts  

Classification and Epidemiology of Systemic Vasculitis;  
Immune Complex–Mediated Small Vessel Vasculitis

Rainer H. Straub, MD  
Professor of Experimental Medicine  
Laboratory of Experimental Rheumatology and Neuroendocrine Immunology  
University of Regensburg Faculty of Medicine;  
Department of Internal Medicine I  
University Hospital  
Regensburg, Germany  

Neural Regulation of Pain and Inflammation

Susan E. Sweeney, MD, PhD  
Associate Professor of Medicine  
UC San Diego School of Medicine  
La Jolla, California  

Clinical Features of Rheumatoid Arthritis

Nadera J. Sweiss, MD  
Sarcoidosis and Scleroderma Clinic  
University of Illinois at Chicago  
Chicago, Illinois  

Sarcoidosis

Carrie R. Swigart, MD  
Assistant Professor of Orthopaedics and Rehabilitation  
Yale University School of Medicine  
New Haven, Connecticut  

Hand and Wrist Pain

Deborah Symmons, MD, FFPH, FRCP  
Professor of Rheumatology and Musculoskeletal Epidemiology  
School of Medicine;  
Director, Arthritis Research UK Epidemiology Unit  
School of Translational Medicine  
Musculoskeletal Research Group  
University of Manchester;  
Honorary Consultant Rheumatologist  
Central Manchester University Hospitals NHS Foundation Trust  
Manchester, United Kingdom  

Cardiovascular Risk in Rheumatic Disease

Zoltan Szekanecz, MD, PhD, DSc  
Professor of Rheumatology, Medicine, and Immunology  
Department of Rheumatology  
Institute of Medicine  
University of Debrecen Medical Center  
Debrecen, Hungary  

Cell Recruitment and Angiogenesis

Paul-Peter Tak, MD, PhD  
Professor of Medicine  
Department of Clinical Immunology and Rheumatology  
University of Amsterdam Faculty of Medicine  
Academic Medical Center  
Amsterdam, The Netherlands  

Biologic Markers

Peter C. Taylor, MA, PhD, FRCP  
Norman Collisson Professor of Musculoskeletal Sciences  
Kennedy Institute of Rheumatology  
Botnar Research Centre  
Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences  
University of Oxford  
Oxford, United Kingdom  

Cell-Targeted Biologics and Targets: Rituximab, Abatacept, and Other Biologics

Robert Terkeltaub, MD  
Professor of Medicine  
La Jolla;  
Interim, Chief, Division of Rheumatology, Allergy, and Immunology  
UC San Diego School of Medicine  
San Diego, California  

Calcium Crystal Disease: Calcium Pyrophosphate Dihydrate and Basic Calcium Phosphate

Argyrios N. Theofilopoulos, MD  
Professor and Chair, Department of Immunology and Microbial Science  
The Scripps Research Institute Kellogg School of Science and Technology  
La Jolla, California  

Autoimmunity

Ranjeny Thomas, MD, MBBS  
Professor of Rheumatology  
School of Medicine  
University of Queensland Faculty of Health Sciences;  
Rheumatologist, University of Queensland Diamantina Institute/Princess Alexandra Hospital  
Queensland, Australia  

Dendritic Cells

Thomas S. Thornhill, MD  
Professor of Orthopedics  
Harvard Medical School;  
Chief of Orthopedics  
Brigham and Women’s Hospital  
Boston, Massachusetts  

Shoulder Pain

Karina D. Torralba, MD, MACM, FACP, FACR  
Assistant Professor of Medicine  
Department of Medicine, Division of Rheumatology  
Keck School of Medicine of USC  
Los Angeles, California  

Occupational and Recreational Musculoskeletal Disorders
Michael J. Toth, PhD  
Associate Professor  
Department of Medicine  
The University of Vermont School of Medicine  
Burlington, Vermont  
Muscle: Anatomy, Physiology, and Biochemistry

Peter Tugwell, MD, MSc, FRCPC  
Professor of Medicine  
Department of Medicine  
Ottawa Health Research Institute  
University of Ottawa Faculty of Medicine  
Ottawa, Ontario, Canada  
Assessment of Health Outcomes

Zuhre Tutuncu, MD  
Rheumatologist  
Scripps Coastal Medical Center  
San Diego;  
Voluntary Assistant Clinical Professor of Rheumatology  
UC San Diego School of Medicine  
La Jolla, California  
Anticytokine Therapies

Katherine S. Upchurch, MD  
Clinical Professor of Medicine  
University of Massachusetts Medical School;  
Clinical Chief, Division of Rheumatology  
Department of Medicine  
UMass Memorial Medical Center  
Worcester, Massachusetts  
Hemophilic Arthropathy

Désirée M.F.M. van der Heijde, MD, PhD  
Professor of Rheumatology  
Department of Rheumatology  
Leiden University Faculty of Medicine  
Leiden, The Netherlands  
Clinical Trial Design and Analysis

Annette H.M. van der Helm-van Mil, MD, PhD  
Internist/Rheumatologist  
Leiden University Medical Center  
Leiden, The Netherlands  
Early Synovitis and Early Undifferentiated Arthritis

Sjef M. van der Linden, MD, PhD  
Professor of Rheumatology  
Department of Medicine  
Maastricht University Faculty of Health, Medicine and Life Sciences  
Maastricht, The Netherlands  
Ankylosing Spondylitis

Jos W.M. van der Meer, MD, PhD, FRCP  
Professor of Internal Medicine  
Department of Medicine, Division of General Internal Medicine  
Radboud University Nijmegen Faculty of Medical Sciences  
Nijmegen, The Netherlands  
Familial Autoinflammatory Syndromes

Jacob M. van Laar, MD, PhD  
Professor of Clinical Rheumatology  
Musculoskeletal Research Group  
Institute of Cellular Medicine  
Newcastle University  
Newcastle upon Tyne, United Kingdom  
Immunosuppressive Drugs

John Varga, MD  
John and Nancy Hughes Professor of Medicine  
Northwestern University Feinberg School of Medicine  
Chicago, Illinois  
Etiology and Pathogenesis of Scleroderma

Mark S. Wallace, MD  
Professor of Clinical Anesthesiology  
Department of Anesthesiology  
UC San Diego School of Medicine;  
Program Director, Center for Pain Medicine  
UC San Diego Medical Center  
La Jolla, California  
Analgesic Agents in Rheumatic Disease

David M. Warshaw, PhD  
Professor and Chair, Department of Molecular Physiology and Biophysics  
The University of Vermont College of Medicine  
Burlington, Vermont  
Muscle: Anatomy, Physiology, and Biochemistry

Lucy R. Wedderburn, MD, PhD, FRCP  
Professor in Paediatric Rheumatology  
Rheumatology Unit  
UCL Institute of Child Health  
University College London;  
Consultant in Paediatric Rheumatology  
Rheumatology Unit  
Great Ormond Street Hospital  
London, United Kingdom  
Etiology and Pathogenesis of Juvenile Idiopathic Arthritis

Victoria P. Werth, MD  
Professor of Dermatology  
University of Pennsylvania School of Medicine;  
Chief of Dermatology  
Philadelphia VA Medical Center  
Philadelphia, Pennsylvania  
The Skin and Rheumatic Diseases

Fredrick M. Wigley, MD  
Professor of Medicine  
Associate Director, Division of Rheumatology  
Department of Medicine  
Johns Hopkins University School of Medicine  
Baltimore, Maryland  
Clinical Features and Treatment of Scleroderma
Christopher M. Wise, MD
W. Robert Irby Professor of Medicine
Department of Medicine, Division of Rheumatology,
Allergy, and Immunology
Virginia Commonwealth University School of Medicine,
Medical College of Virginia Campus
Richmond, Virginia
Arthrocentesis and Injection of Joints and Soft Tissue

David Wofsy, MD
Professor of Medicine and Microbiology/Immunology
Department of Medicine
UCSF School of Medicine
San Francisco, California
Clinical Features of Systemic Lupus Erythematosus

Frederick Wolfe, MD
Director, National Data Bank for Rheumatic Diseases;
Clinical Professor of Medicine
Department of Medicine
University of Kansas School of Medicine
Wichita, Kansas
Fibromyalgia

Frank A. Wollheim, MD, PhD, FRCP
Emeritus Professor of Rheumatology
University of Lund Faculty of Medicine
Lund, Sweden
Enteropathic Arthritis

Robert L. Wortmann, MD
Professor of Medicine
Department of Medicine, Section of Rheumatology
Geisel School of Medicine at Dartmouth
Lebanon, New Hampshire
Clinical Features and Treatment of Gout

Edward Yelin, PhD
Professor in Residence of Medicine and Health Policy
Department of Medicine, Division of Rheumatology, and
Philip R. Lee Institute for Health Policy Studies
UCSF School of Medicine
San Francisco, California
Economic Burden of Rheumatic Diseases

David Yu, MD
Emeritus Professor of Medicine
David Geffen School of Medicine at UCLA
Los Angeles, California
Pathogenesis of Ankylosing Spondylitis and Reactive Arthritis

John B. Zabriskie, MD
Professor Emeritus
Rockefeller University
New York, New York
Poststreptococcal Arthritis and Rheumatic Fever

Robert B. Zurier, MD
Professor of Medicine Emeritus
Department of Medicine, Division of Rheumatology
University of Massachusetts Medical School
Worcester, Massachusetts
Prostaglandins, Leukotrienes, and Related Compounds

Anne-Marie Zuurmond, PhD
TNO Quality of Life
Business Unit Biomedical Research
Leiden, The Netherlands
Biologic Markers
Rheumatology continues to evolve and inspire as a discipline that occupies the forefront of molecular medicine and novel targeted therapies. The previous edition of the Textbook built on a proud heritage of excellence but was distinguished by change: new editors, more color, new online access, and many other features. Matching the extraordinary pace of change in our field, this new edition continues a grand tradition by accelerating our commitment to excellence in the face of the changing world of publishing. The most obvious examples are the editors for this edition. Three distinguished and longtime colleagues, “The Three Amigos” who were the heart and soul of the Textbook for a generation, have stepped down: Shaun Ruddy, John Sergent, and Ted Harris. Ted passed away recently but left a legacy that will endure (see dedication to the 9th edition). Finding new editors of such high caliber was daunting, but fortunately we met the challenge when we convinced Jim O’Dell and Sherine Gabriel to join our intrepid crew. They brought incredible new strength and expertise, especially in clinical medicine, clinical trials, outcomes research, and epidemiology.

The 9th edition includes a multitude of new authors and chapters. Improved graphics and more easily accessible online content are also features of this edition. The print edition now limits the number of references because we preferred to use allotted pages for scientific content rather than long lists of articles. The complete citations are, however, still available online.

The initial preparative stages of the book occurred, like the last edition, in Costa Rica, where we slaved away for days on the Table of Contents and in selecting an outstanding group of authors. We admit that some fun and entertainment occurred as well, organized and supervised by Linda Lyons Firestein, MD. We also thank the Elsevier staff who braved the rigors of tropical paradise with us, Pam Hetherington and Janice Gaillard. But mostly we want to thank the authors who put in countless hours writing chapters and putting up with our constant haranguing out of love for our discipline, readers, and students.

We hope that you enjoy the Textbook as much as we enjoyed preparing it. The journey has been a formidable and gratifying collegial effort. Because our “Textbook of Rheumatology Costa Rica” Headquarters was sold in 2011, we are searching the globe for alternative sites when it is time to prepare the 10th edition. Although we do not yet know how the next edition will evolve, one certainty is that it will continue the tradition of excellence.

The Editors
Joints also can be classified according to the connective tissues present. Symphyses have a fibrocartilaginous disk separating bone ends that are joined by firm ligaments (e.g., symphysis pubis and intervertebral joints). In synchondroses, the bone ends are covered with articular cartilage, but no synovium or significant joint cavity is present (e.g., sternomanubrial joint). In syndesmoses, the bones are joined directly by fibrous ligaments without a cartilaginous interface (the distal tibiofibular articulation is the only joint of this type outside the cranial vault). In synostoses, bone bridges are formed between bones, producing ankylosis.

Synovial joints, which are classified further according to their shapes, include ball-and-socket (hip), hinge (interphalangeal), saddle (first carpometacarpal), and plane (patellofemoral) joints. These configurations reflect varying functions, as the shapes and sizes of opposing surfaces determine the direction and extent of motion. The various designs permit flexion, extension, abduction, adduction, or rotation. Certain joints can act in one (humeroulnar), two (wrist), or three (shoulder) axes of motion.

This chapter concentrates on the developmental biology and relationship between structure and function of a “prototypic,” “normal” human diarthrodial joint—the joint most likely to develop arthritis. Most research that has been done concerns the knee because of its accessibility, but other joints are described when appropriate.

**DEVELOPMENTAL BIOLOGY OF THE DIARTHRODIAL JOINT**

Skeletal development is initiated by the differentiation of mesenchymal cells that arise from three sources: (1) neural crest cells of the neural ectoderm that gives rise to craniofacial bones; (2) the sclerotome of the paraxial mesoderm, or somite compartment, which forms the axial skeleton; and (3) the somatopleure of the lateral plate mesoderm, which yields the skeleton of the limbs. The appendicular skeleton develops in the human embryo from limb buds, which first are visible at around 4 weeks of gestation. Structures resembling adult joints are generated at approximately 4 to 7 weeks of gestation. Many other crucial phases of musculoskeletal development follow, including vascularization of
epiphyseal cartilage (8 to 12 weeks), appearance of villous folds in synovium (10 to 12 weeks), evolution of bursae (3 to 4 months), and appearance of periarticular fat pads (4 to 5 months).

The upper limbs develop approximately 24 hours earlier than analogous portions of the lower limbs. Proximal structures, such as the glenohumeral joint, develop before more distal ones, such as the wrist and hand. As a consequence, insults to embryonic development during limb formation affect a more distal portion of the upper limb than of the lower limb. Long bones are formed as a result of replacement of the cartilage template by endochondral ossification. The stages of limb development are well described by O'Rahilly and Gardner and are shown in Figure 1-2. The developmental sequence of events occurring during synovial joint formation and some of the regulatory factors and extracellular matrix components involved are summarized in Figures 1-3 and 1-4.

**Interzone Formation and Joint Cavitation**

The morphology of the developing synovial joint and the process of joint cavitation have been described in many classic studies done on the limbs of mammalian and avian embryos. In the human embryo, cartilage condensations, or chondrifications, can be detected at stage 17, when the embryo is small, approximately 11.7 mm long. In the region of the future joint, following formation of the homogeneous chondrogenic interzone at 6 weeks (stages 18 and 19), a three-layered interzone is formed at approximately 7 weeks (stage 21), which consists of two chondrogenic, perichondrium-like layers that cover the opposing surfaces of the cartilage anlagen and are separated by a narrow band of densely packed cellular blastema that remains and forms the interzone. Cavitation begins in the central interzone at about 8 weeks (stage 23).

Although these cellular events associated with joint formation have been recognized for many years, only recently have the genes regulating these processes been elucidated. These genes include growth differentiation factor (GDF)-5, Wnt-14, bone morphogenetic protein (BMP)-2, BMP-4, BMP-6, BMP-7, and the GDF-BMP antagonists. In addition, joint formation is accompanied by the expression of several fibroblast growth factor (FGF) family members, including FGF-2 and FGF-4. The balance of signaling between BMP and FGF determines the rate of proliferation, adjusting the pace of differentiation. Two transcription factors, Cux-1, a homeobox factor, and the ETS factor ERG/C-1-1, are expressed concurrently with GDF-5 and Wnt-14 at the onset of joint formation. Hartmann and Tabin have proposed two major roles for Wnt-14. First, it acts at the onset of joint formation as a negative regulator of chondrogenesis. Second, it facilitates interzone formation and cavitation by inducing expression of GDF-5 (also known as cartilage-derived morphogenetic protein-1 [CDMP-1]), Wnt-4, chordin, and the hyaluronan receptor, CD44. Paradoxically, application of GDF-5 to developing joints in mouse embryo limbs in organ culture causes joint fusion, suggesting that temporospatial interactions among distinct cell populations are important for the correct
characterize the matrix produced by mesenchymal cells, which switch to the production of types II, IX, and XI collagen that typify the cartilaginous matrix at the time of condensation. The messenger RNAs (mRNAs) encoding the small proteoglycans, biglycan and decorin, may be expressed at this time, but the proteins do not appear until after cavitation in the regions destined to become articular cartilage. The current view is that GDF-5 is required at early stages of condensation, where it stimulates recruitment and differentiation of chondrogenic cells, and later, when its expression is restricted to the interzone. The distribution of collagen types and keratan sulfate in developing avian and rodent joints has been characterized by immunohistochemistry. The stages of diarthrodial joint formation and the temporal pattern of expression of the genes involved in regulation at different stages. BMP, bone morphogenetic protein; COMP, cartilage oligomeric matrix protein; FGF, fibroblast growth factor; FgFR, fibroblast growth factor receptor; Hh, hedgehog; Hox, homeobox; IGF, insulin-like growth factor; Ihh, Indian hedgehog; Lef, lymphoid enhancer binding factor; Ptc, patched; PTHrP, parathyroid hormone–related protein; Runx, runt domain binding protein; Sox, SRY-related high-mobility group-box protein; Stat, signal transducer and activator of transcription; TCF, T cell–specific factor; TGF-β, transforming growth factor-β; VEGF, vascular endothelial growth factor; Wnt, wingless type. Development of long bones from cartilage anlagen. BMP, bone morphogenetic protein; C-1-1, Erg3 variant; CD44, cell determinant 44; Cux, cut-repeat homeobox protein; Erg5, ETS-related gene 5; FGF, fibroblast growth factor; GDF, growth and differentiation factor; Gli, glioma-associated oncogene homolog; Hox, homeobox; IGF, insulin-like growth factor; Ihh, Indian hedgehog; Lmx1b, LIM homeodomain transcription factor 1b; PTHrP, parathyroid hormone–related protein; RA, retinoic acid; r-Fng, radical fringe; Runx, runt domain binding protein; Shh, Sonic hedgehog; Sox, SRY-related high-mobility group-box protein; TCF, T cell–specific factor; TGF-β, transforming growth factor-β; Wnt, wingless type. response. The current view is that GDF-5 is required at early stages of condensation, where it stimulates recruitment and differentiation of chondrogenic cells, and later, when its expression is restricted to the interzone. The distribution of collagen types and keratan sulfate in developing avian and rodent joints has been characterized by immunohistochemistry. Collagen types I and III
cartilage.\textsuperscript{23} Interzone regions are marked by the expression of type IIA collagen by chondrocyte progenitors in the perichondrial layers, type IIB and XI collagens by overt chondrocytes in the cartilage anlagen, and type I collagen in the interzone and in the developing capsule and perichondrium (Figure 1-5).\textsuperscript{24}

The interzone region contains cells in two outer layers that are destined to differentiate into chondrocytes and become incorporated into the epiphyses, and in a thin intermediate zone that are programmed to undergo joint cavitation and may remain as articular chondrocytes.\textsuperscript{25} Fluid and macromolecules accumulate in this space, creating a nascent synovial cavity. Blood vessels appear in the surrounding capsulosynovial blastemal mesenchyme before separation of adjacent articulating surfaces.\textsuperscript{26} Although it was first assumed that these interzone cells should undergo necrosis or programmed cell death (apoptosis),\textsuperscript{27} some investigators have found no evidence of DNA fragmentation preceding cavitation.\textsuperscript{24,25,28,29} Evidence that metalloproteinases are involved in loss of tissue strength in the region undergoing cavitation is also lacking.\textsuperscript{30} Instead, the actual joint cavity seems to be formed by mechanospatial changes induced by the synthesis of hyaluronan via uridine diphosphoglucose dehydrogenase (UDPGD) and hyaluronan synthase. The interaction of hyaluronan with its cell surface receptor, CD44, modulates cell migration, but it is thought that the accumulation of hyaluronan and the associated mechanical influences play a major role in forcing the cells apart and inducing rupture of the intervening extracellular matrix by tensile forces.\textsuperscript{20,30} This mechanism accounts for the observation that joint cavitation is incomplete in the absence of movement.\textsuperscript{31,32} Equivalent data from human embryonic cartilage.\textsuperscript{23}

![Figure 1-5](image-url) In situ hybridization of a 13-day-old (stage 39) chicken embryo middle digit, proximal interphalangeal joint, midfrontal sections. A, Bright-field image shows developing joint and capsule (C). B, Equivalent paraffin section of opposite limb of same animal shows onset of cavitation laterally (arrow). C, Expression of type IIA collagen mRNA in articular surface cells, perichondrium, and capsule. D, Type IIB collagen mRNA is expressed only in chondrocytes of the anlagen. E, Type XI collagen mRNA is expressed in the surface cells, perichondrium, and capsule, with lower levels in chondrocytes. F, Type I collagen mRNA is present in cells of the interzone and capsule. C through F images are dark field. Calibration bar = 1 \( \mu \text{m} \). (From Nalin AM, Greenele TK Jr, Sandell LK: Collagen gene expression during development of avian synovial joints: transient expression of types II and XI collagen genes in the joint capsule, Develop Dyn 203:352–362, 1995.)
joints are difficult to obtain. In all large joints in humans, complete joint cavities are apparent at the beginning of the fetal period.

**CARTILAGE FORMATION AND ENDOCHONDRAL OSSIFICATION**

The skeleton develops from the primitive, avascular, densely packed cellular mesenchyme, termed the skeletal blastema. Common precursor mesenchymal cells divide into chondrogenic, myogenic, and osteogenic lineages that determine the differentiation of cartilage centrally, muscle peripherally, and bone. Surrounding tissues, particularly epithelium, influence the differentiation of mesenchymal progenitor cells to chondrocytes in cartilage anlagen. The cartilaginous nodules appear in the middle of the blastema; simultaneously, cells at the periphery become flattened and elongated to form the perichondrium. In the vertebral column, cartilage disks arise from portions of the somites surrounding the notochord, and nasal and auricular cartilage and the embryonic epiphysis form from the perichondrium. In the limb, the cartilage remains as a resting zone that later becomes the articular cartilage, or it undergoes terminal hypertrophic differentiation to become calcified (growth plate formation) and is replaced by bone (endochondral ossification). The latter process requires extracellular matrix remodeling and vascularization (angiogenesis). These events are controlled exquisitely by cellular interactions with the surrounding matrix, growth and differentiation factors, and other environmental factors that initiate or suppress cellular signaling pathways and transcription of specific genes in a temporospatial manner.

**Condensation and Limb Bud Formation**

Formation of cartilage anlagen occurs in four stages: (1) cell migration, (2) aggregation regulated by mesenchymal-epithelial cell interactions, (3) condensation, and (4) overt chondrocyte differentiation, or chondrification. Interactions with the epithelium determine mesenchymal cell recruitment and migration, proliferation, and condensation. The aggregation of chondrogenic mesenchymal cells into precartilage condensations was first described by Fell and depends on signals initiated by cell-cell and cell-matrix interactions, the formation of gap junctions, and changes in the cytoskeletal architecture. Before condensation, the prechondrocytic mesenchymal cells produce extracellular matrix that is rich in hyaluronan and type I collagen and type II collagen, which contains the exon-2-encoded aminopropeptide found in noncartilage collagens. The initiation of condensation is associated with increased hyaluronidase activity and the appearance of cell adhesion molecules, neural cadherin (N-cadherin), and the neural cell adhesion molecule (N-CAM), all of which facilitate cell-cell interactions.

Before chondrocyte differentiation, cell-matrix interactions are facilitated by fibronectin binding to syndecan, downregulating N-CAM and setting condensation boundaries. Increased cell proliferation and extracellular matrix remodeling, with the disappearance of type I collagen, fibronectin, and N-cadherin, and the appearance of tenascins, matrilins, and thrombospondins, including cartilage oligomeric protein, initiate the transition from chondrogenic progenitor cells to a fully committed chondrocyte. N-cadherin and N-CAM disappear in differentiating chondrocytes and are detectable later only in perichondrial cells. The differentiated chondrocytes can proliferate and undergo the complex process of hypertrophic maturation or remain within cartilage elements in articular joints.

Zwilling proposed that positional information for organization of the limb bud was impacted by diffusible agents generated at the tip of the limb bud and along its posterior margin, promoting the development of a cartilaginous anlage along proximal-distal and anterior-posterior axes. Limb buds develop from the lateral plate mesoderm. Patternning of limb mesenchyme is the result of interactions between the mesenchyme and the overlying epithelium. The embryonic limb possesses two signaling centers: the apical ectodermal ridge (AER) and the zone of polarizing activity (ZPA), which produce signals responsible for directing proximal-distal outgrowth (AER) and anterior-posterior patterning (ZPA).

Much of our current understanding of limb development is based on early studies in chickens and more recently in mice. Regulatory events are controlled by interacting patterning systems involving FGF, hedgehog, BMP, and Wnt pathways, each of which functions sequentially over time (see Figure 1-3). Wnt signaling via β-catenin is required to induce FGFs, such as FGF-10 and FGF-8, which act in positive feedback loops. BMP-2, BMP-4, and BMP-8 (induced by Wnt-3A), from specialized epithelial cells in the AER that are covering the limb bud tip, control proximal-distal (shoulder/finger) outgrowth. The homeobox (Hox) transcription factors encoded by HoxA and HoxD gene clusters, which are crucial for early events of limb patterning in the undifferentiated mesenchyme, are required for the expression of FGF-8 and Sonic hedgehog (Shh), and they modulate the proliferation of cells within the condensations. Among the Hox genes, Hoxa13 and Hoxd13 enhance and Hoxa11 and Hoxd11 suppress early events in the formation of cartilage anlagen.

Wnt-7A is expressed early during limb bud development, where it acts to maintain Shh expression. Shh, produced by a small group of cells in the posterior zone of the ZPA (in response to retinoic acid in the mesoderm and FGF-4 in the AER), plays a key role in directing anterior-posterior (e.g., little finger/thumb) patterning and in stimulating expression of BMP-2, BMP-4, BMP-7, and Hox genes. Shh signaling, which is required for early limb patterning, but not for limb formation, is mediated by the Shh receptor Patched (Ptc1), which activates another transmembrane protein, Smoothered (Smo), and inhibits processing of the GlI3 transcription factor to a transcriptional repressor. Dorsal-ventral (e.g., knuckles/palm) patterning depends on secretion of Wnt-7A and expression of the following transcription factors: radical fringe (β-Fng) by the dorsal ectoderm, and engrailed (En-1) and Lmx1b (which is induced by Wnt-7A) by the ventral endoderm. BMP-2, BMP-4, and BMP-7 coordinately regulate the patterning of limb elements within condensations depending on the temporal and spatial expression of BMP receptors and BMP antagonists, such as noggin and chordin, as well as the availability of SMADs (signaling mammalian homologs of Drosophila mothers against decapentaplegic).
In vitro and in vivo studies have shown that BMP signaling is required for the formation of precartilaginous condensations and for the differentiation of precursors into chondrocytes. Growth of the condensation ceases when noggin inhibits BMP signaling and permits overt differentiation to chondrocytes, which often are designated as chondroblasts. The cartilage formed serves as a template for formation of cartilage elements in the vertebra, sternum, and rib, and for limb elongation or endochondral bone formation.

**Molecular Signals in Cartilage Morphogenesis and Growth Plate Development**

The cartilage anlagen grow by cell division and deposition of the extracellular matrix and by apposition of proliferating cells from the inner chondrogenic layer of the perichondrium. The nuclear transcription factor, Sox9, is one of the earliest markers expressed in cells undergoing condensation and is required for the subsequent stage of chondrogenesis characterized by the deposition of matrix containing collagens II, IX, and XI and aggrecan in the cartilage anlagen. Two additional Sox family members, L-Sox5 and Sox6, which are not present in early mesenchymal condensations but are coexpressed with Sox9 during chondrocyte differentiation, have a high degree of sequence identity with each other but have no sequence homology with Sox9, except in the high-mobility group (HMG) box. They can form homodimers or heterodimers, which bind more efficiently to pairs of HMG box sites than to single sites, and in contrast to Sox9, they maintain no transcriptional activation domain. The expression of SOX proteins depends on BMP signaling via BMPR1A and BMPR1B, which are functionally redundant and active in chondrocyte condensations but not in the perichondrium. L-Sox5 and Sox6 are required for the expression of Col9a1, aggrecan, link protein, and Col2a1 during overt chondrocyte differentiation.

The runt domain transcription factor, Runx2 (also known as core binding factor, Cbfα1), is expressed in all condensations, including those that are destined to form bone.

Throughout chondrogenesis, the balance of signaling by BMPs and FGFs determines the rate of proliferation while adjusting the pace of differentiation. In the long bones, long after condensation, BMP-2, BMP-3, BMP-4, BMP-5, and BMP-7 are expressed primarily in the perichondrium, and only BMP-7 is expressed in the proliferating chondrocytes. BMP-6 is found later exclusively in hypertrophic chondrocytes along with BMP-2. More than 23 FGFs have been identified so far. The specific ligands that activate each FGF receptor (R) during chondrogenesis in vivo have been difficult to identify because signaling depends on the temporal and spatial location of not only the ligands, but also the receptors. FGF2 is upregulated early in condensing mesenchyme and is present later in the periphery of the condensation along with FGF1, which is expressed in surrounding loose mesenchyme. FGF3 is associated with proliferation of chondrocytes in the central core of the mesenchymal condensation and may overlap with FGF2. Proliferation of chondrocytes in the embryonic and postnatal growth plate is regulated by multiple mitogenic stimuli, including FGFs, which converge on cyclin D1.

In the growth plate, FGFR3 serves as a master inhibitor of chondrocyte proliferation via phosphorylation of the Stat1 transcription factor, which increases expression of the cell cycle inhibitor p21. More recent studies suggest that FGF-18 is the preferred ligand of FGFR3 because Fgf18-deficient mice have an expanded zone of proliferating chondrocytes similar to that in Fgfr3-deficient mice, and that FGF-18 can inhibit Indian hedgehog (Ihh) expression. FGF-18 and FGF-9 are expressed in the perichondrium and periosteum and form a functional gradient from the proximal proliferating zone, where FGF-18 acts via FGFR3 to downregulate proliferation and subsequent maturation. FGF-18 and FGF-9 interact with FGFR1 in the prehypertrophic and hypertrophic zones, where more recent evidence indicates that they regulate vascular invasion by inducing the expression of vascular endothelial growth factor (VEGF) and VEGFR1. As the epiphyseal growth plate develops, FGFR3 disappears, and FGFR1 expression is upregulated in prehypertrophic and hypertrophic chondrocytes, suggesting a role for FGFR1 in the regulation of cell survival and differentiation and possibly cell death.

The proliferation of chondrocytes in the lower proliferative and prehypertrophic zones is under the control of a local negative feedback loop involving signaling by parathyroid hormone–related protein (PTHrP) and Ihh. Ihh expression is restricted to the prehypertrophic zone, and the PTHrP receptor is expressed in the distal zone of periarticular chondrocytes. Adjacent, surrounding perichondrial cells express the Hedgehog receptor patched (Ptc), which on Ihh binding, similar to Shh in the mesenchymal condensations, activates Smo and induces Gli transcription factors; this can feedback regulate Ihh target genes in a positive (Gli1 and Gli2) or negative (Gli3) manner. Ihh induces expression of PTHrP in the perichondrium, and PTHrP signaling stimulates cell proliferation via its receptor expressed in the periarticular chondrocytes. These interactions are modulated by a balance of BMP and FGF signaling that adjusts the pace of chondrocyte terminal differentiation to the proliferation rate. FGF-18 or FGFR3 signaling can inhibit Ihh expression, and BMP signaling upregulates the expression of Ihh in cells that are beyond the range of the PTHrP-induced signal. Evidence indicates that Ihh acts independently of PTHrP on periarticular chondrocytes to stimulate differentiation of columnar chondrocytes in the proliferative zone, whereas PTHrP acts by preventing premature differentiation into prehypertrophic and hypertrophic chondrocytes, suppressing premature expression of Ihh. Ihh and PTHrP, by transiently inducing proliferation markers and repressing differentiation markers, function in a temporospatial manner to determine the number of cells that remain in the chondrogenic lineage versus the number that enter the endochondral ossification pathway.

**Endochondral Ossification**

The development of long bones from the cartilage anlagen occurs by a process termed endochondral ossification, which involves terminal differentiation of chondrocytes to the hypertrophic phenotype, cartilage matrix calcification, vascular invasion, and ossification (see Figure 1-4). This process is initiated when cells in the central region of the
anlage begin to hypertrophy, increasing cellular fluid volume by almost 20 times. Ihh plays a pivotal role in regulating endochondral bone formation by synchronizing perichondrial maturation with chondrocyte hypertrophy, which is essential for initiating the process of vascular invasion.\(^{18,26}\) Ihh is expressed in prehypertrophic chondrocytes as they exit the proliferative phase and enter the hypertrophic phase, at which time they begin to express the hypertrophic chondrocyte marker, type X collagen (Col10a1), and alkaline phosphatase. These cells are responsible for laying down the cartilage matrix, which subsequently undergoes mineralization.

Runx2, which serves as a positive regulatory factor in chondrocyte maturation to hypertrophy,\(^{81}\) is expressed in the adjacent perichondrium and in prehypertrophic chondrocytes, but less in late hypertrophic chondrocytes,\(^{82,83}\) overlapping with Ihh, Col10a1, and BMP-6.\(^{77,84}\) BMP-induced Smad1 interacts with Runx2, and Runx2 and Smad1 are important for chondrocyte hypertrophy.\(^{81,85,86}\) An essential role for Runx2 in the process of chondrocyte hypertrophy is supported by the observation that terminal differentiation is blocked in Runx2-deficient mice.\(^{84,87}\) Interactions with components of the extracellular matrix also contribute to regulation of the process of chondrocyte terminal differentiation. Matrix metalloproteinase (MMP)-13, a downstream target of Runx2, is expressed by terminal hypertrophic chondrocytes,\(^{88,91}\) and MMP-13 deficiency results in significant interstitial collagen accumulation, leading to a delay in endochondral ossification in the growth plate with increased length of the hypertrophic zone.\(^{92,93}\)

In contrast, Col10a1 knockout mice and transgenic mice with a dominant interference Col10a1 mutation have subtle growth plate phenotypes with compressed proliferative and hypertrophic zones and altered mineral deposition.\(^{80}\) Mutations in the COL10a1 gene are associated with the dwarfism observed in human chondrodysplasias. These mutations affect regions of the growth plate that are under great mechanical stress, and it has been suggested that the defect in skeletal growth may be due in part to alteration of the mechanical integrity of the pericellular matrix in the hypertrophic zone, although a role for defective vascularization also has been proposed.\(^{95}\) The extracellular matrix remodeling that accompanies chondrocyte terminal differentiation is thought to induce an alteration in the environmental stress experienced by hypertrophic chondrocytes, which eventually undergo apoptosis.\(^{77,86,97}\) Together these studies indicate that the composition and remodeling of the extracellular matrix play an important role in processes associated with chondrocyte hypertrophy, vascular invasion, and, as discussed subsequently, osteoblast recruitment and subsequent bone formation.\(^{90}\)

Vascular invasion of the hypertrophic zone is required for the replacement of calcified cartilage by bone.\(^{94,91}\) The angiogenic factor, VEGF, promotes vascular invasion by specifically activating localized receptors, including Flk expressed in endothelial cells in the perichondrium or surrounding soft tissues, neuropilin 1 (Npn1) expressed in late hypertrophic chondrocytes, or Npn2 expressed exclusively in the perichondrium.\(^{92}\) VEGF is expressed as three different isoforms: VEGF188, a matrix-bound form, is essential for metaphyseal vascularization, whereas the soluble form, VEGF120 (VEGFA), regulates chondrocyte survival and epiphyseal cartilage angiogenesis.\(^{96,100}\) VEGF164 can be soluble or matrix bound and may act directly on chondrocytes via Npn2. VEGF is released from the extracellular matrix by MMPs, including MMP-9, membrane-type (MT)1-MMP (MMP-14), and MMP-13. MMP-9 is expressed by endothelial cells that migrate into the central region of the hypertrophic cartilage.\(^{97}\) MMP-14, which has a broader range of expression than MMP-9, is essential for chondrocyte proliferation and secondary ossification,\(^{101}\) whereas MMP-13 is found exclusively in late hypertrophic chondrocytes.\(^{82}\) These events of cartilage matrix remodeling and vascular invasion are required for the migration and differentiation of osteoblasts and osteoclasts, which remove the mineralized cartilage matrix and replace it with bone.

### Development of the Joint Capsule and Synovium

The interzone and the contiguous perichondrial envelope, of which the interzone is a part, contain the mesenchymal cell precursors that give rise to other joint components, including the joint capsule, synovial lining, menisci, intracapsular ligaments, and tendons.\(^{3,4,102,103}\) The external mesenchymal tissue condenses as a fibrous capsule. The peripheral mesenchyme becomes vascularized and is incorporated as the synovial mesenchyme, which differentiates into a pseudomembrane at about the same time as cavitation begins in the central interzone (stage 23, approximately 8 weeks). The menisci arise from eccentric portions of the articular interzone. In common usage, the term synovium refers to the true synovial lining and the subjacent vascular and areolar tissue, up to—but excluding—the capsule. Synovial lining cells can be distinguished as soon as multiple cavities within the interzone begin to coalesce. At first, these are exclusively fibroblast-like (type B) cells.

As the joint cavity increases in size, synovial lining cell layers expand through proliferation of fibroblast-like cells and recruitment of macrophage-like (type A) cells from the circulation.\(^{104}\) The synovial lining cells express the hyaluronan receptor, CD44, and UDPGD, the levels of which remain elevated after cavitation. This increased activity likely contributes to the high concentration of hyaluronic acid in joint fluids.\(^{30,105}\) Further synovial expansion results in the appearance of synovial villi at the end of the second month, early in the fetal period, which greatly increases the surface area available for exchange between the joint cavity and the vascular space. Cadherin-11 is an additional molecule expressed by synovial lining cells.\(^{106,107}\) It is essential for establishment of synovial lining architecture during development, where its expression correlates with cell migration and tissue outgrowth of the synovial lining.

The role of innervation in the developing joint is not well understood. A dense capillary network develops in the subsynovial tissue, with numerous capillary loops that penetrate into the true synovial lining layer. The human synovial microvasculature is already innervated by 8 weeks (stage 23) of gestation, around the time of joint cavitation,\(^{102}\) as is shown by immunoreactivity for the neuronal “housekeeping” enzymes.\(^{108}\) Evidence of neurotransmitter
function is not found until much later, however, with the appearance of the sensory neuropeptide, substance P, at 11 weeks. The putative sympathetic neurotransmitter, neuropeptide Y, appears at 13 weeks of gestation, along with the catecholamine-synthesizing enzyme tyrosine hydroxylase. The finding that the Slit2 gene, which functions for the guidance of neuronal axons and neurons, is expressed in the mesenchyme adjacent to the AER (stages 20 to 22) and in the peripheral mesenchyme of the limb bud (stages 23 to 28) suggests that innervation is an integral part of synovial joint development.\textsuperscript{109}

**Development of Nonarticular Joints**

In contrast to articular joints, the temporomandibular joint develops slowly, with cavitation at a crown-rump length of 57 to 75 mm (i.e., well into the fetal stage).\textsuperscript{110} This slow development may occur because this joint develops in the absence of a continuous blastema and involves the insertion between bone ends of a fibrocartilaginous disk that arises from muscular and mesenchymal derivatives of the first pharyngeal arch.

The development of other types of joints, such as synarthroses, is similar to that of diarthrodial joints, except that cavitation does not occur and synovial mesenchyme is not formed. In these respects, synarthroses and amphiarthroses resemble the “fused” peripheral joints induced by paralyzing chick embryos,\textsuperscript{111} and they may develop as they do because there is relatively little motion during their formation.

Human vertebrae and intervertebral disks develop as units, each derived from a homogeneous blastema arising from a somite. Each embryonic intervertebral disk serves as a rostral and caudal chondrogenic zone for the two adjacent evolving vertebral bodies. The periphery of the embryonic “disk” is replaced by the annulus fibrosus.\textsuperscript{112} The intervertebral disk bears many similarities to the joint; the annulus is the joint capsule, the nucleus pulposus is the joint cavity, and the vertebral end plates are the cartilage-covered bone ends composing the articulation. The proteoglycans and collagens expressed during development of the intervertebral disk have been mapped and reflect the complex structure-function relationships that allow flexibility and resistance to compression in the spine.\textsuperscript{113-116}

**Development of Articular Cartilage**

In the vertebrate skeleton, cartilage is the product of cells from three distinct embryonic lineages. Craniofacial cartilage is formed from cranial neural crest cells, the cartilage of the axial skeleton (intervertebral disks, ribs, and sternum) forms from paraxial mesoderm (somites), and the articular cartilage of the limbs is derived from the lateral plate mesoderm.\textsuperscript{7} In the developing limb bud, mesenchymal condensations, followed by chondrocyte differentiation and maturation, occur in digital zones, whereas undifferentiated mesenchymal cells in the interdigital web zones undergo cell death.\textsuperscript{117} Embryonic cartilage is destined for one of several fates: It can remain as permanent cartilage, as on the articular surfaces of bones, or it can provide a template for the formation of bones by endochondral ossification. During development, chondrocyte maturation expands from the central site of the original condensation, which forms the cartilage anlage resembling the shape of the future bone, toward the ends of the forming bones. During joint cavitation, the peripheral interzone is absorbed into each adjacent cartilaginous zone, evolving into the articular surface. The articular surface is destined to become a specialized cartilaginous structure that does not normally undergo vascularization and ossification.

More recent evidence indicates that postnatal maturation of the articular cartilage involves an appositional growth mechanism originating from progenitor cells at the articular surface, rather than by an interstitial mechanism.\textsuperscript{114} The chondrocytes of mature articular cartilage are terminally differentiated cells that are capable of expressing cartilage-specific matrix molecules, such as type II collagen and aggrecan (see following section).\textsuperscript{19,21,24} Through the processes described previously, the articular joint spaces are developed and are lined on all surfaces by cartilage or by synovial lining cells. These two different tissues merge at the enthesis, the region at the periphery of the joint where the cartilage melds into bone, and where ligaments and the capsule are attached.\textsuperscript{118} In the postnatal growth plate, differentiation of the perichondrium also is linked to differentiation of the chondrocytes in the epiphysis to form the different zones of the growth plate, contributing to longitudinal bone growth.\textsuperscript{25,27}

**ORGANIZATION AND PHYSIOLOGY OF THE MATURE JOINT**

The unique structural properties and biochemical components of diarthrodial joints make them extraordinarily durable load-bearing devices.\textsuperscript{119} The mature diarthrodial joint is a complex structure, influenced by its environment and by mechanical demands (see Chapter 6). Structural differences between joints are determined by their different functions. The shoulder joint, which demands an enormous range of motion, is stabilized primarily by muscles, whereas the hip, requiring motion and antigravity stability, has an intrinsically stable ball-and-socket configuration. Components of the “typical” synovial joint include synovium, muscles, tendons, ligaments, bursae, menisci, articular cartilage, and subchondral bone. The anatomy and physiology of muscles are described in detail in Chapter 5.

**Synovium**

The synovium lines the joint cavity and is the sight of production of synovial fluid that provides nutrition for the articular cartilage and lubricates the cartilage surfaces. It is a thin membrane between the fibrous joint capsule and the fluid-filled synovial cavity that attaches to skeletal tissues at the bone-cartilage interface and does not encroach on the surface of the articular cartilage. It is divided into functional compartments: the lining region (synovial intima), the subintimal stroma, and the neurovasculature (Figure 1-6). The synovial intima, also termed synovial lining, is the superficial layer of the normal synovium that is in contact with the intra-articular cavity.\textsuperscript{105,120} The synovial lining is loosely attached to the subintima, which contains blood vessels, lymphatics, and nerves. Capillaries and arterioles generally
are located directly underneath the synovial intima, whereas venules are located closer to the joint capsule.

A transition from loose to dense connective tissue occurs from the joint cavity to the capsule. Most cells in the normal subintimal stroma are fibroblasts and macrophages, although adipocytes and occasional mast cells are present. These compartments are not circumscribed by basement membranes but nonetheless have distinct functions; they are separated from each other by chemical barriers, such as membrane peptidases, which limit the diffusion of regulatory factors between compartments. Synovial compartments are unevenly distributed within a single joint. Vascularity is high at the enthesis, where synovium, ligament, and cartilage coalesce. Far from being a homogeneous tissue in continuity with the synovial cavity, synovium is highly heterogeneous, and synovial fluid may be poorly representative of the tissue-fluid composition of any synovial tissue compartment. In rheumatoid arthritis, the synovial lining of diarthrodial joints is the site of the initial inflammatory process. This lesion is characterized by proliferation of synovial lining cells, increased vascularization, and infiltration of the tissue by inflammatory cells, including lymphocytes, plasma cells, and activated macrophages (see Chapter 53).

Synovial Lining

The synovial lining, a specialized condensation of mesenchymal cells and extracellular matrix, is located between the synovial cavity and the stroma. In normal synovium, the lining layer is two to three cells deep, although intra-articular fat pads usually are covered by only a single layer of synovial cells, and ligaments and tendons are covered by synovial cells that are widely separated. At some sites, lining cells are absent, and extracellular connective tissue constitutes the lining layer. Such “bare areas” become increasingly frequent with advancing age. Although the synovial lining is often referred to as the synovial membrane, the term membrane is more correctly reserved for endothelial and epithelial tissues that have basement membranes, tight intercellular junctions, and desmosomes. Instead, synovial lining cells lie loosely in a bed of hyaluronate interspersed with collagen fibrils. This is the macromolecular sieve that imparts the semipermeable nature of the synovium. The absence of any true basement membrane is a major determinant of joint physiology.

Early electron microscopic studies characterized lining cells as macrophage-derived type A synoviocytes and fibroblast-derived type B synoviocytes. High UDGD activity and CD55 are used to distinguish type B synovial cells, whereas nonspecific esterase and CD68 typify type A cells. Normal synovium is lined predominantly by fibroblast-like cells, whereas macrophage-like cells account for only 10% to 20% of lining cells (see Figure 1-6).

Type A, macrophage-like synovial cells contain vacuoles, a prominent Golgi apparatus, and filopodia, but they have little rough endoplasmic reticulum. These cells express numerous cell surface markers of the monocyte-macrophage lineage, including CD11b, CD68, CD14, CD163, and the immunoglobulin (Ig)G Fc receptor, FcγRIIa. Synovial intimal macrophages are phagocytic and may provide a mechanism by which particulate matter can be cleared from the normal joint cavity. Similar to other tissue macrophages, these cells have little capacity to proliferate and are likely localized to the joint during development. The op/op osteopetrotic mouse that is deficient in macrophages because of an absence of macrophage colony-stimulating factor also lacks synovial macrophages. This finding provides further evidence that type A synovial cells are of a common lineage with other tissue macrophages. Although they represent only a small percentage of cells in the normal synovium, macrophages are recruited from the circulation during synovial inflammation, in part from subchondral bone marrow through vascular channels near the enthesis.

The type B, fibroblast-like synovial cell contains fewer vacuoles and filopodia than type A cells and has abundant protein-synthetic organelles. Similar to other fibroblasts, lining cells express the collagen synthesis enzyme and synthesize extracellular matrix components, including collagens, sulfated proteoglycans, fibronectin, fibrillin-1, and...
tenascin. They have the potential to proliferate, although proliferation markers are rarely seen in normal synovium. In contrast to stromal fibroblasts, synovial intimal fibroblasts express UDPGD and synthesize hyaluronic acid, an important constituent of synovial fluid. They also synthesize lubricin, which, together with hyaluronic acid, is necessary for the low-friction interaction of cartilage surfaces in the diarthrodial joint. Synovial lining cells bear abundant membrane peptidases on their surface, capable of degrading a wide range of regulatory peptides, such as substance P and angiotensin II. These enzymes may be important in limiting the diffusion of these potent peptide mediators away from the immediate vicinity of their site of release and action.

Normal synovial lining cells also express a rich array of adhesion molecules, including CD44, the principal receptor for hyaluronan; vascular cell adhesion molecule (VCAM)-1; and intercellular adhesion molecule (ICAM)-1. They are essential for cellular attachment to specific matrix components in the synovial lining region, preventing loss into the synovial cavity of cells subjected to deformation and shear stresses during joint movement. Adhesion molecules such as VCAM-1 and ICAM-1 potentially are involved in the recruitment of inflammatory cells during the evolution of arthritis. Cadherins mediate cell-cell adhesion between adjacent cells of the same type. The identification of cadherin-11 as a key adhesion molecule that regulates formation of the synovial lining during development and the synoviocyte function postnatally has provided the opportunity to examine its role in inflammatory joint disease. Recent studies have shown that cadherin-11 is highly expressed in fibroblast-like cells at the pannus-cartilage interface in rheumatoid synovium, where it plays a role in the invasive properties of the synovial fibroblasts; treatment with a cadherin-11 antibody or a cadherin-11 fusion protein has been shown to reduce synovial inflammation and cartilage erosion in an animal model of arthritis.

Synovial Vasculature

The subintimal synovium contains blood vessels, which provide the blood flow required for solute and gas exchange in the synovium itself and for generation of synovial fluid. The avascular articular cartilage also depends on nutrition in the synovial fluid, derived from the synovial vasculature. The vascularized synovium behaves similarly to an endocrine organ, generating factors that regulate synoviocyte function and serving as a selective gateway that recruits cells from the circulation during stress and inflammation. Finally, synovial blood flow plays an important role in regulating intra-articular temperature.

The synovial vasculature can be divided, on morphologic and functional grounds, into arterioles, capillaries, and venules. In addition, lymphatics accompany arterioles and larger venules. Arterial and venous networks of the joint are complex and are characterized by artheriovenous anastomoses that communicate freely with blood vessels in periosteum and periarticular bone. As large synovial arteries enter the deep layers of the synovium near the capsule, they give off branches, which bifurcate to form microvascular units in the subsynovial layers. The synovial lining region, the surfaces of intra-articular ligaments, and the entheses (in the angle of ligamentous insertions into bone) are particularly well vascularized.

The distribution of synovial vessels, which were formed largely as a result of vasculogenesis during development of the joint, displays considerable plasticity. Vasculogenesis is a dynamic process that depends on cellular interactions with regulatory factors and the extracellular matrix that are also important in angiogenesis. In inflammatory arthritis, the density of blood vessels decreases relative to the growing synovial mass, creating a hypoxic and acidic environment. Angiogenic factors such as VEGF, acting via VEGF receptors 1 and 2 (Flt-1 and Flk-1), and basic FGF promote proliferation and migration of endothelial cells—a process that is facilitated by matrix-degrading enzymes and adhesion molecules such as integrin αvβ3 and E-selectin, expressed by activated endothelial cells. Vessel maturation is facilitated by angiopoietin-1 acting via the Tie-2 receptor. Angiogenic molecules are restricted to the capillary epithelium in normal synovium, but their levels are elevated in inflamed synovium in periarticular sites and areas remote from vessels.

Regulation of Synovial Blood Flow

Synovial blood flow is regulated by intrinsic (autocrine and paracrine) and extrinsic (neural and humoral) systems. Locally generated factors, such as the peptide vasoconstrictors angiotensin II and endothelin-1, act on adjacent arteriolar smooth muscle to regulate regional vascular tone. Normal synovial arterioles are richly innervated by sympathetic nerves containing vasoconstrictors, such as norepinephrine and neuropeptide Y, and by “sensory” nerves that also play an efferent vasodilatory role by releasing neuropeptides, such as substance P and calcitonin gene–related peptide. Arterioles regulate regional blood flow. Capillaries and postcapillary venules are sites of fluid and cellular exchange. Correspondingly, regulatory systems are differentially distributed along the vascular axis. Angiotensin-converting enzyme, which generates angiotensin II, is localized predominantly in arteriolar and capillary endothelium and is decreased during inflammation. Specific receptors for angiotensin II and for substance P are abundant on synovial capillaries, with lower densities on adjacent arterioles. Dipeptidyl peptidase IV, a peptide-degrading enzyme, is specifically localized to the cell membranes of venular endothelium. The synovial vasculature not only is functionally compartmentalized from the surrounding stroma, but is also highly specialized along its arteriovenous axis. Other unique characteristics of the normal synovial vasculature include the presence of inducible nitric oxide synthase–independent 3-nitrotyrosine, a reaction product of peroxynitrite, and localization of the synoviocyte-derived CXCL12 chemokine on heparan sulfate receptors on endothelial cells, suggesting physiologic roles for these molecules in normal vascular function.

Joint Innervation

Dissection studies have shown that each joint has a dual nerve supply, consisting of specific articular nerves that penetrate the capsule as independent branches of adjacent
peripheral nerves and articular branches that arise from related muscle nerves. The definition of joint position and the detection of joint motion are monitored separately and through a combination of multiple inputs from different receptors in varied systems. Nerve endings in muscle and skin and in the joint capsule mediate the sensation of joint position and movement. Normal joints have afferent (sensory) and efferent (motor) innervations. Fast-conducting, myelinated A fibers innervating the joint capsule are important for proprioception and detection of joint movement; slow-conducting, unmyelinated C fibers transmit diffuse pain sensation and regulate synovial microvascular function.

Normal synovium is richly innervated by fine, unmyelinated nerve fibers that follow the courses of blood vessels and extend into the synovial lining layers. These nerve fibers do not have specialized endings and are slow-conducting fibers; they may transmit diffuse, burning, or aching pain sensation. Sympathetic nerve fibers surround blood vessels, particularly in the deeper regions of normal synovium. They contain and release classic neurotransmitters, such as noradrenaline, and neuropeptides that constrict synovial blood vessels. Neuropeptides that are markers of sensory nerves include substance P, calcitonin gene–related peptide, neuropeptide Y, and vasoactive intestinal peptide.

Afferent nerves containing substance P also have an efferent role in the synovium. Substance P is released from peripheral nerve terminals into the joint, and specific, G protein–coupled receptors for substance P are localized to microvascular endothelium in normal synovium. Abnormalities of articular innervation that are associated with inflammatory arthritis may contribute to the failure of synovial inflammation to resolve. Excessive local neuropeptide release may result in loss of nerve fibers owing to neuropeptide depletion. Synovial tissue proliferation without concomitant growth of new nerve fibers may lead to an apparent partial denervation of synovium. Studies in patients suggest that free nerve endings containing substance P may modulate inflammation and the pain pathway in osteoarthritis. Afferent nerve fibers from the joint play an important role in the reflex inhibition of muscle contraction. Trophic factors generated by motor neurons, such as the neuropeptide calcitonin gene–related peptide, are important in maintaining muscle bulk and a functional neuromuscular junction. Decreases in motor neuron trophic support during articular inflammation probably contribute to muscle wasting.

Mechanisms of joint pain have been reviewed in detail. In a noninflamed joint, most sensory nerve fibers do not respond to movement within the normal range; these are referred to as silent nociceptors. In an acutely inflamed joint, however, these nerve fibers become sensitized by mediators, such as bradykinin, neuropeptide Y, and prostaglandins (peripheral sensitization), such that normal movements induce pain. Pain sensation is upregulated or downregulated further in the central nervous system, at the level of the spinal cord, and in the brain by central sensitization and gating of nociceptive input. Although the normal joint may respond predictably to painful stimuli, poor correlation has been noted between apparent joint disease and perceived pain in chronic arthritis. Pain associated with joint movements within the normal range is a characteristic symptom described by patients with chronically inflamed joints caused by rheumatoid arthritis. Chronically inflamed joints may not be painful at rest, however, unless acutely inflamed.

**Tendons**

Tendons are functional and anatomic bridges between muscle and bone. They focus the force of a large mass of muscle into a localized area on bone and, by splitting to form numerous insertions, may distribute the force of a single muscle to different bones. Tendons are formed of longitudinally arranged collagen fibrils embedded in an organized, hydrated proteoglycan matrix with blood vessels, lymphatics, and fibroblasts. Cross-links between adjacent collagen chains or molecules contribute to the tensile strength of the tendon. Tendon collagen fibrillogenesis is initiated during early development through a highly ordered process of alignment involving the actin cytoskeleton and cadherin-11. Many tendons, particularly those with a large range of motion, run through vascularized, discontinuous sheaths of collagen lined with mesenchymal cells resembling synovium. Gliding of tendons through their sheaths is enhanced by hyaluronic acid produced by the lining cells. Tendon movement is essential for the embryogenesis and maintenance of tendons and their sheaths. Degenerative changes appear in tendons, and fibrous adhesions are formed between tendons and sheaths when inflammation or surgical incision is followed by long periods of immobilization. At the myotendinous junction, recesses between muscle cell processes are filled with collagen fibrils, which blend into the tendon. At its other end, collagen fibers of the tendon typically blend into fibrocartilage or mineralize, and merge into bone through a fibrocartilaginous transition zone termed the enthesis, or insertion site. Tendon fibroblasts synthesize and secrete collagens, proteoglycans, and other matrix components, such as fibronectin and tenasin C, and MMPs and their inhibitors, which can contribute to the breakdown and repair of tendon components. Collagen fibrils in tendon are composed primarily of type I collagen with some type III collagen, but regional differences in the distribution of other matrix components have been noted. The compressed region contains the small proteoglycans—biglycan, decorin, fibromodulin, and lumican—and a large proteoglycan—versican. Major components in the tensile region of the tendon are decorin, microfibrillar type VI collagen, fibromodulin, and the proline and arginine-rich end leucine-rich repeat protein (PRELP). The presence of cartilage oligomeric matrix protein, aggrecan, biglycan, and collagen types II, IX, and XI is indicative of fibrocartilage. The collagen fiber orientation at the tendon-to-bone enthesis is important for maintaining microarchitecture by reducing stress concentrations and shielding the outward splay of insertion from the highest stresses. Understanding the structure has implications for tendon repair because motion between a tendon graft and a bone tunnel may impair early graft incorporation, leading to tunnel widening secondary to bone resorption.

Failure of the muscle-tendon apparatus is rare, but when it does occur, it is secondary to enormous, quickly generated...
inflammation may lead to the development of new bursae, but they are generally less well vascularized than synovium. Sparsely with mesenchymal cells similar to synovial cells, movement of its tendon. Bursae are closed sacs, lined one tissue over another, much as a tendon sheath facilitates movement of its tendon. Bursae are closed sacs, lined sparsely with mesenchymal cells similar to synovial cells, but they are generally less well vascularized than synovium. Most bursae differentiate concurrently with synovial joints during embryogenesis. During life, however, trauma or inflammation may lead to the development of new bursae, hypertrophy of previously existing ones, or communication between deep bursae and joints. In patients with rheumatoid arthritis, communications may exist between the subacromial bursae and the glenohumeral joint, between the gastrocnemius or semimembranous bursae and the knee joint, and between the iliopectos bursa and the hip joint. It is unusual, however, for subcutaneous bursae, such as the prepatellar bursa or olecranon bursa, to develop communication with the underlying joint.

Menisci
The meniscus, a fibrocartilaginous, wedge-shaped structure, is best developed in the knee, but also is found in the acromioclavicular and sternoclavicular joints, the ulnocarpal joint, and the temporomandibular joint. Until recently, menisci were thought to have little function and a quiescent metabolism with no capability of repair, although early observations indicated that removal of menisci from the knee may lead to premature arthritic changes in the joint. Evidence from an arthroscopic study of patients with anterior cruciate ligament insufficiency indicates that the pathology of the medial meniscus correlates with that of the medial femoral cartilage. The meniscus is now considered to be an integral component of the knee joint that has important functions in joint stability, load distribution, shock absorption, and lubrication.

The microanatomy of the meniscus is complex and age dependent. The characteristic shape of the lateral and medial menisci is achieved early in prenatal development. At that time, the menisci are cellular and highly vascularized; with maturation, vascularity decreases progressively from the central margin to the peripheral margin. After skeletal maturity, the peripheral 10% to 30% of the meniscus remains highly vascularized by a circumferential capillary plexus and is well innervated. Tears in this vascularized peripheral zone may undergo repair and remodeling.

The central portion of the mature meniscus is an avascular fibrocartilage, however, without nerves or lymphatics, consisting of cells surrounded by an abundant extracellular matrix of collagens, chondroitin sulfate, dermatan sulfate, and hyaluronic acid. Tears in this central zone heal poorly, if at all.

Collagen constitutes 60% to 70% of the dry weight of the meniscus and is mostly type I collagen, with lesser amounts of types III, V, and VI. A small quantity of cartilage-specific type II collagen is localized to the inner, avascular portion of the meniscus. Collagen fibers in the periphery are mostly circumferentially oriented, with radial fibers extending toward the central portion. Elastin content is around 0.6%, and proteoglycan content is around 2% to 3% dry weight. Aggrecan and decorin are the major proteoglycans in the adult meniscus. Decorin is the predominant proteoglycan synthesized in the meniscus from young individuals, whereas the relative proportion of aggrecan synthesis increases with age. Although the capacity of the meniscus to synthesize sulfated proteoglycans decreases after the teenage years, age-related increases in expression of decorin and aggrecan mRNA suggest that the resident cells are able to respond quickly to alterations in the biomechanical environment.

The meniscus was defined originally as a fibrocartilage, based on the rounded or oval shape of most of the cells and
the fibrous microscopic appearance of the extracellular matrix.\textsuperscript{207} Based on molecular and spatial criteria, three distinct populations of cells are recognized in the meniscus of the knee joint\textsuperscript{205}:  

1. The fibrochondrocyte is the most abundant cell in the middle and inner meniscus, synthesizing primarily type I collagen and relatively small amounts of type II and III collagen. It is round or oval in shape and has a pericellular filamentous matrix containing type VI collagen.  

2. The fibroblast-like cells lack a pericellular matrix and are located in the outer portion of the meniscus. They are distinguished by long, thin, branching cytoplasmic projections that stain for vimentin. They make contact with other cells in different regions via connexin 43–containing gap junctions. The presence of two centrosomes, one associated with a primary centrium, suggests a sensory, rather than motile, function that could enable the cells to respond to circumferential tensile loads, rather than compressive loads.\textsuperscript{208}  

3. The superficial zone cells have a characteristic fusiform shape with no cytoplasmic projections. Occasional staining of these cells in the uninjured meniscus with α-actin and their migration into surrounding wound sites suggest that they are specialized progenitor cells that may participate in a remodeling response in the meniscus and surrounding tissues.\textsuperscript{209, 210}  

**MATURE ARTICULAR CARTILAGE**  

Articular cartilage is a specialized connective tissue that covers the weight-bearing surfaces of diarthrodial joints.\textsuperscript{199, 211} The principal functions of cartilage layers covering bone ends are to permit low-friction, high-velocity movement between bones, to absorb transmitted forces associated with locomotion, and to contribute to joint stability. Lubrication by synovial fluid provides frictionless movement of the articulating cartilage surfaces. Chondrocytes (see Chapter 3) are the single cellular components of adult hyaline articular cartilage and are responsible for synthesizing and maintaining the highly specialized cartilage matrix macromolecules. The cartilage extracellular matrix is composed of an extensive network of collagen fibrils, which confers tensile strength, and an interlocking mesh of proteoglycans, which provides compressive stiffness through the ability to absorb and extrude water. Numerous other noncollagenous proteins also contribute to the unique properties of cartilage (Table 1-1). Histologically, the tissue appears to be fairly homogeneous and is clearly distinguished from calcified cartilage and underlying subchondral bone (Figure 1-7). The organization of articular cartilage and the structure-function relationships of cartilage matrix components are described in Chapter 3.  

**Subchondral Bone Interactions with Articular Cartilage**  

Subchondral bone is not a homogeneous tissue; it consists of a layer of compact cortical bone and an underlying system of cancellous bone organized into a trabecular network.\textsuperscript{212, 213} The subchondral bone is separated from the overlying articular cartilage by a thin zone of calcified cartilage. The so-called tidemark defines the transition zone between articular and calcified cartilage. This complex biocomposite of bone and calcified cartilage provides an optimal system for distributing loads that are transmitted from the weight-bearing surfaces lined by hyaline articular cartilage. Although the tidemark was originally believed to form a barrier to fluid flow, evidence suggests that biologically active molecules can transit this zone, providing a mechanism by which products produced by chondrocytes or bone cells can influence the activity of the other cell type.\textsuperscript{214}  

| Table 1-1 Extracellular Matrix Components of Articular Cartilage\textsuperscript{a} |
|--------------------------|--------------------------|
| **Collagens** |
| Type II |
| Type IX |
| Type XI |
| Type VI |
| Types XII, XIV |
| Type X (hypertrophic chondrocyte) |
| **Proteoglycans** |
| Aggrecan |
| Versican |
| Link protein |
| Biglycan (DS-PGII) |
| Decorin (DS-PGII) |
| Epiphycan (DS-PGIII) |
| Fibromodulin |
| Lumican |
| Proline/arginine-rich and leucine-rich repeat protein (PRELP) |
| Chondroadherin |
| Perlecan |
| Lubricin (SZP) |
| **Other Noncollagenous Proteins (Structural)** |
| Cartilage oligomeric matrix protein (COMP) or thrombospondin-5 |
| Thrombospondin-1 and thrombospondin-3 |
| Cartilage matrix protein (matrilin-1) and matrilin-3 |
| Fibronectin |
| Tenascin-C |
| Cartilage intermediate layer protein (CILP) |
| Fibrilin |
| Elastin |
| **Other Noncollagenous Proteins (Regulatory)** |
| Glycoprotein (gp)-39, YKL-40 |
| Matrix Gla protein (MGP) |
| Chondromodulin-I (SCGP) and chondromodulins-II |
| Cartilage-derived retinoic acid–sensitive protein (CD-RAP) |
| Growth factors |
| **Cell Membrane–Associated Proteins** |
| Integrins (α1β1, α2β1, α3β1, α5β1, α6β1, α10β1, αvβ3, αvβ5) |
| Anchorin CII (annexin V) |
| Cell determinant 44 (CD44) |
| Syndecan-1, -3, and -4 |
| Discoidin domain receptor 2 |

\textsuperscript{a}The collagens, proteoglycans, and other noncollagenous proteins in the cartilage matrix are synthesized by chondrocytes at different stages during development and growth of cartilage. In mature articular cartilage, proteoglycans and other noncollagen proteins are turned over slowly, whereas the collagen network is stable unless exposed to proteolytic cleavage. Proteins that are associated with chondrocyte cell membranes also are listed because they permit specific interactions with extracellular matrix proteins. The specific structure-function relationships are discussed in Chapter 3 and are described in Table 3-1.

*DS-PG, dermanan sulfate proteoglycan; SCGP, small cartilage–derived glycoprotein; SZP, superficial zone protein; YKL-40, 40kD chitinase 3-like glycoprotein.*
Under physiologic conditions, the composition and structural organization of subchondral bone and calcified cartilage are optimally adapted to transfer loads, but multiple conditions can lead to changes in the structural and functional properties of these tissues. For example, with advancing age, the zone of calcified cartilage may expand and advance into the deep zones of the overlying articular cartilage, producing thinning of the cartilage layer and alterations in load transfer and fibrillation and disruption of the cartilage surface.

Maintenance of the structural and functional integrity of articular cartilage and subchondral bone under physiologic loading provides evidence of the unique and intimate interaction of these tissues, but there remains controversy regarding the relationship between them in the pathogenesis of osteoarthritis. Radin and Rose proposed that the initiation of early alterations in articular cartilage is caused by an increase in subchondral bone stiffness that adversely affects the function of articular chondrocytes, leading to deterioration in the properties of the articular cartilage and susceptibility to mechanical disruption. Alternatively, it has been proposed that changes in subchondral bone stiffness may be secondary to cartilage deterioration.

The alterations in subchondral bone and cartilage that accompany the osteoarthritis process are not restricted to these tissues, but also affect the zone of calcified cartilage, where evidence reveals vascular invasion, advancement of the calcified cartilage, and duplication of the tidemark, which contributes further to a decrease in articular cartilage thickness. A more recent study showed that angiogenesis in the osteochondral junction is independent of synovial angiogenesis and synovitis, but is associated with cartilage changes and clinical disease activity. These structural alterations in the articular cartilage and periarticular bone may lead to modification of the contours of adjacent articulating surfaces, further contributing to the adverse biomechanical environment.

Analyses of periarticular bone in patients with osteoarthritis reveal that the structural and functional properties of subchondral cortical and trabecular bone are dependent on the stage of osteoarthritis progression. Several studies have investigated therapies that target bone remodeling to prevent these changes. Examples include the use of calcitonin, bisphosphonates, and estrogen. To date, no study has been performed in patients with osteoarthritis to investigate the efficacy of targeting receptor activator of nuclear factor κB (NFκB) ligand (RANKL), which mediates osteoclast differentiation and activity, and its receptor RANK, a member of the tumor necrosis factor receptor family. RANK is expressed in adult articular chondrocytes, but exogenous RANKL does not activate NFκB nor stimulate the production of collagenase or nitric oxide. Inhibition of RANKL expression does not block cartilage destruction in inflammatory models, although RANKL may have indirect effects on cartilage through its protective effect on bone.

**SYNOVIAL FLUID AND NUTRITION OF JOINT STRUCTURES**

The volume and composition of synovial fluid are determined by the properties of the synovium and its vasculature. Fluid in normal joints is present in small quantities (2.5 mL in the normal knee) sufficient to coat the synovial surface, but not to separate one surface from the other. Tendon
sheath fluid and synovial fluid are biochemically similar. Both are essential for the nutrition and lubrication of adjacent avascular structures, including tendon and articular cartilage, and for limiting adhesion formation and maintaining movement. Characterization and measurement of synovial fluid constituents have proved useful for the identification of locally generated regulatory factors, markers of cartilage turnover, and the metabolic status of the joint, and for assessment of the effects of therapy on cartilage homeostasis. Interpretation of such data requires, however, an understanding of the generation and clearance of synovial fluid and its various components.

**Generation and Clearance of Synovial Fluid**

Synovial fluid concentrations of a protein represent the net contributions of synovial blood flow, plasma concentration, microvascular permeability, and lymphatic removal and its production and consumption within the joint space. Synovial fluid is a mixture of a protein-rich ultrafiltrate of plasma and hyaluronan synthesized by synoviocytes. Generation of this ultrafiltrate depends on the difference between intracapillary and intra-articular hydrostatic pressures and between colloid osmotic pressures of capillary plasma and synovial tissue fluid. Fenestrations, small pores covered by a thin membrane, in the synovial capillaries and the macromolecular sieve of hyaluronic acid facilitate rapid exchange of small molecules, such as glucose and lactate, assisted—in the case of glucose—by an active transport system. Proteins are present in synovial fluid at concentrations inversely proportional to molecular size, with synovial fluid albumin concentrations being about 45% of those in plasma (Figure 1-8). Concentrations of electrolytes and small molecules are equivalent to those in plasma.

Synovial fluid is cleared through lymphatics in the synovium, assisted by joint movement. In contrast to ultrafiltration, lymphatic clearance of solutes is independent of molecular size. In addition, constituents of synovial fluid, such as regulatory peptides, may be degraded locally by enzymes, and low-molecular-weight metabolites may diffuse along concentration gradients into plasma. The kinetics of delivery and removal of a protein must be determined (e.g., using albumin as a reference solute) to assess the significance of its concentration in the joint.

Hyaluronic acid is synthesized by fibroblast-like synovial lining cells, and it appears in high concentrations in synovial fluid, at around 3 g/L, compared with a plasma concentration of 30 µg/L. Lubricin, a glycoprotein that assists articular lubrication, is another constituent of synovial fluid that is generated by the lining cells. It is now believed that hyaluronan functions in fluid-film lubrication, whereas lubricin is the true boundary lubricant in synovial fluid (see following). Because the volume of synovial fluid is determined by the amount of hyaluronan, water retention seems to be the major function of this large molecule.

Despite the absence of a basement membrane, synovial fluid does not mix freely with extracellular synovial tissue fluid. Hyaluronan may trap molecules within the synovial cavity by acting as a filtration screen on the surface of the synovial lining, resisting the movement of synovial fluid out from the joint space. Synovial fluid and its constituent proteins have a rapid turnover time (around 1 hour in normal knees), and equilibrium is not usually reached among all parts of the joint. Tissue fluid around fenestrated endothelium reflects plasma ultrafiltrate most closely, with a low content of hyaluronate compared with synovial fluid. Alternatively, locally generated or released peptides, such as endothelin and substance P, may attain much higher perivascular concentrations than those measured in synovial fluid. The turnover time for hyaluronan in the normal joint (13 hours) is an order of magnitude slower, however, than that for small solutes and proteins. Association with hyaluronan may result in trapping of solutes within synovial fluid.

In normal joints, intra-articular pressures are slightly subatmospheric at rest (0 to −5 mm Hg). During exercise, hydrostatic pressure in the normal joint may decrease further. Resting intra-articular pressures in rheumatoid joints are around 20 mm Hg, whereas during isometric exercise, they may increase to greater than 100 mm Hg—well above capillary perfusion pressure and, at times, above arterial pressure. Repeated mechanical stresses can interrupt synovial perfusion during joint movement, particularly in the presence of a synovial effusion.

**Synovial Fluid as an Indicator of Joint Function**

In the absence of a basement membrane separating synovium or cartilage from synovial fluid, measurements made on synovial fluid may reflect the activity of these structures. A
wide range of regulatory factors and products of synoviocyte metabolism and cartilage breakdown may be generated locally within the joint, resulting in marked differences between the composition of synovial fluid and that of plasma ultrafiltrate. Because there is little capacity for the selective concentration of solutes in synovial fluid, solutes present at higher concentrations than in plasma are probably synthesized locally. It is necessary to know the local clearance rate, however, to determine whether solutes present in synovial fluid at lower concentrations than in plasma are generated locally.236 Although microvascular permeability to protein in highly inflamed rheumatoid joints is more than twice that in osteoarthritic joints, synovial fluid protein concentrations vary little between the two joint diseases241 because enhanced entry of proteins through the microvasculature is largely offset by the increased lymphatic clearance.242 Because clearance rates from synovial fluid may be slower than those from plasma, however, synovial fluid levels of drugs or urate may remain elevated after plasma levels have declined.234

Comparisons of synovial fluid constituents between disease groups are often limited by the sparseness of data on normal synovial fluid as a result of difficulties in its collection. Extrapolation from synovial fluid concentrations to local synthetic rates is complicated further by variations in clearance rates and in synovial fluid volume. Plasma proteins are less effectively filtered in inflamed synovium, perhaps because of the increased size of endothelial cell fenestrations, or because interstitial hyaluronate-protein complexes are fragmented by enzymes associated with the inflammatory process.235 Concentrations of proteins, such as α2-macroglobulin (the principal proteinase inhibitor of plasma), fibrinogen, and IgM, are elevated in inflammatory synovial fluids (see Figure 1-8), as are associated protein-bound cations. Membrane peptidases may limit the diffusion of regulatory peptides from their sites of release into synovial fluid. In inflammatory arthritis, fibrin deposits may retard flow between tissue and liquid phases. Cautious interpretation of synovial fluid analysis has important implications in understanding how to use data on biomarkers of cartilage damage and repair in rheumatoid arthritis and osteoarthritis (see Chapter 53).

Recently, Gobeje and co-workers243 have utilized high-throughput mass spectroscopy-based proteomic analysis to define the protein expression profiles of high-abundance synovial fluid proteins in healthy subjects and patients with early and late osteoarthritis. They identified 18 proteins that were significantly differentially expressed between osteoarthritic and control groups. Although all of the differentially expressed proteins are present in the blood and could therefore enter the joint through alterations in vascular permeability associated with the disease state, these molecules are also products of synovial cells and chondrocytes, suggesting that they could be locally produced within the joint. Proteins associated with oxidative damage and activation of mitogen-activated protein kinases were among the high-abundance molecules in osteoarthritic synovial fluids. Members of the proinflammatory complement cascade were also identified in the synovial fluid. Of interest, these molecules have been implicated in the pathophysiology of both osteoarthritis and rheumatoid arthritis.

**Lubrication and Nutrition of the Articular Cartilage**

**Lubrication**

Synovial fluid serves as a lubricant for articular cartilage and as a source of nutrition for the chondrocytes within. Lubrication is essential for protecting cartilage and other joint structures from friction and shear stresses associated with movement under loading. Two basic categories of joint lubrication are known. In fluid-film lubrication, cartilage surfaces are separated by an incompressible fluid film; hyaluronan functions as the lubricant. In boundary lubrication, specialized molecules attached to the cartilage surface permit surface-to-surface contact, while decreasing the coefficient of friction.

During loading, a noncompressible fluid film trapped between opposing cartilage surfaces prevents the surfaces from touching. Irregularities in the cartilage surface and its deformation during compression may augment this trapping of fluid. This stable film is approximately 0.1 μm thick in the normal human hip joint, but it can be much thinner in the presence of inflammatory synovial fluids or with increased cartilage porosity.244,245

Lubricin is the major boundary lubricant in the human joint.246 It is a glycoprotein, also called superficial zone protein or proteoglycan 4, that is synthesized by synovial cells and chondrocytes.247,248 Recent studies have demonstrated that lubricin is also produced by meniscal and tendon cells.251,252 It has a molecular weight of 225,000, a length of 200 nm, and a diameter of 1 to 2 nm.251 Dipalmitoyl phosphatidylcholine, which constitutes 45% of the lipid in normal synovial fluid, acts together with lubricin as a boundary lubricant.254 More recent work indicates that lubricin functions as a phospholipid carrier via a mechanism that is common to all tissues.255,256 In cartilage, lipid composes 1% to 2% of the dry weight,257 and experimental treatment of cartilage surfaces with fat solvents impairs lubrication qualities.258

**Nutrition**

As observed by Hunter in 1743,259 normal adult articular cartilage contains no blood vessels. Vascularization of cartilage would be expected to alter its mechanical properties. Blood flow would be repeatedly occluded during weight bearing and exercise, with reactive oxygen species generated during reperfusion, resulting in repeated damage to cartilage matrix and chondrocytes. Chondrocytes synthesize specific inhibitors of angiogenesis that maintain articular cartilage as an avascular tissue.260,262 As a result of the lack of adjacent blood vessels, the chondrocyte normally lives in an hypoxic and acidic environment, with extracellular fluid pH values around 7.1 to 7.2,263 and it uses anaerobic glycolysis for energy production.263 High lactate levels in normal synovial fluid, compared with paired plasma measurements, partially reflect this anaerobic metabolism.265 There are two sources of nutrients for articular cartilage: synovial fluid and subchondral blood vessels.

The synovial fluid and, indirectly, the synovial lining, through which synovial fluid is generated, are the major sources of nutrients for articular cartilage. Nutrients may
enter cartilage from synovial fluid by diffusion or by mass transport of fluid during compression-relaxation cycles.\textsuperscript{256} Molecules as large as hemoglobin (65 kD) can diffuse through normal articular cartilage,\textsuperscript{257} and the solutes needed for cellular metabolism are much smaller. Diffusion of uncharged small solutes, such as glucose, is not impaired in matrices containing large quantities of glycosaminoglycans, and the diffusivity of small molecules through hyaluronate is enhanced.\textsuperscript{268,269}

Intermittent compression may serve as a pump mechanism for solute exchange in cartilage. The concept has arisen from observations that joint immobilization or dislocation leads to degenerative changes. In contrast, exercise increases solute penetration into cartilage in experimental systems.\textsuperscript{261} During weight bearing, fluid escapes from the load-bearing region by flow to other cartilage sites. When the load is removed, cartilage re-expands and draws back fluid, exchanging nutrients with waste materials.\textsuperscript{270}

In a growing child, the deeper layers of cartilage are vascularized, such that blood vessels penetrate between columns of chondrocytes in the growth plate. It is likely that nutrients diffuse from these tiny end capillaries through the matrix to chondrocytes. Diffusion from subchondral blood vessels is not considered a major route for the nutrition of normal adult articular cartilage because of the barrier provided by its densely calcified lower layer. Nonetheless, partial defects may normally exist in this barrier,\textsuperscript{271} and in arthritis, neovascularization of the deeper layers of articular cartilage may contribute to cartilage nutrition and to entry of inflammatory cells and cytokines.\textsuperscript{221,222} In aging and osteoarthritis, tidemark "duplication" may indicate communication between bone and cartilage.\textsuperscript{218,219} Experimental studies have indicated that cartilage lesions of chondromalacia may develop if the subchondral blood supply of the patella is compromised.\textsuperscript{274}

## SUMMARY AND CONCLUSION

Normal human synovial joints are complex structures that comprise interacting connective tissue elements that permit constrained and low-friction movement of adjacent bones. The development of synovial joints in the embryo is a highly ordered process involving complex cell-cell and cell-matrix interactions, leading to the formation of cartilage anlagen and interzone and joint cavitation. Understanding of the cellular interactions and molecular factors involved in cartilage morphogenesis and limb development has provided clues to understanding the functions of the synovium, articular cartilage, and associated structures in the mature joint.

The synovial joint is uniquely adapted to responding to environmental and mechanical demands. The synovial lining is composed of two to three cell layers, and no basement membrane separates the lining cells from underlying connective tissue. The synovium produces synovial fluid, which provides nutrition and lubrication to the avascular articular cartilage. Normal articular cartilage contains a single cell type, the articular chondrocyte, which is responsible for maintaining the integrity of the extracellular cartilage matrix. This matrix consists of a complex network of collagens, proteoglycans, and other noncollagenous proteins, which provide tensile strength and compressive resistance. Proper distribution and relative composition of these proteins are required for the function of cartilage in protecting the subchondral bone from adverse environmental influences.

Maintenance of the unique composition and organization of each joint tissue is crucial for normal joint function, which is compromised in response to inflammation, biomechanical injury, and aging. Knowledge of normal structure-function relationships within joint tissues is essential for understanding the pathogenesis and consequences of joint disease.

## Selected References


65. Liu Z, Xu J, Colvin JS, Ornitz DM: Coordination of chondrogenesis requires the homeobox gene L-Sox5, Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene, EMBO J 17:5718–5733, 1998.


CHAPTER 1 | BIOLOGY OF THE NORMAL JOINT 19.e5

19.1 The joints and synovial fluid

19.1.1 The synovial lining

19.1.2 Synovial fluid and joint lubrication

19.1.3 Synovial fluid mechanics

19.1.4 Synovial fluid composition and turnover

19.1.5 Synovial fluid clearance

19.1.6 Synovial fluid nutrition and metabolism

19.1.7 Synovial fluid in joint metabolism

19.1.8 Synovial fluid in joint function

19.1.9 Synovial fluid in joint health and disease

19.1.10 Synovial fluid and joint comfort

19.1.11 Synovial fluid and joint resilience

19.1.12 Synovial fluid and joint repair

19.1.13 Synovial fluid and joint regeneration

19.1.14 Synovial fluid and joint senescence

19.1.15 Synovial fluid and joint longevity

19.1.16 Synovial fluid and joint evolution

19.1.17 Synovial fluid and joint innovation

19.1.18 Synovial fluid and joint revolution

19.1.19 Synovial fluid and joint revolution

19.1.20 Synovial fluid and joint revolution

19.1.21 Synovial fluid and joint revolution

19.1.22 Synovial fluid and joint revolution

19.1.23 Synovial fluid and joint revolution

19.1.24 Synovial fluid and joint revolution

19.1.25 Synovial fluid and joint revolution

19.1.26 Synovial fluid and joint revolution

19.1.27 Synovial fluid and joint revolution

19.1.28 Synovial fluid and joint revolution

19.1.29 Synovial fluid and joint revolution

19.1.30 Synovial fluid and joint revolution

19.1.31 Synovial fluid and joint revolution

19.1.32 Synovial fluid and joint revolution

19.1.33 Synovial fluid and joint revolution

19.1.34 Synovial fluid and joint revolution

19.1.35 Synovial fluid and joint revolution

19.1.36 Synovial fluid and joint revolution

19.1.37 Synovial fluid and joint revolution

19.1.38 Synovial fluid and joint revolution

19.1.39 Synovial fluid and joint revolution

19.1.40 Synovial fluid and joint revolution

19.1.41 Synovial fluid and joint revolution

19.1.42 Synovial fluid and joint revolution

19.1.43 Synovial fluid and joint revolution

19.1.44 Synovial fluid and joint revolution

19.1.45 Synovial fluid and joint revolution

19.1.46 Synovial fluid and joint revolution

19.1.47 Synovial fluid and joint revolution

19.1.48 Synovial fluid and joint revolution

19.1.49 Synovial fluid and joint revolution

19.1.50 Synovial fluid and joint revolution

19.1.51 Synovial fluid and joint revolution

19.1.52 Synovial fluid and joint revolution

19.1.53 Synovial fluid and joint revolution

19.1.54 Synovial fluid and joint revolution

19.1.55 Synovial fluid and joint revolution

19.1.56 Synovial fluid and joint revolution

19.1.57 Synovial fluid and joint revolution

19.1.58 Synovial fluid and joint revolution

19.1.59 Synovial fluid and joint revolution

19.1.60 Synovial fluid and joint revolution

19.1.61 Synovial fluid and joint revolution

19.1.62 Synovial fluid and joint revolution

19.1.63 Synovial fluid and joint revolution

19.1.64 Synovial fluid and joint revolution

19.1.65 Synovial fluid and joint revolution

19.1.66 Synovial fluid and joint revolution

19.1.67 Synovial fluid and joint revolution

19.1.68 Synovial fluid and joint revolution

19.1.69 Synovial fluid and joint revolution

19.1.70 Synovial fluid and joint revolution

19.1.71 Synovial fluid and joint revolution

19.1.72 Synovial fluid and joint revolution

19.1.73 Synovial fluid and joint revolution

19.1.74 Synovial fluid and joint revolution

19.1.75 Synovial fluid and joint revolution

19.1.76 Synovial fluid and joint revolution

19.1.77 Synovial fluid and joint revolution

19.1.78 Synovial fluid and joint revolution

19.1.79 Synovial fluid and joint revolution

19.1.80 Synovial fluid and joint revolution

19.1.81 Synovial fluid and joint revolution

19.1.82 Synovial fluid and joint revolution

19.1.83 Synovial fluid and joint revolution

19.1.84 Synovial fluid and joint revolution

19.1.85 Synovial fluid and joint revolution

19.1.86 Synovial fluid and joint revolution

19.1.87 Synovial fluid and joint revolution

19.1.88 Synovial fluid and joint revolution

19.1.89 Synovial fluid and joint revolution

19.1.90 Synovial fluid and joint revolution

19.1.91 Synovial fluid and joint revolution

19.1.92 Synovial fluid and joint revolution

19.1.93 Synovial fluid and joint revolution

19.1.94 Synovial fluid and joint revolution

19.1.95 Synovial fluid and joint revolution

19.1.96 Synovial fluid and joint revolution
Synovium

BARRY BRESNIHAN • ADRIENNE M. FLANAGAN • GARY S. FIRESTEIN

KEY POINTS

The synovium provides nutrients to cartilage and produces lubricants for the joint.

The intimal lining of the synovium includes macrophage-like and fibroblast-like synoviocytes.

The sublining in normal synovium contains scattered immune cells, fibroblasts, blood vessels, and fat cells.

Fibroblast-like synoviocytes in the intimal lining produce specialized enzymes that synthesize lubricants such as hyaluronic acid.

STRUCTURE

The synovium is a membranous structure that extends from the margins of articular cartilage and lines the capsule of diarthrodial joints, including the temporomandibular joint and the facet joints of vertebral bodies (Figure 2-1). The healthy synovium covers intra-articular tendons and ligaments, as well as fat pads, but not articular cartilage or meniscal tissue. Synovial also ensheathes tendons where they pass beneath ligamentous bands and bursas that cover areas of stress such as the patella and the olecranon. The synovial membrane is divided into general regions—the intima, or synovial lining, and the subintima, otherwise referred to as the sublining. The intima represents the interface between the cavity containing synovial fluid and the subintimal layer. No well-formed basement membrane separates the intima from the subintima. It is not a true lining, in contrast to the pleura or pericardium, because it lacks tight junctions, epithelial cells, and a well-formed basement membrane. The subintima is composed of fibrovascular connective tissue and merges with the densely collagenous fibrous joint capsule.

Synovial Lining Cells

The synovial intimal layer is composed of synovial lining cells (SLCs), which are arrayed on the luminal aspect of the joint cavity. SLCs, termed synoviocytes, are one to three cells deep, depending on the anatomic location, and extend 20 to 40 µm beneath the lining layer surface. The major and minor axes of SLCs measure 8 to 12 µm and 6 to 8 µm, respectively. The SLCs are not homogeneous and are conventionally divided into two major populations, namely, type A (macrophage-like) synoviocytes and type B (fibroblast-like) synoviocytes.

Ultrastructure of Synovial Lining Cells

Transmission electron microscopic analysis shows that the intimal cells form a discontinuous layer, so that the subintimal matrix can directly contact the synovial fluid (Figure 2-2). The existence of two distinct cell types—type A and type B SLCs—was originally described by Barland and associates, and several lines of evidence, including animal models, detailed ultrastructural studies, and immunohistochemical analyses, indicate that these cells represent macrophages (type A SLCs) and fibroblasts (type B SLCs). Studies of SLC populations in a variety of species, including humans, have found that macrophages make up anywhere from 20% and fibroblast-like cells approximately 80% of the lining cell. The existence of the two cell types has been substantiated by similar findings in a wide variety of species, including hamsters, cats, dogs, guinea pigs, rabbits, mice, rats, and horses. Distinguishing different cell populations that form the synovial lining requires immunohistochemistry or transmission light microscopy. At an ultrastructural level, type A cells are characterized by a conspicuous Golgi apparatus, large vacuoles, and small vesicles, and they contain little rough endoplasmic reticulum, giving them a macrophage-like phenotype (Figure 2-3A and B). The plasma membrane of type A cells possesses numerous fine extensions, termed filopodia, which are characteristic of macrophages. Type A cells occasionally cluster at the tips of the synovial villi; this uneven distribution explains at least in part early reports that suggested type A cells were the predominant intimal cell type. However, the distribution is highly variable and can differ depending on the joint evaluated or even within an individual joint.

Type B SLCs have prominent cytoplasmic extensions that extend onto the surface of the synovial lining (Figure 2-3C and D). Frequent invaginations are seen along the plasma membrane; a large indented nucleus relative to the area of the surrounding cytoplasm is also a feature. Type B cells have abundant rough endoplasmic reticulum widely distributed in the cytoplasm, and the Golgi apparatus, vacuoles, and vesicles are generally inconspicuous, although some cells have small numbers of prominent vacuoles at their apical aspect. Type B SLCs are known to contain longitudinal bundles of different-sized filaments, supporting their classification as fibroblasts. Desmosomes and gap-like junctions have been described in rat, mouse, and rabbit synovium, but the existence of these structures in human SLCs has never been documented. Although occasional reports describe an intermediate synoviocyte phenotype, it is likely that these cells are functionally conventional type A or B cells.
Immunohistochemical Profile of Synovial Cells

Synovial Macrophages. Synovial macrophages and fibroblasts express lineage-specific molecules, which can be detected by immunohistochemistry. Synovial macrophages express common hematopoietic antigen CD45 (Figure 2-4A); monocyte/macrophage receptors CD163 and CD97; and lysosomal enzymes CD68 (Figure 2-4B), neuron-specific esterase, and cathepsins B, L, and D. Cells expressing CD14, a molecule that acts as a co-receptor for the detection of bacterial lipopolysaccharide and expressed by circulating monocytes and monocytes newly recruited to tissue, are rarely seen in the healthy intimal layer, but small numbers are found close to venules in the subintima.18-24

The Fcγ receptor, FcγRIII (CD16), expressed by Kupffer cells of the liver and type II alveolar macrophages of the lung, is expressed on a subpopulation of synovial macrophages.25-27 The synovial macrophage population also expresses the major histocompatibility complex (MHC) class II molecule, which plays an important role in the immune response. More recently, the macrophages, which are responsible for the removal of debris, blood, and particulate material from the joint cavity and possess antigen processing properties, have been found to express the complement-related protein, Z39lg, a cell surface receptor and immunoglobulin superfamily member involved in the induction of human leukocyte antigen (HLA)-DR and in implicated phagocytosis and antigen-mediated immune responses.28-30

Expression of the β2 integrin chains—CD18, CD11a, CD11b, and CD11c—varies; CD11a and CD11c may be absent or weakly expressed on a few lining cells.31,32 Osteoclasts, which are tartrate resistant and acid phosphatase positive and express the α,β3 vitronectin and calcitonin receptors, do not appear in the normal synovium.

Synovial Intimal Fibroblasts. Synovial intimal and subintimal fibroblasts are indistinguishable by light microscopy. They generally are considered to be closely related in terms of cell lineage, but because of their different microenvironments, they do not always share the same phenotype. They possess prominent synthetic capacity and produce the essential joint lubricants hyaluronic acid (HA) and lubricin.33 Intimal fibroblasts express uridine diphosphoglucose dehydrogenase (UDPGD), an enzyme involved in HA synthesis that is a relatively specific marker for this cell type. UDPGD converts UDP-glucose to UDP-glucuronate, one of the two substrates required by HA synthase for assembly of the HA polymer.34 CD44, the nonintegrin receptor for HA, is expressed by all SLCs.12,13,35,36

Synovial fibroblasts also synthesize normal matrix components, including fibronectin, laminin, collagens, proteoglycans, lubricin, and other identified and unidentified proteins. They have the capacity to produce large quantities of metalloproteinases, metalloproteinase inhibitors, prostaglandins, and cytokines. This capacity must provide essential biologic advantages, but the complex physiologic mechanisms relevant to normal function are incompletely delineated. Expression of selected adhesion molecules on synovial fibroblasts probably facilitates the trafficking of some cell populations, such as neutrophils, into the synovial fluid, and retention of others, such as mononuclear leukocytes, in the synovial tissue. Expression of metalloproteinases, cytokines, adhesion molecules, and other cell surface molecules is strikingly increased in inflammatory states.

Specialized intimal fibroblasts express many other molecules that might be expressed by the intimal macrophage population or by most subintimal fibroblasts, including decay-accelerating factor (CD55); vascular cell adhesion molecule-133,37,40; and cadherin-11.41,42 Pgp95, a neuronal marker, might be specific for type B synoviocytes in some species.43 Decay-accelerating factor, also expressed on many other cells (most notably erythrocytes) as well as bone marrow cells, interacts with CD97, a glycoprotein that is present on the surface of activated leukocytes, including intimal macrophages, and is thought to be involved in signaling processes early after leukocyte activation.44,45 In contrast, FcγRIII is expressed by macrophages only when they are in close contact with decay-accelerating factor–positive fibroblasts or decay-accelerating factor–coated fibrillin-1 microfibrils in the extracellular matrix.46

Cadherins are a class of tissue-restricted transmembrane proteins that play important roles in homophilic intercellular adhesion and are involved in maintaining the integrity of tissue architecture. Cadherin-11, which was cloned from
rheumatoid arthritis synovial tissue, is expressed in normal synovial intimal fibroblasts, but not in intimal macrophages. The finding that fibroblasts transfected with cadherin-11 are induced to form a lining-like structure in vitro implicates this molecule in the architectural organization of the synovial lining. This suggestion is supported by the observation that cadherin-deficient mice have a hypoplastic synovial lining and are resistant to inflammatory arthritis. When fibroblasts expressing cadherin-11 are embedded in laminin microparticles, they migrate to the surface and form an intimal lining–like structure. If macrophage lineage cells are included in the culture, they can co-localize with fibroblasts on the surface. These data suggest that the organization of the synovial lining, including the distribution of type A and B cells, is orchestrated by the fibroblast-like synoviocytes.

β1 and β3 integrins are present on all SLCs, forming receptors for laminin (CD49f and CD49b), collagen types I and IV (CD49b), vitronectin (CD51), CD54 (a member of the immunoglobulin superfamily), and fibronectin (CD49d and CD49e). CD31 (platelet–endothelial cell adhesion molecule), a member of the immunoglobulin superfamily expressed on endothelial cells, platelets, and monocytes, is weakly expressed on SLCs.

**Turnover of Synovial Lining Cells**

Proliferation of SLCs in humans is low, as is seen when normal human synovial explants, exposed to a pulse of 3H thymidine, cause SLCs to have a labeling index of approximately 0.05% to 0.3%; this bears a striking contrast to labeling indices of approximately 50% for bowel crypt
In addition to immunohistochemistry, several lines of evidence have added weight to the concept that type A SLCs are recruited from the bone marrow:

1. The op/op mouse, a spontaneously occurring mutant that fails to produce macrophage colony-stimulating factor because of a missense mutation in the csf-1 gene, has low numbers of circulating and resident macrophage colony-stimulating factor–dependent macrophages, including those in the synovium.

2. Type A cells in rat synovium do not populate the joint until after the development of synovial blood vessels.

3. Others have reported that type A SLCs were conspicuous around vessels in the synovium in neonatal mice.

4. When synovial explants are placed in culture, the reduction in type A SLCs is explained in part by their migration into the culture medium—an observation that reflects the process of migration of macrophages into the synovial fluid in vivo.

5. Macrophages are found around venules in disease states and constitute 80% of the intimal cells in inflammatory conditions such as rheumatoid arthritis. Type B intimal cells represent a resident fibroblast population in the synovial lining, but little is known about the cells from which they derive, and about how their recruitment is regulated. The existence of mesenchymal stem cells in the synovium suggests that these might differentiate into the synovial lining fibroblast. To date, a specific transcription factor directing mesenchymal stem cell differentiation into the synovial fibroblast, similar to factors required for commitment by this multipotential population into bone (cbfa-1), cartilage (Sox 9), and fat (peroxisome proliferator-activated receptor γ [PPARγ]), has not been identified.

**Origin of Synovial Lining Cells**

There is little doubt that the type A SLC population is bone marrow derived and represents cells of the mononuclear phagocyte system. Studies in the Beige (bg) mouse, which harbors a homozygous mutation that confers the presence of giant lysosomes in macrophages, have confirmed the bone marrow origin of these cells. Normal mice, bone marrow depleted through irradiation, were rescued with bone marrow cells obtained from the bg mouse. Electron microscopic analysis of the synovium from recipient animals revealed that type A SLCs contained the giant lysosomes of the donor bg mouse, and that these structures were never identified in type B cells. These findings provide powerful evidence that type A SLCs represent macrophages, that they are recruited from the bone marrow, and that they are a distinct lineage from type B SLCs.

In addition to immunohistochemistry, several lines of evidence have added weight to the concept that type A SLCs are recruited from the bone marrow:

1. The op/op mouse, a spontaneously occurring mutant that fails to produce macrophage colony-stimulating factor because of a missense mutation in the csf-1 gene, has low numbers of circulating and resident macrophage colony-stimulating factor–dependent macrophages, including those in the synovium.

2. Type A cells in rat synovium do not populate the joint until after the development of synovial blood vessels.

3. Others have reported that type A SLCs were conspicuous around vessels in the synovium in neonatal mice.

4. When synovial explants are placed in culture, the reduction in type A SLCs is explained in part by their migration into the culture medium—an observation that reflects the process of migration of macrophages into the synovial fluid in vivo.

5. Macrophages are found around venules in disease states and constitute 80% of the intimal cells in inflammatory conditions such as rheumatoid arthritis. Type B intimal cells represent a resident fibroblast population in the synovial lining, but little is known about the cells from which they derive, and about how their recruitment is regulated. The existence of mesenchymal stem cells in the synovium suggests that these might differentiate into the synovial lining fibroblast. To date, a specific transcription factor directing mesenchymal stem cell differentiation into the synovial fibroblast, similar to factors required for commitment by this multipotential population into bone (cbfa-1), cartilage (Sox 9), and fat (peroxisome proliferator-activated receptor γ [PPARγ]), has not been identified.

**Subintimal Layer**

SLCs are not separated from the underlying subintima by a well-formed basement membrane composed of the typical trilaminar structure seen beneath epithelial mucosa.
Nevertheless, most components of basement membrane are present in the extracellular matrix surrounding SLCs. These components include tenascin X, perlecan (a heparan sulfate proteoglycan), collagen type IV, laminin, and fibrillin-1.\(^60,61\) Of note is the absence of laminin-5 and integrin \(\alpha_3\beta_3\gamma_2\), which are components of epithelial hemidesmosomes.\(^62\)

The subintima is composed of loose connective tissue of variable thickness and variable proportions of fibrous/collagenous and adipose tissue, depending on the anatomic site. Under normal healthy conditions, inflammatory cells are virtually absent from the subintima, apart from a sprinkling of macrophages and scattered mast cells.\(^63\) Human synovial tissue is a rich source of mesenchymal stem cells, and although it is unknown which compartment contains this cell population, some cells have the ability to self-renew and differentiate into bone, cartilage, and fat in vitro—a phenomenon that reflects the ability of the cell to regenerate in vivo.\(^64-66\)

Three categories of subintima are well defined: areolar, fibrous, and fatty/adipose types. Under the light microscope, areolar-type subintima, the most commonly studied, generally is found in larger joints in which there is free movement (Figure 2-5A). It is composed of fronds with a cellular intimal lining and loose connective tissue in the subintima, with little in the way of dense collagen fibers, and a rich vasculature. The fibrous subintima is composed of scant dense fibrous, poorly vascularized connective tissue and has an attenuated layer of SLCs (Figure 2-5B). The adipose type contains abundant mature fat cells and has a single layer of SLCs. This is seen more commonly with aging and in intra-articular fat pads (Figure 2-5C).

The subintima contains collagen types I, III, V, and VI; glycosaminoglycans; proteoglycans; and extracellular matrices including tenascin and laminins. Integrin receptors for collagens, laminin, and vitronectin are absent or at best weakly expressed by subintimal cells. In contrast, receptors for fibronectin (CD49d and CD49e) are detected, and CD44, the HA receptor, is strongly expressed in most subintimal cells. \(\beta_2\) integrins are largely limited to perivascular areas, particularly in the subintimal zone, as is CD54.\(^67\)

### Subintimal Vasculature

The vascular supply to the synovium is provided by many small vessels and is shared in part by the joint capsule, epiphyseal bone, and other perisynovial structures. Arteriovenous anastomoses communicate freely with the vascular supply to the peristeum and periarticular bone. As large synovial arteries enter the deep layers of the synovium near the capsule, they branch to form microvascular units in the more superficial subsynovial layers. Precapillary arterioles probably play a major role in controlling circulation to the lining layer. The surface area of the synovial capillary bed is large, and because it runs only a few cell layers deep to the surface, it has a role in trans-synovial exchange of molecules. The intimal lining, however, is devoid of blood vessels.

Numerous physical factors influence synovial blood flow. Heat promotes blood flow through synovial capillaries. Exercise enhances synovial blood flow to normal joints but may reduce the clearance rate of small molecules from the joint space. Experiments have shown a substantial vascular reserve capacity in normal articulations. Immobilization reduces synovial blood flow, and pressure on the synovial membrane can act to tamponade the synovial blood supply.

Vascular endothelial lining cells express CD34 and CD31 (Figure 2-6A). They also express receptors for the
major components of basement membrane, including laminin and collagen IV, and the integrin receptors CD49a (laminin and collagen receptors), CD49d (fibronectin receptor), CD41, CD51 (vitronectin receptor), and CD61 (the β3 integrin subunit). Endothelial cells express CD44, the HA receptor, and CD62, P-selectin, which acts as a receptor that supports binding of leukocytes to activated platelets and endothelium. They are only weakly positive in uninflamed synovium, however, for expression of CD54, intercellular adhesion molecule-1, a receptor for β2
integrins expressed by many leukocytes. The endothelial cells of capillaries in the superficial zone of the subintima are strongly positive for HLA-DR expression by immunohistochemistry, whereas cells in the larger vessels in the deep aspect of the membrane are negative.\textsuperscript{32,34}

**Subintimal Lymphatics**

Detailed analysis of the number and distribution of lymphatic vessels has been made possible by the use of the antibody to the lymphatic vessel endothelial HA receptor (LYVE-1) (Figure 2-6B).\textsuperscript{68} This antibody is highly specific for lymphatic endothelial cells in lymphatic vessels and lymph node sinuses and does not react with endothelial cells of capillaries and other blood vessels that express CD34 and factor VIII–related antigen. Expression of LYVE-1 in lymphatic endothelial cells has been used as a marker to show that lymphatic vessels are less common in the fibrous synovium compared with areolar and adipose variants of human subsynovial tissue. Detection of this molecule reveals that lymphatics are present in the superficial, intermediate, and deeper layers of synovial membrane in synovium from normal individuals or patients with osteoarthritis and rheumatoid arthritis joints, although the number in the superficial subintimal layer is low in normal synovium. Little difference in the distribution and number is noted between normal and osteoarthritis synovium, which is characterized by lack of villous hypertrophy. Lymphatic channels are plentiful, however, in the subintimal layer in the presence of villous edema hypertrophy and chronic inflammation.

**Subintimal Nerve Supply**

The synovium has a rich network of sympathetic and sensory nerves. The former, which are myelinated and detected with the antibody against S-100 protein, terminate close to blood vessels, where they regulate vascular tone (Figure 2-6C through E). Sensory nerves respond to proprioception and pain via large myelinated nerve fibers and via small (≤5 μm) unmyelinated or myelinated fibers with unmylelinated free nerve ends (nociceptors). The latter are immunoreactive in the synovium for neuropeptides, including substance P, calcitonin gene–related peptide, and vasoactive intestinal peptides.\textsuperscript{69,70}

**FUNCTION**

Known synthetic and protective functions of individual synovial cell populations are multiple and complex. The composite synovial structure, which includes cell populations and their products, vasculature, nerves, and the intercellular matrix, possesses several specialized functions that are essential for normal joint movement, synovial fluid formation, chondrocyte nutrition, and cartilage protection at multiple anatomic locations. These functions must be preserved over a lifetime to maintain maximal mobility and independence. Absence of essential constituents of synovial fluid, or inadequate cartilage protection, results in early articular malfunction, which may progress to local or generalized joint failure.

**Joint Movement**

Four characteristics of the synovium are essential for joint movement: deformability, porosity, nonadherence, and cartilage lubrication. In health, the synovium is a highly deformable structure that facilitates movement between other adjacent, nondeformable structures within the joint. This unique facility of the synovium to enable movement between, rather than within, tissues has been emphasized\textsuperscript{71} and can be attributed to the presence of a free surface that allows synovial tissue to remain separated from adjacent tissues. The ensuing space is maintained by the presence of synovial fluid.

**Deformability**

The deformability of normal synovium is considerable because it must accommodate the extreme positional range available to the joint and its adjacent tendons, ligaments, and capsule. When a finger is flexed, the palmar synovium of each interphalangeal joint contracts, while the dorsal synovium expands, and as the finger extends, the reverse occurs. This normal contraction and expansion of synovium seems to involve a folding and unfolding component and an elastic stretching and relaxation of the tissue. It is essential that during repeated rapid movement, synovial lining does not become pinched between cartilage surfaces and can successfully retain its integrity and the integrity of synovial blood vessels and lymphatics. Deformability also limits the extent of synovial ischemia-reperfusion injury during joint motion by maintaining a relatively low intra-articular pressure.

**Porosity**

The synovial microvasculature and the intimal lining must be porous to permit robust diffusion of nutrients to cartilage. The structure of the intimal lining is ideal for this requirement because of the relatively disorganized basement membrane and lack of tight junctions. Plasma components freely diffuse into the intra-articular space, and most plasma components, including proteins, are present in synovial fluid at about one-third to one-half the plasma concentration.

**Nonadherence**

The third important characteristic of the synovium that facilitates joint movement is its nonadherence to opposing surfaces. Intimal cells on the synovial surface adhere to underlying cells and matrix but do not adhere to opposing synovial and cartilage surfaces. The mechanism that preserves this phenomenon of nonadherence is unknown and might involve the arrangement of cell surface and tissue matrix molecules, such as collagen, fibronectin, and HA. Alternatively, nonadherence may result in part from regular movements of the normal synovial lining.

**Lubrication**

The fourth characteristic of synovium that is essential for joint motion is an efficient lubrication mechanism to
facilitate movement of cartilage on cartilage. The mechanisms of joint lubrication are complex and are an integral component of synovial physiology. In an articulating joint, cartilage is subjected to numerous compressive and frictional forces every day. Friction and wear can never be eliminated from a functioning joint. Adult chondrocytes do not normally divide in vivo, and damaged cartilage has limited capacity for self-repair. For a joint to maintain its function throughout a lifetime of use, protective biologic mechanisms, such as lubrication, help minimize wear and damage that result from normal daily activities.

**Boundary lubrication** refers to the protective effect of particular lubricating molecules adsorbing to a surface and repelling its opposing interface. Bearing surfaces must generate a mutual repulsion to be lubricated in the boundary mode. Boundary lubricants exert their effects by changing the physicochemical characteristics of a surface and reduce articular friction and wear by providing a smooth and slippery coating. Friction is reduced by an interposed film of protective fluid that allows one surface to ride freely over another. The cartilage matrix is integral to this phenomenon because it is fluid filled and compressible. Loaded cartilage extrudes lubricant fluid from its surface, and expressed fluid contributes to the separation of the two articulating surfaces. Scanning electron microscopy has shown a continuous film of fluid, only 100 nm thick, which separates one surface from the other, preventing direct abrasive contact. This ultrathin coating of lubricant resists distraction of the two articulating surfaces, enhancing joint stability. Another essential advantage of an intra-articular lubrication system is the effective prevention of pinching of adjacent, well-vascularized synovial membrane.

**Hyaluronic Acid**

HA, a high-molecular-weight polysaccharide, is a major component of synovial fluid and cartilage. It is produced in large amounts by mechanosensitive, fibroblast-like synoviocytes. HA, of which three mammalian forms are designated HAS1, HAS2, and HAS3, is synthesized by HA synthase at the plasma membrane and is extruded directly into the extracellular compartment. HA synthase activity and HA secretion are stimulated by proinflammatory cytokines, including interleukin-1β and transforming growth factor-β. HA also is synthesized by many other skeletal cells and is an important component of extracellular matrices. It is simultaneously a solid phase matrix element of cartilage and other tissues, and a fluid phase element in the synovial space under normal and abnormal conditions.

HA has many biologic functions, which include effects on cell growth, migration, and adhesion. The regulatory role of HA is mediated through HA-binding proteins and receptors, including CD44, which are present on the cell surfaces of chondrocytes, lymphocytes, and other mononuclear cell populations. HA plays a crucial role in morphogenesis and in wound healing. Additionally, HA is a vital structural component of the synovial lining, and it has an essential role in the induction of joint cavity during embryogenesis. HA, produced by synovium, was originally thought to be primarily a joint lubricant, and it is generally accepted that it plays a major physiologic role in maintaining synovial fluid viscosity. It is important in normal joint function, not least through its capacity to provide effective shock absorption. It has been suggested that HA is a particularly important viscohydrodynamic lubricant at low-load interfaces, such as synovium-on-synovium and synovium-on-cartilage. Synovial fluid HA, acting in combination with albumin, has a role in the attenuation of fluid loss from the joint cavity, particularly during periods of increased pressure, which can occur during sustained joint flexion.

**Lubricin.** Compelling evidence suggests that lubricin, first described in the 1970s, is the factor primarily responsible for boundary lubrication of diarthrodial joints. Lubricin, a large secreted, mucin-like proteoglycan with an apparent molecular weight of 280 kD, is a product of the gene proteoglycan 4 (PRG4). It is a major component of synovial fluid and is present at the cartilage surface. The gene is highly expressed by human synovial fibroblasts and by superficial zone chondrocytes. Lubricin is closely related to superficial zone protein, megakaryocyte-stimulating factor, and hemangiopoietin, which are encoded by the same gene but can differ in terms of posttranslational modification. Superficial zone protein is expressed by SLCs and by superficial zone chondrocytes at the cartilage surface, but not by intermediate or deep zone chondrocytes. It has been suggested that lubricin may bind to the much longer hyaluronate polymers, distributing shear stress and stabilizing essential lubricant molecules.

In an experimental model, lubricin seemed to have multiple functions in articulating joints and tendons, including protection of cartilage surfaces from protein deposition and cell adhesion and inhibition of synovial cell overgrowth. In other studies of lubricin biology and joint integrity, experimental injury resulted in reduced synovial fluid lubricin concentrations, decreased boundary lubricating ability, and increased cartilage matrix degradation, each of which could be attributed to trauma-induced inflammatory processes.

Others have argued against the primacy of lubricin in joint lubrication by proposing that surface-active phospholipid, also secreted by intimal fibroblasts, is the essential boundary lubricant that reduces cartilage friction to remarkably low levels. It was hypothesized that lubricin acts as the carrier of surface-active phospholipid to articular cartilage but is not the lubricant per se, a function that is similar to that of the well-characterized alveolar surfactant binding proteins in the lung.

**Formation of Synovial Fluid**

In health, a constant volume of synovial fluid is important during joint movement as a cushion for synovial tissue and
Figure 2-7  Clinical appearance and radiographic changes in Prg4−/− mice. A and B, Photographs of the hind paws of 6-month-old Prg4−/− (A) and wild-type (B) mice. Note the curved digits in the mutant mouse and swelling at the ankle joint. C and D, Radiographs of the ankle joint of 9-month-old wild-type (C) and Prg4−/− mice (D). Structures corresponding to the tibia (t) and talus (ta) are indicated. Note the calcification of structures adjacent to the ankle (arrows in D). E, Lateral knee x-ray of a 4-month-old wild-type mouse. Structures corresponding to the patella (p), femoral condyle (f), tibial plateau (t), and fibula (fib) are indicated. F, Lateral knee x-ray of a 4-month-old Prg4−/− mouse. Note the increased joint space between the patella and femur (arrow) and osteopenia of the patella, femoral condyles, and tibial plateau. G, Shoulder x-ray of a 4-month-old wild-type mouse. Structures corresponding to the humeral head (h), glenoid fossa of the scapula (s), and lateral portion of the clavicle (c) are indicated. H, Shoulder x-ray of a 4-month-old Prg4−/− mouse. Note the increased joint space between the humerus and scapula (arrow) and osteopenia of the humeral head. (From Rhee DK, Marcelino J, Baker M, et al: The secreted glycoprotein lubricin protects cartilage surfaces and inhibits synovial cell overgrowth, J Clin Invest 115:622–631, 2005.)

Figure 2-8  Clinical features of camptodactyly–arthropathy–coxa vara–pericarditis (CACP) syndrome. A, The characteristic deformity of the hands is shown. B, Chest x-ray shows an enlarged cardiac outline caused by pericarditis. C, X-ray of the pelvis highlights coxa vara in a boy with CACP. (B and C, Courtesy Ronald Laxer, MD, Hospital for Sick Children, Toronto, Ontario, Canada.)
as a reservoir of lubricant for cartilage. Many of the soluble components and proteins in synovial fluid exit the synovial microcirculation through pores or fenestrations in the vascular endothelium, then diffuse through the interstitium before entering the joint space. Synovial permeability to most small molecules is determined by a process of free diffusion through the double barrier of endothelium and interstitium, limited mainly by the intercellular space between synovial lining cells. Fat-soluble molecules can diffuse through, and between, cell membranes; their passage across the synovial surface is less restricted. Additional components, including hyaluronan and lubricin, are produced by synovial lining cells.

![Figure 2-9](image)

**Figure 2-9** Schematic representation of the formation of synovial fluid. Many of the soluble components and proteins in synovial fluid exit the synovial subintimal microcirculation through pores or fenestrations in the vascular endothelium, then diffuse through the interstitium before entering the joint space. Synovial permeability to most small molecules is determined by a process of free diffusion through the double barrier of endothelium and interstitium, limited mainly by the intercellular space between synovial lining cells. Fat-soluble molecules can diffuse through, and between, cell membranes; their passage across the synovial surface is less restricted. Additional components, including hyaluronan and lubricin, are produced by synovial lining cells.

In contrast to hydrophilic molecules, fat-soluble molecules can diffuse through and between cell membranes, and their passage across the synovial surface is less restricted. The entire surface area of the synovium is available to lipophilic molecules that diffuse into and out of the joint space. Physiologically, the most important fat-soluble molecules are the respiratory gases, oxygen and carbon dioxide. When the joint is inflamed, synovial fluid may exhibit low partial pressure of oxygen, high partial pressure of carbon dioxide, decreased pH, and increased lactate production. The resultant hypoxia and acidosis can have serious implications for the synovial microcirculation and chondrocyte metabolism.

**Nutrition of Chondrocytes**

Another important function of synovium is to enhance the nutrition of chondrocytes, which are resident in articular cartilage (see Chapter 3). Because articular cartilage is avascular, delivery of nutrients to chondrocytes and removal of metabolic breakdown products from the cartilage are believed to occur through synovial fluid and synovial tissue arterioles and venules, as well as through subchondral bone. Morphologic, physiologic, and pathologic studies have confirmed that solutes pass easily from the synovial fluid into cartilage, and that cartilage does not survive without synovial fluid contact in vivo. Within the cartilage matrix, three potential mechanisms for nutrient transfer have been proposed: diffusion, active transport by chondrocytes, and pumping by intermittent compression of cartilage matrix. A large proportion of hyaline cartilage lies within 50 µm of a synovial surface and its rich supply of blood vessels. Chondrocytes are oxygen sensitive and are well adapted to living in hypoxic conditions. Low oxygen tension promotes expression of the chondrocyte phenotype and cartilage-specific matrix formation. Reactive oxygen species also may play a crucial role in the regulation of some normal chondrocytic activities, such as cell activation, proliferation, and matrix remodeling.

**SUMMARY**

The normal human synovial membrane is a highly specialized, multifunctional organ that is vital for mobility, independence, and survival. The intimal layer is composed of two distinct cell phenotypes with characteristics of macrophage and fibroblast lineages. Synovial macrophages express CD45, CD163 and CD97, CD68, neuron-specific esterase, and cathepsins B, L, and D. Cells expressing CD14 are rarely seen in the healthy intimal layer. FcyRIII
(CD16), expressed by Kupffer cells of the liver and type II alveolar macrophages of the lung, is expressed on a subpopulation of synovial macrophages. Synovial macrophages also express the MHC class II molecule and play a central role in phagocytosis and in antigen-mediated immune responses.

Synovial intimal fibroblasts possess prominent synthetic capacity and produce the essential joint lubricants HA and lubricin. They also synthesize normal matrix components, including fibronectin, laminin, collagens, proteoglycans, lubricin, and other identified and unidentified proteins. They have the capacity to produce large quantities of metalloproteinases, metalloproteinase inhibitors, prostaglandins, and cytokines. Expression of selected adhesion molecules on synovial fibroblasts probably facilitates the trafficking of some cell populations, such as polymorphs, into synovial fluid, and the retention of others, such as mononuclear leukocytes, within synovial tissue.

The subintimal layer is composed of a loose connective tissue matrix and contains branching blood and lymphatic vessels; a nerve supply; and a variety of resident cell populations, including infiltrating macrophages and fibroblasts. The nerve supply is important in regulating synovial blood flow. Lymphatic vessels allow egress of metabolic breakdown products from the synovium and synovial fluid. The morphology of the subintimal layer varies according to the anatomic location and the local functional requirements.

Coordinated functions of the composite synovial membrane are essential for normal joint movement, formation of synovial fluid, nutrition of chondrocytes, and protection of cartilage. These functions must be preserved over a lifetime at multiple anatomic locations. Absence of essential constituents of synovial fluid, such as lubricin, or inadequate cartilage protection results in early articular malfunction, which may progress to variable degrees of joint failure. The characteristics of lubricin deficiency have been elegantly described in animal models and in humans. Additional studies may define novel clinical categories of degenerative polyarthritis that are associated with other specific disorders of synovial membrane function.

References

The references for this chapter can also be found on www.expertconsult.com.
Cartilage and Chondrocytes

MARY B. GOLDRING

KEY POINTS

Articular cartilage matrix is heterogeneous and contains numerous extracellular matrix (ECM) proteins, of which the large aggregating proteoglycan aggrecan and collagen types II, IX, and XI are the major constituents.

The collagen network of cartilage confers tensile strength, and aggrecan provides resistance to compression.

Adult articular chondrocytes are nonmitotic cells that survive at low oxygen tension in the absence of a vascular supply.

In response to trauma or inflammation, the metabolic activity of the chondrocyte is increased in response to catabolic and anabolic factors that regulate remodeling of the ECM.

Under physiologic conditions, the chondrocyte maintains low-turnover repair of proteoglycans, but the repair capacity, responses to anabolic factors, cell survival, and quality of the matrix decline with age.

Hyaline cartilage, including the articular cartilage of diarthrodial joints, consists of a single cellular component, the chondrocyte, which is embedded in a unique and complex matrix. Adult articular chondrocytes are considered to be fully differentiated cells that maintain matrix constituents in a low-turnover state of equilibrium. Chondrocytes serve diverse functions during development and postnatal life. In the embryo, the chondrocyte arises from mesenchymal progenitors from diverse sources, including the cranial neural crest of the neural ectoderm, cephalic mesoderm, sclerotome of the paraxial mesoderm, and somatopleure of the lateral plate mesoderm, depending upon the ultimate location of the cartilage. The chondrocyte synthesizes the templates, or cartilage anlagen, through a process termed chondrogenesis.

After mesenchymal condensation and chondroprogenitor cell differentiation, the chondrocytes undergo proliferation, terminal differentiation to chondrocyte hypertrophy, and apoptosis through a process termed endochondral ossification, whereby the hypertrophic cartilage is replaced by bone. A similar sequence of events occurs in the postnatal growth plate and leads to rapid growth of the skeleton. Processes that control the different stages of skeletal development are described in Chapter 1.

In adults, the anatomic distribution of cartilage is restricted primarily to the joints, trachea, and nasal septum, where the major function is structural support. In joints, cartilage has the additional function of providing low-friction articulation. Adult articular cartilage comprises a specialized matrix of collagens, proteoglycans, and other cartilage-specific and nonspecific proteins. Adult articular chondrocytes, remnants of the resting, or reserve, chondrocytes that laid down the original cartilage matrix during chondrogenesis, are inactive metabolically, owing partially to the absence of a vascular supply and innervation in the tissue. The clinical importance of the adult chondrocyte resides in its capacity to respond to mechanical stimuli, growth factors, and cytokines that may influence normal homeostasis in a positive or negative manner. In rheumatoid arthritis (RA), cartilage destruction occurs primarily in areas contiguous with the proliferating synovial pannus, although evidence indicates that the chondrocyte can respond to the inflammatory milieu and participate in degrading its own matrix. In osteoarthritis (OA), the chondrocyte plays a key role by reacting to structural changes in the surrounding cartilage matrix through the production of catabolic cytokines and anabolic factors, which act in an autocrine-paracrine manner. Nevertheless, the chondrocyte has limited capacity, which declines with age, to regenerate the normal cartilage architecture with zonal variations in the matrix network that was formed originally. This chapter focuses on the structure and function of normal articular cartilage and the role of the chondrocyte in maintaining cartilage homeostasis and responding to adverse environmental insults that may modify cartilage integrity.

CARTILAGE STRUCTURE

Normal articular cartilage is a specialized tissue characterized macroscopically by its milky, shelled-almond (hyaline) appearance. It is an avascular tissue nourished by diffusion from the vasculature of the subchondral bone and from the synovial fluid. Articular cartilage is more than 70% water, and it is hypocellular compared with other tissues; chondrocytes constitute only 1% to 2% of its total volume. Most of the dry weight of cartilage consists of two components: type II collagen and the large aggregating proteoglycan, aggrecan. Several "minor" collagens and small proteoglycans also seem to play a role in cartilage-matrix organization, however.

Organic constituents represent only about 20% of the wet weight of cartilage. Collagen, primarily type II, accounts for approximately 15% to 25% of the wet weight and about half of the dry weight except in the superficial zone, where it represents most of the dry weight. Proteoglycans, primarily aggrecan, account for 10% of the wet weight and about 25% of the dry weight. The highly cross-linked type II collagen–containing fibrils form a systematically oriented network that traps the highly negatively charged proteoglycan aggregates. Histochemical analysis of cartilage shows that proteoglycans can be stained reliably with safranin O,
Collagen also can be stained efficiently, but differentiation of collagen types requires immunostaining with specific antibodies.

Despite its thinness (≤7 mm) and apparent homogeneity, mature articular cartilage is a heterogeneous tissue with four distinct regions: (1) the superficial tangential (or gliding) zone, (2) the middle (or transitional) zone, (3) the deep (or radial) zone, and (4) the calcified cartilage zone, which is located immediately below the tidemark and above the subchondral bone (Figure 3-1).6,9,11,12 In the superficial zone, the chondrocytes are flattened, and the matrix comprises thin collagen fibrils in tangential array, associated with a high concentration of the small proteoglycan decorin and a low concentration of aggrecan. The middle zone, composing 40% to 60% of cartilage weight, consists of rounded chondrocytes surrounded by radial bundles of thick collagen fibrils. In the deep zone, chondrocytes frequently are grouped in columns or clusters (Figure 3-2). In this region, collagen bundles are the thickest and are arranged in a radial fashion.

Cell density progressively decreases from the surface to the deep zone, where it is one-half to one-third the density in the superficial zone13; chondrocytes in the deep and middle zones have a cell volume that is twice that of superficial chondrocytes.14 Water is 75% to 80% of the wet weight in the superficial zone and progressively decreases to 65% to 70% with increasing depth. Greater amounts of collagen relative to proteoglycans are present in the superficial zone, compared with the middle and deep zones, and type I collagen may be synthesized, in addition to type II collagen.15,16 With increasing depth, the proportion of proteoglycan increases to 50% of the dry weight in the deep zone.14,17,19 The calcified zone is formed as a result of endochondral ossification and persists after growth plate closure as the histologically defined tidemark.20 The calcified zone serves as an important mechanical buffer between uncalcified articular cartilage and subchondral bone.

The physical properties of articular cartilage are determined by the unique fibrillar collagen network, which provides tensile strength, interspersed with proteoglycan aggregates that bestow compressive resilience.21-23

Figure 3-1 The structure of human adult articular cartilage, showing zones of cellular distribution and the pericellular, territorial, and interterritorial regions of matrix organization. Insets show the relative diameters and orientations of collagen fibrils in the different zones. The positions of the tidemark and subchondral bone and other special features of matrix composition also are noted. (From Poole AR, Kojima T, Yasuda T, et al: Composition and structure of articular cartilage: a template for tissue repair, Clin Orthop Relat Res [391S]:S26–S33, 2001. Copyright Lippincott Williams & Wilkins.)
35

CHAPTER 3  |  CARTILAGE AND CHONDROCYTES

Figure 3-2  Light micrograph of vertically sectioned adult human cartilage (femoral condyle), illustrating its subdivision into superficial (S), transitional (T), upper radial (U), lower radial (L), and calcified cartilage (M) zones; the last mentioned abuts on the subchondral bone plate (B). Saw-cut, 100 µm thick surface stained with basic fuchsin, McNeil’s tetrachrome, and toluidine blue O. (From Hunziker EB: Articular cartilage structure in humans and experimental animals. In Kuettner KE, Schleyerbach R, Peyron JG, et al, editors: Articular cartilage and osteoarthritis, New York, 1992, Raven, pp 183–199.)

Proteoglycans are associated with large quantities of water bound to the hydrophilic glycosaminoglycan chains. This cartilaginous extracellular matrix (ECM), with its tightly bound water, provides a high degree of resistance to deformation by compressive forces. The capacity to resist compressive forces is associated with the ability to extrude water as the cartilage compresses. When compression is released, the proteoglycans (now depleted of balancing counter ions that were removed with the water) contain sufficient fixed charge to reabsorb osmotically the water and small solutes into the matrix, which then rebounds to its original dimensions.24,25

STRUCTURE-FUNCTION RELATIONSHIPS OF CARTILAGE MATRIX COMPONENTS

ECM components synthesized by chondrocytes include highly cross-linked fibrils of triple-helical type II collagen molecules that interact with other collagens, aggrecan, small proteoglycans, and other cartilage-specific and nonspecific matrix proteins (Table 3-1).5,7,8,11 The importance of these structural proteins may be observed in heritable disorders, such as chondrodysplasias, or in transgenic animals in which mutations or deficiencies in cartilage genes result in cartilage abnormalities.26-30 Deficiencies or disruptions in genes that encode the cartilage-specific collagens result, in some cases, in premature OA.28 Knowledge of the composition of the cartilage matrix has permitted the development of methods for identifying molecular markers in serum and synovial fluid that can be used to monitor changes in cartilage metabolism and to assess cartilage damage in OA or RA.31-34 Changes in the structural composition of cartilage can markedly affect its biomechanical properties (see Chapter 6).

Cartilage Collagens

The major component of the collagen network in adult articular cartilage is the triple-helical type II collagen molecule, which is composed of three identical α chains (α1[II]). These molecules are assembled in fibrils in a quarter-stagger array that can be observed by electron microscopy.7,15,16 These fibrils are thinner than type I collagen–containing fibrils in skin because of the higher numbers of hydroxylysine residues that can form cross-links and the presence of other collagen and noncollagen

Table 3-1  Extracellular Matrix Components of Cartilage

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Structure and Size</th>
<th>Function and Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Collagens</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type II</td>
<td>[α1(II)]3; fibril-forming</td>
<td>Tensile strength; major component of collagen fibrils</td>
</tr>
<tr>
<td>Type IX</td>
<td>[α1(IX)α2(IX)α3(IX)]; single CS or DS chain; α1(II) gene encodes α3(IX); FACIT</td>
<td>Tensile properties, interfibrillar connections; cross-links to surface of collagen fibril, NC4 domain projects into matrix</td>
</tr>
<tr>
<td>Type XI</td>
<td>[α1(XI)α2(XI)α3(XI)]; fibril-forming</td>
<td>Nucleation/control of fibril formation; within collagen fibril</td>
</tr>
<tr>
<td>Type VI</td>
<td>[α1(VI)α2(VI)α3(VI)]; microfibrils</td>
<td>Forms microfibrillar network, binds hyaluronan, biglycan, decorin; pericellular</td>
</tr>
<tr>
<td>Type X</td>
<td>[α1(X)3]; hexagonal network</td>
<td>Support for endochondral ossification; hypertrophic zone and calcified cartilage</td>
</tr>
<tr>
<td>Type XII</td>
<td>[α1(XII)3]; FACIT large cruciform NC3 domain</td>
<td>Associated with type I collagen fibrils in perichondrium and articular surface</td>
</tr>
<tr>
<td>Type XIV</td>
<td>[α1(XIV)3]; FACIT</td>
<td>Associated with type I collagen; superficial zone</td>
</tr>
<tr>
<td>Type XVI</td>
<td>[α1(XVI)3]; FACIT</td>
<td>Integrates with collagen II/IX fibrils</td>
</tr>
<tr>
<td>Type XXVII</td>
<td>Col27α1 gene: 156 kb, 61 exons</td>
<td>Fibril-forming; developing cartilage</td>
</tr>
</tbody>
</table>

Continued
components in the fibril. Type IIB collagen in articular cartilage is a product of alternative splicing and lacks a 69 amino acid, cysteine-rich domain of the amino-terminal propeptide, which is encoded by exon 2 in the human type II collagen gene (COL2A1). This domain is found in type IIA procollagen, which is expressed by chondroprogenitor cells during development, and in the amino propeptides of other interstitial collagen types, and may play a feedback-inhibitory role in collagen biosynthesis. The reappearance of type IIA collagen in the midzone pericellular matrix and type X collagen, the hypertrophic chondrocyte marker, in the deep zone of OA cartilage suggests reversion to a

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Structure and Size</th>
<th>Function and Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggrecan</td>
<td>255 kD core protein; CS/HS side chains; C-terminal EGF and lectin-like domains</td>
<td>Compressive stiffness through hydration of fixed charge density; binding through G1 domain to HA stabilized by link protein</td>
</tr>
<tr>
<td>Versican</td>
<td>265-370 kD core protein; CS/DS side chains; C-terminal EGF, C-type lectin, and CRP-like domains</td>
<td>Low levels in articular cartilage throughout life; calcium-binding and selectin-like properties</td>
</tr>
<tr>
<td>Perlecan</td>
<td>400-467 kD core protein; HS/CS side chains; no HA binding</td>
<td>Cell-matrix adhesion; pericellular</td>
</tr>
<tr>
<td>Biglycan</td>
<td>38 kD; LRR core protein with two DS chains (76 kD)</td>
<td>Binds collagen VI and TGF-β1; pericellular</td>
</tr>
<tr>
<td>Decorin</td>
<td>36.5 kD; LRR core protein with one CS or DS side chain (100 kD)</td>
<td>Controls size/shape of collagen fibrils, binds collagen II and TGF-β1, interterritorial</td>
</tr>
<tr>
<td>Asporin</td>
<td>40 kD; LRR core protein; N-terminal extension of 15 aspartate residues</td>
<td>Binds collagen, modulates TGF-β1 function</td>
</tr>
<tr>
<td>Fibromodulin</td>
<td>42 kD; containing KS chains in central LRR region and N-terminal tyrosine sulfate domains</td>
<td>Same as decorin</td>
</tr>
<tr>
<td>Lumican</td>
<td>38 kD; structure similar to fibromodulin</td>
<td>Same as decorin</td>
</tr>
<tr>
<td>PRELP</td>
<td>44 kD; LRR core protein; proline-rich and arginine-rich N-terminal binding domain for heparin and HS</td>
<td>Mediates cell binding through HS in syndecan</td>
</tr>
<tr>
<td>Chondroadherin</td>
<td>45 kD; LRR core protein without N-terminal extension</td>
<td>Binding to cells via α2β1 integrin</td>
</tr>
</tbody>
</table>

**Other Molecules**

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Structure and Size</th>
<th>Function and Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyaluronic acid (HA); Hyaluronan</td>
<td>1000-3000 kD</td>
<td>Retention of aggrecan within matrix</td>
</tr>
<tr>
<td>Link protein</td>
<td>38.6 kD</td>
<td>Stabilizes attachment of aggrecan G1 domain to HA</td>
</tr>
<tr>
<td>Cartilage oligomeric matrix protein (COMP)</td>
<td>550 kD; five 110-kD subunits; thrombospondin-like</td>
<td>Interterritorial in articular cartilage; stabilizes collagen network or promotes collagen fibril assembly; calcium binding</td>
</tr>
<tr>
<td>Cartilage matrix protein (CMP, or matrilin-1); matrilin-3</td>
<td>Three 50-kD subunits with vWF and EGF domains</td>
<td>Tightly bound to aggrecan in immature cartilage</td>
</tr>
<tr>
<td>Cartilage intermediate layer protein (CILP)</td>
<td>92 kD; homology with nucleotide pyrophosphohydrolase without active site</td>
<td>Restricted to middle/deep zones of cartilage; increase in early and late osteoarthritis</td>
</tr>
<tr>
<td>Glycoprotein (gp)-39, YKL-40, or chitinase 3-like protein 1 (CH3L1)</td>
<td>39 kD; chitinase homology</td>
<td>Marker of cartilage turnover; chondrocyte proliferation; superficial zone of cartilage</td>
</tr>
<tr>
<td>Fibronectin</td>
<td>Dimer of 220 kD subunits</td>
<td>Cell attachment and binding to collagen and proteoglycans; increased in osteoarthritis cartilage</td>
</tr>
<tr>
<td>Tenascin-C</td>
<td>Six 200-kD subunits forming hexabrachion structure</td>
<td>Binds syndecan-3 during chondrogenesis; angiogenesis</td>
</tr>
<tr>
<td>Superficial zone protein (SZP), lubricin, or proteoglycan (PRG) 4</td>
<td>225 kD, 200 nm length</td>
<td>Joint lubrication; superficial zone only</td>
</tr>
</tbody>
</table>

**Membrane Proteins**

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Structure and Size</th>
<th>Function and Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD44</td>
<td>Integral membrane protein with extracellular HS/CS side chains</td>
<td>Cell-matrix interactions; binds HA</td>
</tr>
<tr>
<td>Syndecan-1, -3, -4</td>
<td>N-terminal HS attachment site; cytoplasmic tyrosine residues</td>
<td>Syndecan-3 is receptor for tenascin-C during cartilage development; cell-matrix interactions</td>
</tr>
<tr>
<td>Anchorin CII (or annexin V)</td>
<td>34 kD; homology to calcium-binding proteins calpain and lipocortin</td>
<td>Cell surface attachment to type II collagen; calcium binding</td>
</tr>
<tr>
<td>Integrins (α1, α2, α3, α5, α6, α10, β1, β3, β5)</td>
<td>Two noncovalently linked transmembrane glycoproteins (α and β subunits)</td>
<td>Cell-matrix binding: α1β1/collagen I or VI, α2β1 or α3β1/collagen II, α5β1/fibronectin; intracellular signaling</td>
</tr>
<tr>
<td>Discoidin domain receptor 2</td>
<td>Receptor tyrosine kinase</td>
<td>Binds native type II collagen fibrils; Ras/ERK signaling</td>
</tr>
</tbody>
</table>
developmental phenotype in an attempt to repair the damaged matrix. 35,39

Although collagens VI, IX, XI, XII, and XIV are quantitatively minor components, they may have important structural and functional properties. Collagens IX and XI are specific to cartilage, whereas collagens VI, XII, and XIV are widely distributed in other connective tissues. 3 Collagen VI, which is present in cartilage as microfibrils in very small quantities in the pericellular matrix, may play a role in cell attachment and interacts with other matrix proteins, such as hyaluronic, decorin, and biglycan. Small amounts of collagen III are found in cartilage, and collagen VI may increase in OA cartilage. 39

Type IX collagen is a proteoglycan and a collagen because it contains a chondroitin sulfate chain attachment site in one of the noncollagen domains. The helical domains of the type IX collagen molecule form covalent cross-links with type II collagen telopeptides and are attached to the fibrillar surface, as observed by the electron microscope. Type IX collagen may function as a structural intermediate between type II collagen fibrils and the proteoglycan aggregates, serving to enhance the mechanical stability of the fibril network and resist the swelling pressure of the trapped proteoglycans. Destruction of type IX collagen accelerates cartilage degradation and loss of function. 7,35,36

The α3 chain of type XI collagen has the same primary sequence as the α1(II) chain, and the heterotrimeric type XI collagen molecule is buried in the same fibril as type II collagen. Type XI collagen may have a role in regulating fibril diameter. The more recently discovered nonfibrillar fibril-associated collagens with interrupted triple helices (FACIT), XII and XIV, which are structurally related to type IX collagen, do not form fibrils by themselves but co-aggregate with fibril-forming collagens and modulate the packing of collagen fibers through domains projecting from their surfaces. 7,35,36

**Cartilage Proteoglycans**

The major proteoglycan in articular cartilage is the large aggregating proteoglycan, or aggrecan, which consists of 100 chondroitin sulfate chains, 30 keratan sulfate chains, and shorter N-linked and O-linked oligosaccharides. 5,8,40,41 Link protein, a small glycoprotein, stabilizes the noncovalent linkage between aggrecan and hyaluronic acid (also called hyaluronan) to form the proteoglycan aggregate that may contain 100 aggrecan monomers. The G1 and G2 N-terminal globular domains of aggrecan and its C-terminal G3 domain have distinct structural properties that function as integral parts of the aggrecan core protein and contribute cleavage products that accumulate with age or in OA. The G2 domain is separated from G1 by a linear interglobular domain and has two proteoglycan tandem repeats. The G3 domain contains sequence homologies to epidermal growth factor, lectin, and complement regulatory protein, and participates in growth regulation, cell recognition, intracellular trafficking, and recognition, assembly, and stabilization of the ECM. About half of the aggrecan molecules in adult cartilage lack the G3 domain, probably as a result of proteolytic cleavage during matrix turnover. Small quantities of other large proteoglycans are found in cartilage, including versican, which forms aggregates with hyaluronic acid, and perlecan, which is nonaggregating; however, these proteoglycans function primarily during skeletal development, where versican is expressed in prechondrogenic condensations, and perlecan is expressed in the cartilage anlagen after expression of type II collagen and aggrecan. 42

The nonaggregating small proteoglycans are not specific to cartilage, but in cartilage they serve specific roles in matrix structure and function, primarily by modulating collagen-fibril formation. 42-44 Of the more than 10 leucine-rich repeat (LRR) proteoglycans discovered so far, only osteoadherin is not present in cartilage. The 24 amino acid central LRR domain is conserved, but the N-terminal and C-terminal domains have patterns of cysteine residues involved in intrachain disulfide bonds that distinguish the four subfamilies: (1) biglycan, decorin, fibromodulin, and lumican; (2) keratocan and proline and arginine-rich end leucine-rich repeat protein (PRELP); (3) chondroadherin; and (4) epiphycan/PG-Lb and mimecan/osteoglycin. Biglycan may have two glycosaminoglycan chains—chondroitin sulfate or dermatan sulfate, or both—at the N-terminus through two closely spaced serine-glycine dipeptides. Decorin contains one only chondroitin sulfate or dermatan sulfate chain. Fibromodulin and lumican contain keratan sulfate chains linked to the central domain of the core protein and several sulfated tyrosine residues in the N-terminus. Negatively charged glycosaminoglycan side chains contribute to the fixed charge density of the matrix and, together with the highly anionic tyrosine-sulfation sites, permit multiple-site linkage between adjacent collagen fibrils, stabilizing the network. Decorin, the most extensively studied LRR proteoglycan, binds to collagens II, VI, XII, and XIV, and to fibronectin and thrombospondin. Biglycan, decorin, and fibromodulin bind transforming growth factor (TGF)-β and the epidermal growth factor receptor and may modulate growth, remodeling, and repair. PRELP and chondroadherin may regulate cell-matrix interactions through binding to syndecan and α2β1 integrin.

**Other Extracellular Matrix and Cell Surface Proteins**

Several other noncollagenous matrix proteins may play important roles in determining cartilage matrix integrity. Cartilage oligomeric protein (COMP), a member of the thrombospondin family, is a disulfide-bonded, pentameric, 550 kD, calcium-binding protein that constitutes approximately 10% of the noncollagenous, nonproteoglycan protein in normal adult cartilage. COMP is located in the interterritorial matrix of adult articular cartilage, where it interacts with the COL3 and NC4 domains of type IX collagen that protrude from the fibril, stabilizing the collagen network. COMP is pericellular in the proliferating region of the growth plate, where it may have a role in cell-matrix interactions. 45 The cartilage matrix protein (or matrilin-1) and matrilin-3 are expressed in cartilage at certain stages of development and are present in tracheal cartilage. Matrilin-1 is present in the pericellular matrix of adult articular cartilage. 46,47
Tenascin-C, a glycoprotein that is regulated in development, is characteristic of nonossifying cartilage. Similar to fibronectin, alternative splicing of tenascin-C mRNA gives rise to different protein products at different stages of chondrocyte differentiation. Both proteins are increased in OA cartilage and may serve specific functions in remodeling and repair. A splice variant of tenascin-C mRNA is found in chondrosarcomas. The cartilage intermediate-layer protein (CILP) is expressed by chondrocytes in the middle to deep zones of articular cartilage as a precursor protein. When cleaved during secretion, CILP has structural similarities with nucleotide pyrophosphohydrolase, although it lacks the catalytic site, and it may play a role in pyrophosphate metabolism and calcification. Asporin is related to decorin and biglycan and, similar to those other LRR proteins, may interact with and sequester growth factors such as TGF-β. YKL-40/HC-gp39, also known as chitinase 3-like protein 1, is found only in the superficial zone of normal cartilage and stimulates proliferation of chondrocytes and synovial cells. Chitinase 3-like protein 1 is induced by inflammatory cytokines and may function as a feedback regulator because it inhibits cytokine-induced cellular responses. Synthesis or release of these proteins or fragments is often increased in cartilage that is undergoing repair or remodeling, and they have been investigated as markers of cartilage damage in arthritis. A related member of the chitinase family, YKL-39, may be a more specific serum marker as a cartilage-derived autoantigen.

MORPHOLOGY, CLASSIFICATION, AND NORMAL FUNCTION OF CHONDROCYTES

**Morphology**

The characteristic feature of the chondrocyte embedded in cartilage matrix is its rounded or polygonal morphology. The exception occurs at tissue boundaries, such as the articular surfaces of joints, where chondrocytes may be flattened or discoid. Intracellular features, including a rough endoplasmic reticulum, a juxtanuclear Golgi apparatus, and deposition of glycosyn, are characteristic of a synthetically active cell. Stockwell and Meachim calculated that the cell density of full-thickness, human, adult, femoral condyle cartilage is maintained at 14.5 \( \pm 3.0 \times 10^3 \) cells/mm\(^2\) from age 20 to 30 years. Because senescence of chondrocytes is known to occur with aging, it is logical to suppose that dead chondrocytes are replaced by mitosis. Mitotic figures are not observed, however, in normal adult articular cartilage.

The morphology, density, and synthetic activity of an adult chondrocyte vary according to its position within the different zones of articular cartilage. In the region of highest cell density, the superficial zone, cells are flattened and are oriented parallel to the surface, along with the collagen fibers. Chondrocytes within the middle zone appear larger and more rounded and display a random distribution within the matrix, where the collagen fibers also are more randomly arranged. Chondrocytes in the deeper zones form columns that, along with the collagen fibers, are oriented perpendicular to the cartilage surface. Chondrocytes may exhibit different behaviors depending on their position within the different layers, and these zonal differences in synthetic properties may persist in primary chondrocyte cultures. Chondrocyte volume in situ increases from the superficial through the deep zones and with the degree of cartilage degeneration. A study using confocal scanning laser microscopy of live, unfixed cartilage has revealed fine cytoplasmic processes extending from the cell bodies of 40% of chondrocytes. These processes are proposed to permit interactions among chondrocytes and the cartilage matrix at near and remote sites. They are distinct from the cilia, which are observed by electron microscopy but not by confocal scanning laser microscopy.

**Classification: Cell Origin and Differentiation**

Chondrocyte arises in the embryo from mesenchymal origin during chondrogenesis, which is the earliest phase of skeletal development involving mesenchymal cell recruitment, migration, and condensation and differentiation of mesenchymal chondroprogenitor cells. As described in detail in Chapter 1, chondrogenesis results in the formation of cartilage anlagen, or templates, at sites where skeletal elements form. This process is controlled by cell-cell and cell-matrix interactions and by growth and differentiation factors that initiate or suppress cellular signaling pathways and transcription of specific genes in a temporospatial manner.

Vertebrate limb development is controlled by interacting patterning systems involving fibroblast growth factor (FGF), hedgehog, bone morphogenetic protein (BMP), and Wnt pathways. Wnt signaling, via the canonical β-catenin pathway and activation of TCF/Lef transcription factors, functions in a cell-autonomous manner to induce osteoblast differentiation and suppress chondrocyte differentiation in early chondroprogenitors. During chondrogenesis, Wnt/β-catenin acts at two stages: at low levels to promote chondroprogenitor differentiation, and later at high levels to promote chondrocyte hypertrophic differentiation and subsequent endochondral ossification. The transcription factor, Sry-type high-mobility group box 9 (Sox9), is an early marker of the differentiating chondrocyte that is required for the onset of expression of type II collagen, aggrecan, and other cartilage-specific matrix proteins, such as type IX collagen. Two other members of the SOX family, L-Sox5 and Sox6, are not present in early mesenchymal condensations but are required during overt chondrocyte differentiation, forming heterodimers that induce transcription more efficiently than Sox9 by itself. Expression of SOX proteins depends on signaling through mammalian homologs of Drosophila mothers against decapentaplegic (SMADs), which are functionally redundant and active in differentiating chondrocytes. A long form of c-Maf interacts with Sox9 at early stages to upregulate COL2A1 gene expression, whereas C/EBPβ and C/EBPα and AP-2α may inhibit chondrocyte differentiation by blocking transcription of COL2A1, aggrecan (ACAN), and other cartilage-specific genes through direct or indirect mechanisms.

In the embryonic or postnatal epiphysial growth plates, upregulation of molecules that promote matrix remodeling and angiogenesis facilitates endochondral ossification, whereby bone replaces the calcified cartilaginous matrix in the hypertrophic zone (see Chapter 1). Differentiated
chondrocytes that remain in the reserve, or resting, zone become the cartilage elements in articular joints, or they can proliferate and undergo the complex process of terminal differentiation to hypertrophy marked by type X collagen. Indian hedgehog and parathyroid hormone–related protein transiently induce proliferation and repress differentiation, determining the number of cells that enter the hypertrophic maturation pathway. The Runx domain transcription factor, Runx2 (also known as core binding factor or Cbfa1), serves as a positive regulatory factor in chondrocyte maturation to the hypertrophic phenotype and subsequent osteogenesis. Runx2 is expressed in the adjacent perichondrium and in prehypertrophic chondrocytes, but less in late hypertrophic chondrocytes, and is required for expression of type X collagen and other markers of terminal differentiation.

Numerous other transcription factors positively or negatively regulate chondrocyte terminal differentiation by controlling the expression or activity of Runx2. BMP-induced Smad 1 and interactions between Smad 1 and Runx2 are required for the induction of chondrocyte hypertrophy. Because no SMAD site is present on the Runx2 promoter, it has been proposed that homeodomain proteins such as Dlx3 may activate Runx2 signaling in response to BMP-2 during endochondral ossification, whereas Dlx5, Dlx6, and Msx2 are known to inhibit Runx2-mediated activation of genes such as osteocalcin at later stages. The homeodomain protein Nkx3.2, which is an early BMP-induced signal required at the onset of chondrogenesis, is a direct transcriptional repressor of Runx2 promoter activity. Histone deacetylase 4 (HDAC4), which is expressed later in prehypertrophic chondrocytes, prevents premature chondrocyte hypertrophy by interfering with Runx2 and inhibiting its activity. The hypoxia-inducible factor (HIF)-1α is required for chondrocyte survival during hypertrophic differentiation, owing partially to its regulation of vascular endothelial cell growth factor (VEGF) expression.

The leucine zipper proteins Fra2, ATF2, and c-Maf are required for gene expression during late-stage hypertrophy (see for review, Goldring and Sandell). One major function of the chondrocyte is growth of the skeleton through increased cell proliferation, production of ECM, and cell volume through hypertrophy. After cessation of growth, the resting chondrocyte remains as part of the supporting structures in articular, tracheal, and nasal cartilages, indicating that the fate of a chondrocyte depends on origin and location (Figure 3-3).

Normal Function of the Adult Articular Chondrocyte

The mature articular chondrocyte embedded in its ECM is a resting cell with no detectable mitotic activity and a low rate of synthetic activity. Because articular cartilage is not vascularized, the chondrocyte must rely on diffusion from the articular surface or subchondral bone for exchange of nutrients and metabolites. Chondrocytes maintain active membrane transport systems for exchange of cations, including Na⁺, K⁺, Ca²⁺, and H⁺, whose intracellular concentrations fluctuate with load and changes in the composition of the cartilage matrix. The chondrocyte cytoskeleton is composed of actin, tubulin, and vimentin filaments, and the composition of these filament systems varies in the different cartilage zones.

Chondrocyte metabolism operates at low oxygen tension within the cartilage matrix, ranging from 10% at the surface to less than 1% in the deep zone. The consumption of oxygen by cartilage on a per-cell basis is only 2% to 5% of that in liver or kidney, although the amounts of lactate produced are comparable. Chondrocytes do not normally contain abundant mitochondria, energy metabolism depends strongly on the glucose supply, and the energy requirements may be modulated by mechanical stress. Glucose serves as the major energy source for chondrocytes and as an essential precursor for glycosaminoglycan synthesis. Facilitated glucose transport in chondrocytes is mediated by several distinct glucose transporter proteins (GLUTs) that may be constitutively expressed (GLUT3 and GLUT8) or cytokine inducible (GLUT1 and GLUT6). A proteomic study of chondrocytes identified 93 different intracellular proteins known to be involved in cell organization (26%), energy (16%), protein fate (14%), metabolism (12%), and cell stress (12%). The relative expression of these proteins may determine the capacity of chondrocytes to survive in cartilage matrix and to modulate metabolic activity in response to environmental changes.

When cultured in a range of oxygen tensions between severe hypoxia (0.1% oxygen) and normoxia (21% oxygen), chondrocytes adapt to low oxygen tension by upregulating HIF-1α. Hypoxia via HIF-1α can stimulate chondrocytes to express GLUTs and angiogenic factors such as VEGF and numerous genes associated with cartilage anabolism and chondrocyte differentiation, including Sox9, TGF-β, and connective tissue growth factor. In the growth plate, hypoxia and HIF-1α are associated with type II collagen production, whereas hyperoxia (55% oxygen) may increase the breakdown of cartilage collagen in articular cartilage, where it maintains tonic activity during physiologic hypoxia in the deeper layers associated with increased proteoglycan synthesis.

Long-term systemic hypoxia (13%) may downregulate collagen and aggrecan gene expression in articular cartilage, whereas hyperoxia (55% oxygen) may increase the breakdown of cartilage collagens in articular cartilage in the presence of vascularized rheumatoid synovium. By modulating the intracellular expression of survival factors such as HIF-1α, chondrocytes have a high capacity to survive in the avascular cartilage matrix and to respond to environmental changes. Findings that catabolic stress and inflammatory cytokines upregulate HIF-1α suggest that it may serve as a survival factor in OA cartilage. More recent studies indicate that another hypoxia-inducible factor, HIF-2α, can become activated by proinflammatory cytokines and stress in the avascular, hypoxic environment of articular cartilage.

The chondrocyte maintains a steady-state metabolism secondary to equilibrium between anabolic processes and catabolic processes that results in the normal turnover of matrix molecules. In normal adult articular cartilage, the turnover of matrix components is low. The turnover of collagen has been estimated to occur with a half-life greater...
than 100 years. In contrast, glycosaminoglycan constituents on the aggrecan core protein are more readily replaced; the half-life of aggrecan subfractions has been estimated to range from 3 to 24 years. Other cartilage ECM components, including biglycan, decorin, COMP, tenascins, and matrilins, incorporated previously into the matrix during development also may be synthesized by chondrocytes under low-turnover conditions. Regional differences in the remodeling activities of chondrocytes have been noted, however, and matrix turnover may be more rapid in the immediate pericellular zones. The metabolic potential of these cells is indicated by their capacity to proliferate in culture and to synthesize matrix proteins after enzymatic release from the cartilage of even elderly individuals.

The complex composition of the articular cartilage matrix is more difficult for the chondrocyte to replicate if severe damage to the collagen network occurs. During initial stages of OA, chondrocytes in vivo respond to structural changes in the surrounding cartilage matrix by increasing cell proliferation and synthesis of matrix proteins, proteinases, and anabolic and catabolic factors. The aberrant behavior of OA chondrocytes is reflected in the appearance of fibrillations; matrix depletion; cell clusters; and changes in quantity, distribution, or composition of matrix proteins. Evidence of phenotypic modulation is reflected in increases in collagens I and III and the appearance of the hypertrophic chondrocyte marker, type X collagen, and other chondrocyte differentiation genes, suggesting recapitulation of a developmental program. However, evidence reveals compensatory increases in type II collagen synthesis in deeper regions of the articular cartilage. Genomic and proteomic analyses of global gene expression in cartilage have confirmed increased COL2A1 mRNA levels in early OA cartilage. Increased levels of factors such as BMP-2 and inhibin βA/activin, members of the TGF-β superfamily, and prostaglandins suggest the presence of anabolic responses. Nevertheless, Aigner and co-workers have shown that COL2A1 expression is suppressed in upper zones of OA cartilage with progressing matrix destruction, whereas global COL2A1 gene

![Figure 3-3](image-url) Schematic representation of cellular phenotypes associated with developmental fates during condensation, chondrogenesis, chondrocyte proliferation, differentiation, and hypertrophy. Some of the regulatory factors active at different stages are listed to the left of the arrows. The major extracellular matrix genes are listed to the right of each cell type in which they are differentially expressed. BMP, bone morphogenetic protein; FGF, fibroblast growth factor; Hh, hedgehog; IGF, insulin-like growth factor; IL-1, interleukin-1; MMP, matrix metalloproteinase; PTHrP, parathyroid hormone–related protein; TGF-β, transforming growth factor-β; VEGF, vascular endothelial growth factor; Wnt, wingless type.)
expression is increased in late-stage OA cartilage compared with normal and early degenerative cartilage. The capacity of the adult articular chondrocyte to regenerate the normal cartilage matrix architecture is limited; the damage becomes irreversible, unless the destructive process is interrupted.

**CULTURE MODELS FOR STUDYING CHONDROCYTE METABOLISM**

Primary cultures of articular chondrocytes isolated from various animal and human sources have served as useful models for studying the mechanisms controlling responses to growth factors and cytokines.\(^{122-125}\) A challenge to researchers in cartilage biology is the maintenance of chondrocyte morphology and cartilage-specific gene expression during in vitro studies. In confluent primary monolayer cultures, chondrocytes maintain a rounded, polygonal morphology (Figure 3-4), but progressive loss of cartilage phenotype is seen with passage of time and after subculture. High-density cultures maintain the gene expression and synthesis of cartilage-specific matrix proteins until they are subcultured, although gene expression of type II collagen is generally more labile than that of aggrecan. During this loss of phenotype or dedifferentiation, chondrocytes lose the rounded, polygonal morphology and express some, but not all, characteristics of the fibroblast phenotype, such as type I collagen. It is possible to expand cultures through a limited number of subcultures and to “redifferentiate” the cells in three-dimensional culture systems, in which chondrocytes regain morphology; cessation of proliferation is associated with increased expression of cartilage-specific matrix proteins. Alternatively, explant cultures of articular cartilage in which chondrocytes remain encased within their own ECM have been used as in vitro models to study cartilage biochemistry and metabolism, as described in the following section.

**Articular Chondrocytes**

**Cartilage Explant (Organ) Cultures**

Based on the pioneering work of Fell,\(^ {126}\) who showed that it was possible to maintain pieces of cartilage in culture, the explant culture system was developed to characterize chondrocyte function in cartilage from various species, including humans, at different ages. Early work in bovine cartilage established the mechanisms of biosynthesis of cartilage proteoglycans under the influence of different serum concentrations and determined the turnover rate whereby the chondrocyte could maintain the balance between anabolic and catabolic pathways.\(^ {127}\) Methods developed for measuring the proteoglycan content in cartilage and incorporation of \(^ {35}S\)-sulfate into newly synthesized proteoglycans are used widely as standard assays for assessing cartilage metabolism.\(^ {128}\) Cartilage organ cultures also maintain constant levels of type II collagen during several weeks of culture, along with the characteristic morphology and banding pattern of collagen fibrils. These cultures have been useful for studying the regulation of cartilage matrix synthesis and degradation by proteases, inflammatory cytokines, retinoic acid, and anabolic growth factors.

**Monolayer Cultures**

Primary monolayer cultures of chondrocytes isolated from young animals that maintain the cartilage-specific phenotype at least throughout primary culture are easily obtained and have been used widely to assess differentiated chondrocyte functions. When chondrocytes are isolated from their matrix and cultured in monolayer, they adhere to the culture dish and readily respond to serum growth factors that stimulate proliferation of normally quiescent cells. Freshly isolated human articular or costal chondrocytes express cartilage-specific type II collagen and continue to do so for several days to weeks in primary monolayer culture.\(^ {129,130}\) In addition to cartilage-specific collagens and aggrecan, chondromodulin and protein S-100 are useful markers expressed in primary chondrocyte cultures.\(^ {131-133}\) Identification of cell surface markers that determine chondrogenic capacity enables the enrichment of subpopulations for further characterization.\(^ {134}\)

Early attempts to culture chondrocytes from various animal and human sources were frustrated by the tendency of these cells to acquire a fibroblast-like morphology associated with the appearance of type I collagen synthesis.\(^ {3,124}\) When plated at high density, the cells maintain a polygonal, although flattened, morphology. At low plating densities,
with prolonged culture, and on expansion in serial subculture, the cells gradually assume a more elongated, “fibroblast-like” morphology. Early work suggested that this change in morphology is associated with loss of phenotype, whereby the synthesis of cartilage-specific matrix molecules, such as type II collagen and aggrecan, decreases or disappears. This “dedifferentiated phenotype,” which has been described so far only in vitro, is marked by the appearance of synthesis of type I and type III collagens, and it can be accelerated by plating the cells at low densities or by treating with cytokines such as interleukin (IL)-1 or retinoic acid. Lack of correlation between cell shape and chondrocyte phenotype has been reported. Dedifferentiation of chondrocytes in monolayer culture seems to be associated, however, with increased expression of genes involved in cell proliferation, such as cyclin D1. The substrate on which chondrocytes are plated can influence the differentiation capacity of articular chondrocytes.

The use of chondrocytes of adult human origin in studies related to the pathogenesis of joint disease has been problematic because the source of the cartilage cannot be controlled, sufficient numbers of cells are not readily obtained through random operative procedures, and the phenotypic stability of adult human chondrocytes is lost more quickly on expansion in serial monolayer cultures than in cells of juvenile human or embryonic or postnatal animal origin. Serum-free defined media of varying compositions, but usually including insulin, have been used, frequently in combination with monolayer and other culture systems, as mentioned in the following section.

Three-Dimensional Culture Systems

Early studies showed that phenotype could be maintained if isolated chondrocytes are placed in suspension cultures in spinner flasks or in dishes coated with nonadherent substrates. Freshly isolated or subcultured chondrocytes also can be embedded in three-dimensional matrices, such as collagen gels or sponges, agarose, or alginate. In these three-dimensional matrices, chondrocytes have a normal spherical shape, synthesize and secrete abundant cell-associated ECM components, and may maintain phenotypic stability for several months. Because articular chondrocytes are unable to proliferate in suspension or three-dimensional culture, expansion in monolayer culture followed by transfer to alginate culture, for example, has been used as a strategy to obtain sufficient numbers of differentiated chondrocytes for study. After prolonged culture in monolayer, however, dedifferentiated chondrocytes may lose irreversibly their chondrogenic potential. The high-density pellet culture system, originally developed to study growth plate hypertrophy, has been used as a three-dimensional model because it permits articular chondrocytes to deposit a well-organized ECM containing type II collagen and aggrecan. Articular chondrocytes in micromass culture show phenotypic plasticity comparable with that of mesenchymal stem cells undergoing chondrogenesis, by recapitulating aspects of chondrocyte hypertrophy. Isolated chondrons containing one or more chondrocytes within a capsule of pericellular matrix have been used for in vitro studies of chondrocyte metabolism within a three-dimensional environment.

Prehypertrophic and Hypertrophic Chondrocytes: Models of the Growth Plate and Terminal Differentiation

Tissues or cells from embryonic or young animals at specific developmental stages or with different developmental fates have been used widely to recapitulate in vitro the transitional stages of chondrogenesis, chondrocyte hypertrophy, and endochondral ossification. A common feature of these models is the requirement for deposition of a collagenous matrix by sufficient numbers of cells after cessation of proliferation of chondrogenic cells in high-density micromass, pellet, or three-dimensional matrix cultures. Epiphyseal chondrocytes isolated from the long bones of postnatal immature rats and rabbits and cultured at high density progress through a differentiation pathway that mimics the transition from a type II collagen–producing, proliferating chondrocyte phenotype to the terminally differentiated type X collagen–producing hypertrophic phenotype associated with growth plate formation and endochondral ossification. Alkaline phosphatase, osteocalcin, and osteopontin have also been used as markers of terminal differentiation.

The pellet and micromass culture systems have been used widely to study terminal differentiation and hypertrophy because they mimic the distribution of cells within the growth plate, and they are sufficiently organized to permit calcification in situ. Arrest of cell proliferation and activation of type X collagen expression occur when the serum concentration is reduced from 10% to 2% or lower. Insulin-like growth factor (IGF)-I or insulin added in serum-free medium or as a constituent of serum seems to be a universal basal requirement in these culture systems. Ascorbic acid and treatments such as thyroxine 1,25-dihydroxy vitamin D₃, retinoic acid, or dexamethasone promote terminal differentiation in vitro. Ectopic matrix mineralization also may require a phosphate donor such as β-glycerophosphate. In contrast, certain BMPs alone or in the presence of ascorbic acid may induce hypertrophy and mineralization in the absence of other additives, if the appropriate progenitor cell population is used.

Chondrocyte Cell Lines

Because primary human chondrocytes in monolayer cultures maintain phenotype only until they are passaged, researchers have attempted to develop chondrocyte cell lines with variable success. Immortalization of chondrocytes from mice, rats, and other species with viral oncogenes has generated cell lines with high proliferative capacities and at least some differentiated chondrocyte properties. Attempts to generate cell lines that can undergo chondrogenesis and terminal differentiation to the hypertrophic phenotype have been more successful because of the availability and plasticity of progenitor cell populations. Cells derived from the ribs of transgenic mice harboring the temperature-sensitive mutant of simian virus 40 (SV40) large T antigen (Tag) are able to undergo hypertrophic differentiation, and bone marrow–derived mesenchymal stem cells derived from these mice contain osteogenic, adipogenic, myogenic, and chondrogenic progenitor cells. Several chondrogenic cell lines from mice or rats, including
ATDC5, C3H10T1/2, RCJ3.1, CK2, and Cl, are now used widely in the field, as reviewed by Johnstone and colleagues.\textsuperscript{156,157}

Human chondrosarcoma cell lines express some aspects of the chondrocyte phenotype but are tumorigenic.\textsuperscript{156,157} Stable transduction of SV40-TAG using plasmid or retroviral vectors has yielded immortalized human chondrocyte cell lines that are useful for studying the regulation of gene expression under defined conditions, but they do not produce sufficient quantities of matrix proteins to form cartilage matrix owing to high rates of proliferation.\textsuperscript{158} Human articular chondrocyte cell lines also have been established using temperature-sensitive SV40-TAg,\textsuperscript{159} the human papillomavirus type 16 early function genes E6 and E7,\textsuperscript{160} and telomerase,\textsuperscript{161} although those derived from adult tissues have low proliferative capacity. A general observation is that phenotypic stability of immortalized chondrocytes is lost during serial subculture in monolayer but can be restored by transfer to three-dimensional culture in alginate\textsuperscript{159} or hyaluronan,\textsuperscript{162} or to suspension culture in poly-2-hydroxyethyl methacrylate–coated dishes.\textsuperscript{161} Immortalized chondrocytes may continue to proliferate in three-dimensional culture, however, and if the scaffold cannot be remodeled, necrotic clusters form. Studies to date indicate that mature chondrocytes removed from the ideal cartilage environment in vivo are incapable of replicating normal phenotype in vitro; the perfect chondrocyte culture model that reproduces the normal cartilage environment has yet to be fabricated, as confirmed by gene profiling studies.\textsuperscript{162,163}

\section*{INTERACTIONS OF CHONDROCYTES WITH THE EXTRACELLULAR MATRIX}

Chondrocytes in vivo respond to structural changes in the cartilage ECM. The ECM not only provides a framework for chondrocytes suspended within it, but its constituents interact with cell surface receptors and provide signals that regulate many chondrocyte functions.\textsuperscript{158,164,165}

\subsection*{Integrins}

The most prominent of the ECM receptors are the integrins, which bind specifically with different cartilage matrix components and induce the formation of intracellular signaling complexes that regulate cell proliferation, differentiation, survival, and matrix remodeling. Integrins also may serve as mechanoreceptors and may mediate responses to normal and abnormal loading of cartilage.\textsuperscript{166,168} Chondrocytes express many different integrins that interact with cartilage ECM ligands, although most are not specific to this cell type.\textsuperscript{158,159,160,161,162,163} They include integrins that are receptors for collagen (\(\alpha\beta_1\), \(\alpha\beta_2\), \(\alpha\beta_3\), \(\alpha\beta_4\), \(\alpha\beta_5\), and \(\alpha\beta_6\)), fibronectin (\(\alpha\beta_1\), \(\alpha\beta_3\), \(\alpha\beta_5\), \(\alpha\beta_6\)), and laminin (\(\alpha\beta_1\). The integrin \(\alpha\beta_1\) has broader ligand specificity than the other collagen-binding integrins and mediates chondrocyte adhesion to pericellular type VI collagen and to the cartilage matrix protein, matrilin-1.\textsuperscript{158,159,160,164,165} The \(\alpha\beta_1\) integrin also binds to chondroadherin. The \(\alpha\beta_1\)-containing integrins bind to vitronectin and osteopontin, in addition to serving as alternative fibronectin receptors. The \(\alpha\beta_1\) and \(\alpha\beta_3\) integrins serve as receptors for different conformations of COMP.\textsuperscript{170}

Because \(\alpha\beta_1\), \(\alpha\beta_2\), and \(\alpha\beta_10\) are receptors for cartilage-specific type II collagen, there is great interest in determining whether they mediate differential responses of chondrocytes to changes in the ECM resulting from normal loading or pathologic changes.\textsuperscript{166-168} The \(\alpha\beta_1\) integrin is the prominent integrin in human adult articular cartilage. Depending on the method of analysis, adult chondrocytes also express \(\alpha\beta_1\) and \(\alpha\beta_5\) integrins accompanied by weaker expression of \(\alpha\beta_1\) and \(\alpha\beta_3\). Normal adult articular chondrocytes express little or no \(\alpha\beta_2\), whereas expression of \(\alpha\beta_5\) and \(\alpha\beta_10\) integrins is associated with a proliferative phenotype, as in fetal chondrocytes and in chondrosarcoma and chondrocyte cell lines.\textsuperscript{171,172} In growth plate chondrocytes, \(\alpha\beta_1\), \(\alpha\beta_5\), and \(\alpha\beta_10\) are important for joint formation, chondrocyte proliferation, hypertrophy, and survival.\textsuperscript{173-177} Knockout of \(\beta_1\) integrin results in severe growth plate abnormalities and chondrodysplasia,\textsuperscript{178} whereas \(\alpha_1\) integrin knockout mice develop OA without growth plate abnormalities.\textsuperscript{179}

Cellular binding to immobilized ECM proteins or integrin receptor aggregation with activating antibodies can promote numerous intracellular signaling events.\textsuperscript{180} As in other cell types, integrin signaling is mediated by interaction with intracellular protein tyrosine kinases, such as pp125 focal adhesion kinase (FAK) and pyk2Pyk2, which interact with the integrin cytoplasmic tail and induce a conformational change in the receptor subunits. Changes in organization of the cytoskeleton are associated with the formation of integrin signaling complexes, which contain scaffolding proteins such as talin, paxillin, and \(\alpha\)-actinin, in addition to FAK and the integrin-linked kinase (ILK). Mice lacking ILK in cartilage display chondrodysplasia, a phenotype similar to that of cartilage-specific \(\beta_1\) integrin knockout mice.\textsuperscript{178,180} Other signaling kinases, such as Src, Ras/Raf, Sos, and Mek family members, may be associated with integrin signaling complexes and mediate downstream signaling cascades in a cell type–specific and ligand-specific manner.

Cooperative signaling among integrins and growth factors is a fundamental mechanism in the regulation of cellular functions. Integrin aggregation and receptor occupancy enhance phosphorylation of growth factor receptors and activation of mitogen-activated protein kinases (MAPKs) in many cell types. The anchorage-dependent mitogenic response to growth factors is thought to be due to synergy between integrin and growth factor signaling. Induction of chondrocyte proliferation by FGF requires fibronectin binding to \(\alpha\beta_1\) integrin.\textsuperscript{175} The \(\beta_1\) integrin subunit also interacts with the IGF-I receptor on treatment of chondrocytes with IGF-I.\textsuperscript{181} Fibronectin increases FGF-stimulated and IGF-I–stimulated proteoglycan synthesis in chondrocytes, and nitric oxide, which disrupts focal adhesion signaling complexes, inhibits this process.\textsuperscript{182} Type II collagen increases TGF-\(\beta\)–induced type II collagen and aggrecan gene expression through a feedback mechanism that is mediated by \(\beta_1\) integrin.\textsuperscript{183}

The primary fibronectin receptor, \(\alpha\beta_1\), may play a role in cartilage degradation by binding to fibronectin fragments that upregulate matrix metalloproteinases (MMPs), such as MMP-3 and MMP-13.\textsuperscript{184-186} Extensive studies in cultured chondrocytes have shown that \(\alpha\beta_1\) is crucial in the hyperpolarization response to mechanical load.\textsuperscript{187}
chondrocytes use \( \alpha 5 \beta 1 \) as a mechanoreceptor, and subsequent to activation of the integrin-signaling cascade by mechanical stimulation is secretion of IL-4, which acts in an autocrine manner via the janus activating kinase (JAK)/signal transducer and activator of transcription (STAT) pathway to increase aggrecan mRNA and decrease MMP-3 mRNA levels.\(^{187} \)

Chondrocyte adhesion to fibronectin or binding to fibronectin fragments increases the production of cytokines, such as IL-1\( \beta \), tumor necrosis factor (TNF), IL-6, and granulocyte-macrophage colony-stimulating factor. Synergies between fibronectin/\( \alpha 5 \beta 1 \) and IL-1\( \beta \) have been shown in chondrocytes.\(^{188,189} \) OA chondrocytes respond to \( \alpha 5 \beta 1 \) ligation by production of IL-1\( \beta \) and other proinflammatory mediators, whereas \( \alpha 2 \beta 3 \) integrin ligation attenuates these responses.\(^{188} \) Nevertheless, good evidence suggests that fibronectin fragments or blocking antibodies to \( \alpha 2 \beta 1 \) and \( \alpha 5 \beta 1 \) integrins can directly stimulate signaling via extracellular signal-regulated kinase (ERK)-1 or ERK-2, c-Jun N-terminal kinase (JNK), and p38 MAPK in chondrocytes and can increase MMP-13 production independent of autocrine production of IL-1\( \beta \).\(^{185} \) This response requires reactive oxygen species.\(^{190} \) Collagen binding to \( \alpha 1 \beta 1 \) and \( \alpha 2 \beta 1 \) integrins also results in activation of distinct signaling pathways and may lead to opposite cellular responses.\(^{191} \) Downregulation of IL-1\( \beta \)-induced responses by dynamic compression is mediated by integrins.\(^{192} \) The specificity of the response may depend on the relative expression of \( \alpha \)-integrin subunits on the chondrocyte cell surface.

**Other Cell Surface Receptors in Chondrocytes**

Other integral membrane proteins found in chondrocytes include cell determinant 44 (CD44), anchorin CII, and syndecans.\(^{205} \) CD44, a receptor for hyaluronan, binds collagen and fibronectin. Through specific interactions with hyaluronan, CD44 has a role in assembly, organization, and maintenance of the chondrocyte pericellular matrix.\(^{194} \) In chondrocyte cultures, assembly of a newly synthesized pericellular matrix can be prevented or reversed by incubation with hyaluronan hexasaccharides or with a CD44 monoclonal antibody. CD44 expression is upregulated in chondrocytes in articular cartilage from RA patients and in experimental OA.\(^{195} \) Hyaluronan binding to CD44 increases MMP-13 and nitric oxide production by chondrocytes.\(^{196} \) Although blocking of CD44 has no effect on attachment of chondrocytes to a cut cartilage surface, recent evidence indicates that CD44 mediates the expression of MMP-1, MMP-2, MMP-9, and MMP-13; MMP-specific cleavage of type II collagen; and the production of nitric oxide induced by the heparin-binding fibronectin fragment in articular cartilage.\(^{195,196} \) Because fibronectin fragments and IL-1\( \beta \) enhance CD44 expression in chondrocytes, cell-matrix interactions mediated by such cell surface receptors represent alternative mechanisms for cartilage damage in joint disease. During joint disease, CD44-mediated co-localization with hyaluronic acid may be one important mechanism for the elimination of residual aggrecan fragments after extracellular degradation.\(^{197} \) Aggrecanase-mediated depletion of proteoglycan does not require CD44, however.\(^{200} \)

Anchorin V, also known as annexin CII, is a 34 kD integral membrane protein that binds type II collagen and shares extensive homology with the calcium-binding proteins calpactin and lipocortin.\(^{201,202} \) Annexins II, V, and VI have been detected in chondrocytes, where they likely play roles in physiologic mineralization of skeletal tissues and in pathologic mineralization of articular cartilage.\(^{203} \) Annexin V was first detected in chick cartilage and was described as a type II collagen–binding protein that anchors chondrocytes to the ECM. In growth plate chondrocytes, annexins are required for calcium ion uptake and subsequent mineralization. Annexin V antibodies block chondrocyte attachment to immobilized type II collagen more effectively than integrin antibodies, but not to a cut cartilage surface, where the N-terminal collagen-binding site may not be exposed.\(^{193} \) In contrast to integrins, annexin V binds to the N-telopeptide of type II collagen, but not to triple-helical fragments.\(^{204} \)

Syndecans have important roles during cartilage development and homeostasis. Syndecans link to the cell surface via glycosyl phosphatidylinositol and bind growth factors, proteases and inhibitors, and matrix molecules through heparan sulfate side chains on the extracellular domain. Syndecan-1, syndecan-3, and syndecan-4 are upregulated in human and mouse OA.\(^{205,207} \) In contrast to integrins, which bind collagen fragments, discoidin domain receptor 2 (DDR2) binds specifically to type II and X collagen fibrils, leading to activation of its integral receptor tyrosine kinase.\(^{208,209} \) DDR2 is upregulated in OA cartilage and induces specifically the expression of MMP-13 associated with cleavage of type II collagen.\(^{210} \) Studies have shown that the serine proteinase, high temperature requirement A1 (HTRA1), which is increased in the articular cartilage of mouse models of OA and human OA cartilage,\(^{211-213} \) is responsible for disrupting the pericellular matrix, which is composed of matrilin-3, fibronectin, biglycan, fibromodulin, COMP, and collagen VI, thereby exposing DDR2 to activation by type II collagen in fibrillar form.\(^{213} \)

**ANGIOGENIC AND ANTIANGIOGENIC FACTORS**

Adult articular cartilage is among the few avascular tissues in mammalian organisms; this property and the presence of angiogenesis inhibitors make it resistant to vascular angiogenesis and invasion by inflammatory and neoplastic cells. In conditions in which extensive remodeling of ECM occurs, as in arthritis, the cartilage becomes susceptible to invasion by vascular endothelial and mesenchymal cells from the synovium and subchondral bone.\(^{214} \) In OA, upregulation of angiogenic factors may contribute to ingrowth of blood vessels, tidemark advancement, and cartilage calcification in the deep zone. In RA, ingrowth of blood vessels and synovial pannus into cartilage contributes to degradation of the cartilage matrix. Troponin I, MMP inhibitors, chondromodulin-I, and endostatin, a 20 kD proteolytic fragment of type XVIII collagen, all function as endogenous angiogenic inhibitors.\(^{215-218} \) VEGF, which is an essential mediator of angiogenesis during endochondral ossification (see Chapter 1), is induced by hypoxia and
mechanical overload. In OA, in which abnormal biomechanics and joint effusions cause severe hypoxia, chondrocytes may produce VEGF, inducing angiogenesis at the chondro-osseous junction and contributing to cartilage destruction. \cite{219,220} Intercellular adhesion molecules also contribute to angiogenesis. These include vascular cell adhesion molecule-1 (VCAM-1) and intercellular adhesion molecule-1 (ICAM-1), which are expressed by human articular chondrocytes and synovial and endothelial cells. Their function on chondrocytes may not be significant, however, unless damage to the matrix permits cell–cell interactions. \cite{221} VCAM-1, VEGF, FGE, and TNF contribute to angiogenesis during synovitis and to activation of chondrocytes during cartilage degradation. \cite{222,223} In RA, VEGF expression may be upregulated by inflammatory cytokines in chondrocytes and synovial cells and by hypoxia. \cite{224,225} The importance of this mechanism is supported by findings in Vegfb knockout mice, which are protected against synovial angiogenesis in experimental inflammatory models. \cite{226}

**ROLES OF GROWTH AND DIFFERENTIATION (ANABOLIC) FACTORS IN NORMAL CARTILAGE METABOLISM**

Growth and differentiation factors generally are considered positive regulators of homeostasis of mature articular cartilage because of their capacity to stimulate chondrocyte anabolic activity and, in some cases, inhibit catabolic activity (see Chapter 1). \cite{227} The best characterized anabolic factors in the context of their production and action in articular cartilage include IGF-I and members of the FGF and TGF-β/BMP families. \cite{228-232} Many of these factors also regulate chondrogenesis and endochondral ossification during skeletal development. In adult cartilage, their expression declines with age—a risk factor for OA—and their activities are downregulated. \cite{227}

**Insulin-like Growth Factor**

IGF-I, also known as somatomedin C, was first discovered as a serum factor controlling sulfate incorporation by articular cartilage in vitro and was later found to have the specific capacity to stimulate or maintain chondrocyte phenotype in vitro by promoting the synthesis of type II collagen and aggrecan. IGF-I is a competence factor for cell proliferation that is categorized more appropriately as a differentiation factor because its limited mitogenic activity seems to depend on the presence of other growth factors, such as FGF-2, a progression factor. \cite{233} IGF-I is considered an essential mediator of cartilage homeostasis through its capacity to stimulate proteoglycan synthesis, promote chondrocyte survival, and oppose the activities of catabolic cytokines in cooperation with other anabolic factors such as BMP-7. \cite{234-236} IGF-I and insulin can activate the cell surface IGF-I tyrosine kinase receptor or the type I insulin receptor at concentrations proportional to their binding affinities. \cite{227}

Specific IGF-binding proteins (IGFBPs) that do not recognize insulin also regulate IGF-I activity. Chondrocytes at different stages of differentiation express IGF-1 and IGF receptors and different arrays of IGFBPs, providing a unique system by which IGF-I can exert different regulatory effects on these cells. IGFBP-2 seems to be a positive regulator in chondrocytes because its induction by TGF-β or estrogen is associated with increased proteoglycan synthesis. \cite{237} Binding of IGFBP-3 to IGF-I is thought to regulate negatively the anabolic functions of IGF-1, although IGFBP-3 may directly inhibit chondrocyte proliferation in an IGF-independent manner.

In OA cartilage, the normal anabolic function of IGF-I may be disrupted because chondrocytes from animals with experimental arthritis and from patients with OA are hyporesponsive to IGF-I, despite normal or increased IGF-I receptor levels. This hyporesponsiveness has been attributed to increased levels of IGFBPs that may interfere with IGF-I actions. \cite{238,239} Disturbances in the balance of IGF-I to IGFBPs that have been reported in OA and RA joints may contribute to defective chondrocyte responses to IGF-I. \cite{240-246} Small-molecule inhibitors of IGF-I/IGFBP interactions that could restore IGF-I–dependent proteoglycan synthesis in cartilage have been proposed for treatment of OA. \cite{244} Although IGF-I can oppose the effects of inflammatory cytokines that promote cartilage degradation and inhibit proteoglycan synthesis, \cite{247-249} these cytokines increase the production of IGFBP-3 by chondrocytes. \cite{250} Overproduction of nitric oxide may contribute to IGF-I resistance by chondrocytes through disruption of integrin signaling, reduction of phosphorylation of the IGF-I receptor, stimulation of cyclic guanosine monophosphate production, or suppression of mitochondrial oxidative metabolism. \cite{251-253} More recent evidence indicates that suppressor of cytokine signaling 3 (SOCS3) acts as a negative feedback regulator during IGF-I desensitization in the absence of nitric oxide by inhibiting insulin receptor substrate-1 phosphorylation. \cite{254}

**Fibroblast Growth Factor**

Members of the FGF family, including FGF-2, FGF-4, FGF-8, FGF-9, FGF-10, and FGF-18, together with the FGF receptors, FGR1, FGR2, and FGR3, coordinate patterning and cell proliferation during chondrogenesis and endochondral ossification in embryonic and postnatal growth plates. \cite{255} The most extensively studied is FGF-2, or basic FGF, which is a potent mitogen for adult articular chondrocytes, but findings on its effects on the synthesis of cartilage matrix are contradictory, showing stimulation, inhibition, or no effect on proteoglycan synthesis. \cite{256-258} Early studies suggested that low concentrations of FGF-2 could stimulate chondrocyte mitogenesis and proteoglycan synthesis, whereas high concentrations might have opposite effects. \cite{258} More recent studies showing that FGF-2 stored in the adult cartilage matrix is released with mechanical injury or with loading suggest a mechanism for modulating chondrocyte proliferation and anabolic activity. \cite{259,260} Although FGF-2 and FGF-9 stimulate the expression of Sox9 and increase the activity of the Sox9-dependent, chondrocyte-specific enhancer in the type II collagen gene, \cite{261,262} FGF-2 can inhibit the anabolic activities of IGF-I and osteogenic protein-1 (OP-1) in vitro. \cite{263} FGF-9 and FGF-18 increase matrix synthesis by mature chondrocytes. \cite{264,265} A study
showed that FGF-18 promotes cartilage repair in a rat meniscal tear model of OA.\textsuperscript{259} FGF-2 stimulates, whereas IGF-I inhibits, expression of the matrix Gla protein, which is a marker of chondrocyte survival during endochondral ossification.\textsuperscript{260,261} Therefore FGFs and FGF receptors are important regulators of cartilage homeostasis during prenatal and postnatal life.

**TGF-β/BMP Superfamily**

Activities of the TGF-β/BMP superfamily in the skeleton were first discovered by Marshall Urist as constituents of demineralized bone that induced new bone formation when implanted into extraskeletal sites in rodents.\textsuperscript{262} These bioactive morphogens subsequently were extracted, purified, and cloned and were found to regulate the early commitment of mesenchymal cells to chondrogenic and osteogenic lineages during cartilage development and endochondral bone formation (Table 3-2). The TGF-β/BMP superfamily includes activins, inhibins, mullerin duct inhibitory substance, and nodal, glial-derived neurotrophic factor, OP-1 (or BMP-7), and growth differentiation factors (GDFs), also called cartilage-derived morphogenetic proteins (CDMPs).\textsuperscript{263} In addition to regulating cartilage condensation and chondrocyte differentiation, members of this superfamily play key roles in site specification and cavitation of synovial joints (see Chapter 1) and in the development of other organ systems. Many of these factors, including BMP-2, BMP-6, BMP-7, BMP-9, TGF-β, and CDMP-1, are able to induce chondrogenic differentiation of mesenchymal progenitor cells in vitro. They also may have direct effects on mature articular chondrocytes in vivo and in vitro.

**Table 3-2  Bone Morphogenetic Protein Superfamily**

<table>
<thead>
<tr>
<th>Bone Morphogenetic Protein</th>
<th>Other Names</th>
<th>Potential Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP-2</td>
<td>BMP-2A</td>
<td>Cartilage and bone morphogenesis</td>
</tr>
<tr>
<td>BMP-3</td>
<td>Osteogenin, GDF-10</td>
<td>Bone formation</td>
</tr>
<tr>
<td>BMP-4</td>
<td>BMP-2B</td>
<td>Cartilage and bone morphogenesis</td>
</tr>
<tr>
<td>BMP-5</td>
<td>Vegetal-related-1 (Vgr-1)</td>
<td>Bone morphogenesis</td>
</tr>
<tr>
<td>BMP-6</td>
<td>Osteogenin protein-1 (OP-1)</td>
<td>Cartilage hypertrophy</td>
</tr>
<tr>
<td>BMP-7</td>
<td>Osteogenin protein-2 (OP-2)</td>
<td>Cartilage and bone morphogenesis</td>
</tr>
<tr>
<td>BMP-8</td>
<td>GDF-2</td>
<td>Bone morphogenesis</td>
</tr>
<tr>
<td>BMP-9</td>
<td>GDF-11</td>
<td>Cartilage morphogenesis</td>
</tr>
<tr>
<td>BMP-10</td>
<td>GDF-7, CDMP-3</td>
<td>Cartilage morphogenesis</td>
</tr>
<tr>
<td>BMP-11</td>
<td>GDF-6, CDMP-2</td>
<td>Cartilage morphogenesis</td>
</tr>
<tr>
<td>BMP-12</td>
<td>GDF-5, CDMP-1</td>
<td>Cartilage morphogenesis</td>
</tr>
</tbody>
</table>

CDMP, cartilage-derived morphogenetic protein; GDF, growth and differentiation factor.

**Transforming Growth Factor-β**

TGF-β was named on the basis of its discovery as a factor that could transform cells to grow in soft agar. TGF-β is not a potent inducer of chondrocyte proliferation, however; rather, it controls early mesenchymal cell condensation, as well as chondrocyte differentiation at early and late stages of chondrogenesis and endochondral ossification (see Chapter 1). Both inhibition and stimulation of the synthesis of aggrecan and type II collagen by TGF-β have been observed in vitro. TGF-β, by itself, cannot rescue the type II collagen phenotype, however, when cells have undergone dedifferentiation during serial passaging. Levels of TGF-β measured in synovial fluids of OA and RA patients may reflect anabolic processes in cartilage and other joint tissues. TGF-β may promote anabolism by inducing the expression of tissue inhibitors of MMP (TIMP).

Nevertheless, TGF-β1, TGF-β2, and TGF-β3 generally are considered as potent stimulators of proteoglycan and type II collagen synthesis in primary chondrocytes and cartilage explants in vitro.\textsuperscript{264} Microarray analysis of chondrocytes in vitro indicates that TGF-β is able to counteract the expression of numerous IL-1–induced genes involved in cartilage injury.\textsuperscript{265} Although intra-articular injection of TGF-β stimulates proteoglycan synthesis and limits cartilage damage in inflammatory arthritis models, injection or adenovirus-mediated delivery of TGF-β1 may result in side effects in joint tissues, such as osteophyte formation, swelling, and synovial hyperplasia.\textsuperscript{266,267} Administration of agents that block TGF-β activity, such as the soluble form of TGF-βRII, inhibitory SMADs, or the physiologic antagonist, latency-associated peptide-1, increases proteoglycan loss and cartilage damage in an experimental model of OA.\textsuperscript{268,269} A more recent finding that TGF-β induces expression of ADAMTS (a disintegrin and metalloproteinase [ADAM family] with thrombospondin-1 domains)-4 in primary human chondrocytes and promotes the degradation of aggrecan suggests that it may be involved in normal turnover of proteoglycans in mature cartilage.\textsuperscript{270} Findings that IL-1 differentially regulates inhibitory SMADs and transcriptional mediators of TGF-β and BMP signaling,\textsuperscript{271} and decreases TGF-β signaling associated with loss of a protective effect of TGF-β during OA progression,\textsuperscript{272} suggest that the balance of inhibitory and stimulatory molecules determines cartilage homeostasis.

**Bone Morphogenetic Proteins**

BMPs constitute a large subclass of the TGF-β superfamily essential for normal appendicular skeletal and joint development.\textsuperscript{128,260,273} The isolation and cloning of the first BMP family members from bone prompted a search for cartilage-derived BMPs, or CDMPs—CDMP-1, CDMP-2, and CDMP-3—which are classified as GDF-5, GDF-6, and GDF-7. The BMPs may be divided into four distinct subfamilies based on the similarity of primary amino acid sequences: (1) BMP-2 and BMP-2B (BMP-4), which are 92% identical in the 7-cysteine region; (2) BMP-3 (osteogenin) and BMP-3B (GDF-10); (3) BMP-5, BMP-6, BMP-7 (OP-1), BMP-8 (OP-2), BMP-9 (GDF-2), BMP-10, and BMP-11 (GDF-11); and (4) BMP-12 (GDF-7 or CDMP-3), BMP-13 (GDF-6 or CDMP-2), BMP-14 (GDF-5 or
CDMP-1), and BMP-15. BMP-1 is not a member of this family but is an astacin-related MMP that cleaves the BMP inhibitor chordin and acts as a procollagen C-proteinase.

Several BMPs, including BMP-2, BMP-7 (OP-1), and GDF-5/CDMP-1, can stimulate differentiation of mesenchymal precursors into chondrocytes and promote the differentiation of hypertrophic chondrocytes. BMP-2, BMP-4, BMP-6, BMP-7, BMP-9, and BMP-13 can enhance the synthesis of type II collagen and aggrecan by articular chondrocytes in vitro. In addition, BMP-7 reverses many of the catabolic responses induced by IL-1β, including induction of MMP-1 and MMP-13, downregulation of TIMP, and downregulation of proteoglycan synthesis in primary human articular chondrocytes. CDMP-2 is found in articular cartilage, skeletal muscle, placenta, and hypertrophic chondrocytes of the epiphyseal growth plate. CDMP-1 and CDMP-2 maintain the synthesis of type II collagen and aggrecan in mature articular chondrocytes, although they are less effective initiators of chondrogenesis than other BMPs in early progenitor cell populations in vitro.

BMP-7 is expressed in mature articular cartilage and is possibly the strongest anabolic stimulus for adult chondrocytes in vitro because it increases aggrecan and type II collagen synthesis more strongly than IGF-I. BMP-2 also is expressed in normal and OA articular cartilage, and it is a molecular marker, along with type II collagen and FGF receptor 3 (FGFR3), for the capacity of adult articular chondrocyte cultures to form stable cartilage in vivo. BMP-2, BMP-7, and BMP-9 are able to oppose many of the detrimental effects of IL-1 on chondrocyte metabolism in vitro and in vivo. BMP-7 can prevent retinoic acid–induced dedifferentiation of articular chondrocytes. BMPs have pleiotropic effects in vivo, however, acting in a concentration-dependent manner. While initiating chondrogenesis in the limb bud, they generally set the stage for bone morphogenesis (see Chapter 1). Several BMPs also are true morphogens for other tissues, such as kidney, eye, heart, and skin.

Receptors, Signaling Molecules, and Antagonists That Mediate Chondrocyte Responses to Growth and Differentiation Factors

Major pathways activated by the growth and differentiation factors discussed earlier involve members of the ERK1/2, p38 MAPK, and phosphatidylinositol-3'-kinase (PI-3K)/Akt murine thymoma viral oncogene homolog (AKT) pathways. As in other cell types, FGF family members activate kinases of the ERK1/2 and p38 MAPK cascades in chondrocytes. Specific inhibitors of these pathways block FGF-2–induced and FGF-18–induced mitogenesis in chondrocytes and prevent FGF-2 induction of Sox9 in primary chondrocytes. The PI-3K pathway is required for the stimulation of proteoglycan synthesis by IGF-I in primary human articular chondrocytes, whereas the ERK1/2 pathway acts as a negative regulator.

TGF-β and BMP family members transduce signals through the formation of heteromeric complexes of ligand-specific receptors, which have serine-threonine kinase activity. The specificity of subsequent signals is determined mainly by the type I receptors. Type I and type II receptors are required for signal transduction (Figure 3-5). Seven types of type I receptors, called activin receptor–like kinases (ALKs), have been identified in mammals and have similar structures. TGF-β interacts with the type II receptor (TβRII), which recruits a TGF-β type I receptor (principally TβRI) to form a heterotrimeric receptor complex. The constitutively active TβRII kinase phosphorylates TβRI at serine and threonine residues. Three type I receptors, BMP type IA (BMPR-IA or ALK-3), BMPR-IB (ALK-6), and ALK-2, mediate BMP signaling. Although BMPR-1 receptors are able to bind ligand in the absence of type II BMP receptors, cooperativity has been shown in binding assays. On ligand binding, analogous to TβRI and TβRII, BMP type I receptors are phosphorylated by the

![Figure 3-5 Bone morphogenetic protein (BMP) receptors and signaling cascade. BMPs are dimeric ligands with a single interchain, disulfide bond. BMPs interact with type I and II BMP receptors (BMPR-I and BMPR-II). BMPR-II phosphorylates BMPR-I and activates the serine/threonine kinase receptor. The BMP-RI protein serine/threonine kinase phosphorylates cytoplasmic signaling substrates Smad 1 and Smad 5. This phosphorylation is modulated and inhibited by inhibitory Smad 6 and Smad 7. Phosphorylated Smad 1 and Smad 5 interact with a common co-Smad 4 and are translocated into the nucleus to initiate the transcription of BMP-response genes. A Smad-interacting protein (SIP) modulates binding of the Smad 1/Smad 4 complex to DNA. The bioavailability of BMP for interaction with cognate receptors depends on BMP binding proteins and antagonists, such as noggin and chordin, and extracellular matrix components, such as collagens and heparan sulfate. (From Reddi AH: Role of morphogenetic proteins in skeletal tissue engineering and regeneration, Nat Biotech 16:247–252, 1998.)](image-url)
BMP type II receptors, which include activin (Act) RII, ActRIIB, and T-ALK. Spatial and temporal differences in the distribution of these receptors in different tissues can govern the response patterns to different members of the TGF-β/BMP family.

The canonical SMAD pathway mediates TGF-β and BMP signaling through phosphorylation of receptor-activated SMADs (R-SMADs). R-SMADs include Smad 1, Smad 5, and Smad 8 induced by BMPs and Smad 2 and Smad 3 induced by TGF-β. The SMADs are related to the Drosophila mothers against decapentaplegic (MAD) and nematode SMA signaling molecules. These regulatory SMADs form complexes with the common Smad 4 and translocate to the nucleus, where they bind to SMAD-specific DNA binding sites in the promoters of target genes. Other BMP-induced transcription factors include JunB, JunD, and inhibitor of DNA binding (ID) and distal less homeo box (DLX) family members, suggesting alternative pathways of signaling. TGF-β and BMPs also can signal by activating TGF-β–activated kinase 1 (TAK1), which interacts with MEK1 and activates p38 and JNK cascades, or by activating Ras/ERK1/2 or RhoA/ROCK signaling.

Findings of numerous studies suggest that the differentiation of chondrocytes from mesenchymal precursors is positively regulated by p38 kinase and negatively regulated by ERK1/2. ERK1/2 activation cross-interacts with BMP-2–induced signaling to regulate chondrogenesis in a positive manner, whereas p38 activity is essential for the TGF-β induction of proteoglycan synthesis in articular chondrocytes. Cytoskeletal compartmentation of SMAD signaling complexes may regulate the differentiation of chondroprogenitor cells into chondrocytes.

Inhibitory Smad 6 and Smad 7, which prevent phosphorylation of the R-SMADs, also control BMP-induced and TGF-β–induced activities. An additional SMAD-interacting molecule, Tob1, negatively regulates signaling by sequestering R-SMADs and is an antiproliferative protein that is downregulated in OA cartilage. BMP antagonists play important roles in spatial and temporal regulation of BMP activities during skeletal development. Originally discovered in Xenopus, they act as antagonists by determining the bioavailability of BMPs for binding to BMP receptors. The roles of noggin and chordin seem to be crucial for determining boundaries during joint morphogenesis. They display different spatial and temporal patterns of expression, binding affinities, and susceptibility to proteinases that release BMP. BMPs bind to chordin and noggin via cysteine-rich domains that are similar to domains in the N-terminal propeptides of fibrillar procollagens I, II, III, and V, which also bind BMPs and are susceptible to cleavage by MMPs. BMP may be released from chordin by cleavage with MMPs or BMP-1/tolloid, whereas noggin binds BMP with high affinity and cannot be cleaved to release BMP.

Follistatin, gremlin, chordin, and chordin-like 2 are upregulated in OA cartilage. Follistatin, which has been linked to inflammatory processes; gremlin, which is associated with hypertrophic phenotype; and chordin appear at different stages of OA and with different topographic distribution. Because each antagonist binds preferentially to different BMPs, the differential expression may serve as a feedback mechanism to balance anabolic activities at different stages.

**ROLE OF THE CHONDROCYTE IN CARTILAGE PATHOLOGY**

The chondrocyte, the unique cell type in mature cartilage, maintains a stable equilibrium between the synthesis and the degradation of matrix components. During aging and joint diseases, such as RA and OA, this equilibrium is disrupted, and the rate of loss of collagens and proteoglycans from the matrix exceeds the rate of deposition of newly synthesized molecules. Cartilage destruction in OA is believed to be chondrocyte mediated in response to biomechanical insult, which may occur directly or indirectly through the production of cytokines and cartilage matrix–degrading proteinases in cartilage and other joint tissues (Figure 3-6).

Cartilage destruction in RA occurs primarily in areas contiguous with the proliferating synovial pannus as a result of the release and activation of proteinases from the synovial cells and, to some extent, at the cartilage surface exposed to matrix-degrading enzymes from polymorphonuclear neutrophils in the synovial fluids. In addition to the direct action of proteinases, RA synovial tissues contribute indirectly to cartilage loss by releasing cytokines and other mediators that act on the chondrocytes to produce dysregulation of chondrocyte function. Understanding of basic cellular mechanisms regulating chondrocyte responses to inflammatory cytokines has been inferred from numerous studies in vitro using cultures of cartilage fragments or isolated chondrocytes and is supported by studies in experimental models of inflammatory arthritis, such as collagen-induced arthritis and antigen-induced arthritis in mice.

Direct analysis of cartilage or chondrocytes from OA patients undergoing joint replacement has yielded more information than is available from RA patients, where cartilage damage is extensive. These studies indicate that chondrocytes produce not only proinflammatory cytokines, but also inhibitory and anabolic cytokines that modulate responses. The impact of cytokines on chondrocyte function, particularly with respect to their various roles in cartilage destruction, has been reviewed extensively. Studies of cartilage degradation in genetically modified mouse strains that are resistant to or accelerate the development of OA spontaneously with aging or when subjected to excessive biomechanical injury have yielded information about the molecular effectors of the disease and potential targets for therapy.

**Cartilage Matrix–Degrading Proteinases**

Chondrocytes synthesize and secrete MMPs in latent forms, which are activated outside the cells via activation cascades. An important cascade in cartilage is initiated by plasmin, the product of plasminogen activator activity, which may be produced by the chondrocyte; plasmin activates latent stromelysin (MMP-3), an activator of latent collagenases. In early studies, chondrocytes were among the first identified sources of TIMP-1, and they are now known to synthesize additional TIMPs. Chondrocytes are assumed to be a major source of the TIMPs and MMPs detected in synovial fluids, where they reflect an adaptive response to the local imbalance caused by increased production of
The role of chondrocyte-derived proteinases in cartilage destruction in osteoarthritis. Although studies in vitro and in vivo have shown that the chondrocyte can respond directly to mechanical loading, to catabolic cytokines such as interleukin-1 (IL-1) and tumor necrosis factor (TNF), and to cartilage breakdown products, the initiating signals and their relative importance have not been defined clearly. MMP, matrix metalloproteinase; TIMP, tissue inhibitor of metalloproteinase; uPA, urinary plasminogen activator. (From Goldring MB: Osteoarthritis and cartilage: the role of cytokines, Curr Rheumatol Rep 2:459–465, 2000.)

Although elevated levels of MMPs in RA synovial fluids likely originate from the synovium, intrinsic chondrocyte-derived chondrolytic activity is present at the cartilage-pannus junction and in deeper zones of cartilage matrix in some RA specimens. MMP-1 is expressed at lower levels than MMP-3 and MMP-13 in the RA synovial pannus but is also produced by chondrocytes. MMP-10, similar to MMP-3, activates procollagenases and is produced by the synovium and chondrocytes in response to inflammatory cytokines. MMP-14, produced principally by the synovial tissue, is important for synovial invasiveness, and antisense mRNA inhibition of this membrane proteinase has been shown to reduce cartilage destruction.

Several of the MMPs, including MMP-3, MMP-8, MMP-14, MMP-19, and MMP-20, are capable of degrading proteoglycans. Members of the replosin-related proteinases of the ADAM family, particularly ADAMTS-4 and ADAMTS-5, are now regarded as the principal mediators of aggrecan degradation. The activities of MMPs and aggrecanases are complementary, however. Of the aggrecanases, ADAMTS-5 is associated with increased susceptibility to OA, as shown in Adamts5-deficient mice. TIMP-3, but not TIMP-1, TIMP-2, or TIMP-4, is a potent inhibitor of ADAMTS-4 and ADAMTS-5 in vitro, and TIMP-3 deficiency results in mild cartilage degradation similar to that seen in patients with OA. A recent study indicates that syndecan-4 by controlling the synthesis of MMP-3 is a positive effector of ADAMTS-5 activation. Hedgehog signaling by modulating Runx2 may promote articular cartilage degradation via ADAMTS-5.

Cysteine proteinases, cathepsins B and L, and the aspartic proteinase, cathepsin D, are lysosomal enzymes that may play a secondary role in cartilage degradation via intracellular digestion of products released by other proteinases. Cathepsin B also may have a role in extracellular degradation of collagen telopeptides, collagens IX and XI, and aggrecan. Cathepsin K is expressed in synovial fibroblasts on the cartilage surface at the pannus-cartilage junction and is upregulated by inflammatory cytokines. Among the known cathepsins, cathepsin K is the only proteinase that is capable of hydrolyzing type I and type II collagens at multiple sites within the triple-helical regions, and its requirement for acidic pH may be provided by the microenvironment between the synovial pannus and the cartilage.
Table 3-3  Chondrocyte Proteinases That Mediate Degradation of Cartilage Matrix

<table>
<thead>
<tr>
<th>Proteinase Class</th>
<th>Cartilage Matrix Substrates</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Matrix Metalloproteinases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collagenases (MMP-1, MMP-8, MMP-13)</td>
<td>Collagens I, II</td>
<td>Fibrillar domain, 3/4 from N-terminus N-telopeptide (MMP-13)</td>
</tr>
<tr>
<td>Stromelysins (MMP-3, MMP-10)</td>
<td>Aggrecan core protein</td>
<td>Asn(^{341})-Phe(^{342}) IGD</td>
</tr>
<tr>
<td>Gelatinases (MMP-2, MMP-9)</td>
<td>Aggrecan core protein, Collagens IX, XI</td>
<td>Telopeptide region</td>
</tr>
<tr>
<td>Membrane-type MMPs</td>
<td>Collagens II, XI, Proteoglycans, link protein</td>
<td>Telopeptide or denatured collagen chains</td>
</tr>
<tr>
<td>Matriylsin (MMP-7)</td>
<td>Link protein</td>
<td></td>
</tr>
<tr>
<td>Enamelysin (MMP-20)</td>
<td>COMP, link protein</td>
<td></td>
</tr>
<tr>
<td><strong>Aggrecanases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADAMTS-1, ADAMTS-4, ADAMTS-5</td>
<td>Aggrecan core protein IGD</td>
<td>Glu(^{373})-Ala(^{374})-Glu(^{1819})-Ala(^{1820})</td>
</tr>
<tr>
<td>Serine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasminogen activators (tPA, uPA)</td>
<td>Aggrecan, fibronectin, proMMPs</td>
<td>Activation of plasminogen gives rise to plasmin</td>
</tr>
<tr>
<td>Cathepsin G</td>
<td>Aggrecan, collagen II, proMMPs</td>
<td></td>
</tr>
<tr>
<td>High temperature requirement A1 (HTRA1)</td>
<td>Matrilin 3, fibronectin, biglycan, fibromodulin, COMP, collagen VI</td>
<td>Degrades pericellular matrix</td>
</tr>
<tr>
<td><strong>Cysteine</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cathepsins B, K, L, S</td>
<td>Collagens IX, XI Link protein, aggrecan</td>
<td>Telopeptides (optimal pH 4.0-6.5)</td>
</tr>
<tr>
<td>Aspartate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cathepsin D</td>
<td>Phagocytosed ECM components</td>
<td>In lysosomes (optimal pH 3.0-6.0)</td>
</tr>
</tbody>
</table>

ADAMTS, a disintegrin and metalloproteinase with thrombospondin-1 domains; COMP, cartilage oligomeric matrix protein; ECM, extracellular matrix; IGD, interglobular domain; MMP, matrix metalloproteinase; MT-MMP, membrane-type MMP; proMMP, proenzyme form of MMP; TNF, tumor necrosis factor; tPA, tissue-type plasminogen activator; uPA, urokinase-type plasminogen activator.

MT1-MMP (MMP-14) also may serve as an activator of other MMPs produced by chondrocytes. Other MMPs, including MMP-16 and MMP-28, and numerous ADAM/ADAMTS family members, including ADAM-17/TACE (TNF converting enzyme), are expressed by chondrocytes, but their roles in cartilage have yet to be defined. Identification of the precise roles of these proteinases and their endogenous inhibitors in chondrocyte-mediated cartilage degradation has provided the opportunity to develop targeted therapies that interfere with the activities of aggrecanases or MMPs without disrupting normal physiology. In that regard, microRNAs are novel endogenous regulators in cartilage. MicroRNA-140 downregulates ADAMTS-5 expression in normal cartilage, and its expression is reduced in OA cartilage and is suppressed by exposing chondrocytes to IL-1β. MicroRNA-140 knockout mice are predisposed to age-related OA development, and overexpression of microRNA-140 in chondrocytes is protective against surgically induced OA.

**Balance of Cytokines in Cartilage Destruction**

Of the cytokines that affect cartilage metabolism, most are pleiotropic factors that were identified originally as immunomodulators but were found to regulate cellular functions in cells of mesenchymal origin. IL-1 and TNF not only stimulate chondrocytes to synthesize cartilage matrix-degrading proteinases, they also regulate matrix protein synthesis and cell proliferation. Considerable redundancy and overlap in biologic activities exist among the individual cytokines, and they do not act alone, but rather in synergy or partnership with or in opposition to other cytokines via cytokine networks. In addition to IL-1 and TNF, other catabolic cytokines have been characterized, as have inhibitory or anabolic cytokines produced by the chondrocytes themselves or by other cells in joint tissues (Table 3-4). Investigations in vitro and in vivo have begun to sort out the complexities of the cytokine networks and to determine how the balance in normal homeostasis can be restored when it is disrupted (Figure 3-7). Studies of type II collagen-induced arthritis and other types of induced arthritis in transgenic animals with overexpressed or deleted genes encoding cytokines, their receptors, or activators have provided further insight into the roles of these factors in cartilage destruction.

**Interleukin-1 and Tumor Necrosis Factor**

IL-1 and TNF are the predominant catabolic cytokines involved in the destruction of articular cartilage. The first
recognition of IL-1 as a regulator of chondrocyte function stems largely from the early work of Fell and others, who identified a soluble factor, termed catabolin, in supernatants of normal, noninflamed porcine synovial fragment cultures that stimulated chondrocytes to degrade the surrounding cartilage matrix. Similar activities in culture supernatants from mononuclear cells and rheumatoid synovium were attributed to IL-1, and the catabolin isoforms were identified as IL-1α and IL-1β. Since those early findings, numerous studies in vitro and in vivo indicate that IL-1 and TNF, originating primarily from the inflamed synovium, are the predominant catabolic cytokines involved in the destruction of articular cartilage in RA.

Major events in OA pathogenesis occur within the cartilage itself, and evidence suggests that chondrocytes participate in this destructive process not only by responding to cytokines released from other joint tissues but also by synthesizing them. They may be exposed continuously to the autocrine and paracrine effects of IL-1 and other inflammatory mediators at high local concentrations. Chondrocytes in OA cartilage, especially those in clonal clusters, are positive for IL-1 immunostaining and express IL-1 receptor antagonist (IL-1Ra), soluble TNF receptor, TNF, tumor necrosis factor. (Adapted from Goldring MB: Osteoarthritis and cartilage: the role of cytokines, Curr Rheumatol Rep 2:459–465, 2000.)

![Figure 3-7](image)

**Figure 3-7** The cytokine balance in cartilage metabolism. Soluble mediators toward the left of the balance promote the loss of cartilage matrix. Mediators on the right side prevent the synthesis or actions of catabolic cytokines and prevent loss of cartilage matrix. Anabolic factors, including insulin-like growth factor I (IGF-I) and bone morphogenetic proteins (BMPs), and prostaglandin E2 (PGE2) maintain or promote cartilage matrix synthesis. IL-1, interleukin-1; IL-1Ra, IL-1 receptor antagonist; LIF, lymphocyte-activating factor; OSM, oncostatin M; sTNF-R, soluble TNF receptor; TNF, tumor necrosis factor. (Adapted from Goldring MB: Osteoarthritis and cartilage: the role of cytokines, Curr Rheumatol Rep 2:459–465, 2000.)

Table 3-4  Cytokines That Regulate Cartilage Destruction

<table>
<thead>
<tr>
<th>Catabolic</th>
<th>Modulatory</th>
<th>Inhibitory</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL-1</td>
<td>IL-6</td>
<td>IL-4</td>
</tr>
<tr>
<td>Tumor necrosis factor</td>
<td>Leukemia inhibitory factor</td>
<td>IL-10</td>
</tr>
<tr>
<td>IL-17</td>
<td>Oncostatin</td>
<td>IL-13</td>
</tr>
<tr>
<td>IL-18</td>
<td>IL-11</td>
<td>IL-1 receptor antagonist</td>
</tr>
</tbody>
</table>

IL, interleukin.

The increased sensitivity of OA chondrocytes to IL-1 and TNF may be associated with increased levels of IL-1R1 and p55 TNF receptor at localized sites. Co-localization of these cytokines, MMPs, and type II collagen cleavage epitopes has been reported in regions of matrix depletion in RA cartilage.

Originally known as cachectin, TNF produces many effects on chondrocytes in vitro that are similar to those of IL-1β, including stimulation of the production of matrix-degrading proteinases and suppression of cartilage matrix synthesis. Although IL-1β is 100-fold to 1000-fold more potent on a molar basis than TNF, strong synergistic effects occur at low concentrations of the two cytokines together, eliciting more severe cartilage damage than injection of either cytokine alone. The concept that TNF drives acute inflammation, whereas IL-1β has a pivotal role in sustaining inflammation and cartilage erosion, has been derived from work in animal models of RA using cytokine-specific neutralizing antibodies, soluble receptors, or receptor antagonists and in transgenic or knockout mouse models. In a surgically induced OA model, IL-1β knockout mice are protected against cartilage damage. A more recent study showed that crossing arthritic human TNF transgenic (hTNFtg) mice with mice deficient in both IL-1α and IL-1β protects against cartilage erosion without affecting synovial inflammation.

**Cytokine Networks**

IL-1β and TNF can induce chondrocytes to produce several other proinflammatory cytokines, including IL-6, leukemia inhibitory factor, IL-17, IL-18, and chemokines. IL-6 seems to play a dual role by increasing expression of IL-1 receptor antagonist (IL-1Ra), soluble TNF receptor, and TIMPs, while also enhancing immune cell function and inflammation. The activity of IL-6 requires soluble IL-6 receptor to synergize with IL-1 to stimulate expression of MMPs and ADAMTS and to downregulate COL2A1 and aggrecan in cultured chondrocytes. IL-6 knockout mice are more susceptible to cartilage degeneration during...
aging, however, suggesting that this cytokine may play a protective role in normal physiology.

Other members of the IL-6 family that act via receptors that homodimerize with gp130 also may serve modulatory roles. IL-11 shares several actions of IL-6, including stimulation of TIMP production, without affecting MMP production by chondrocytes, and it may inhibit cartilage destruction. Leukemia inhibitory factor (LIF) participates in a positive feedback loop by increasing the production of IL-6 by chondrocytes. Oncostatin M (OSM), a product of macrophages and activated T cells, is a potent stimulator of chondrocyte production of MMPs and aggrecanases in synergy with IL-1β or TNF. Direct evidence supporting a role for OSM in contributing to cartilage loss in inflammatory arthritis is provided by studies in animal models.

IL-17 and IL-18 are potent catabolic factors that stimulate the production of IL-1β, MMPs, IL-6, inducible nitric oxide synthase (iNOS), cyclooxygenase (COX)-2, and microsomal PGE synthase-1 (mPGES-1) by human chondrocytes. IL-17 is produced by activated T helper type 1 (Th1), or CD4+, lymphocytes and binds to a receptor that is not related to any known cytokine receptor family. IL-17A, one of at least six family members, is primarily a product of Th17 cells, a newly described subset of T cells, and is a potent inducer of catabolic responses in chondrocytes by itself or in synergy with other cytokines. IL-17 can drive T cell–dependent erosive arthritis in TNF-deficient and IL-1Ra knockout mice, and treatment of collagen-induced arthritis or antigen-induced arthritis mice with neutralizing IL-17 antibody effectively inhibits cartilage destruction in these models of RA. A role for IL-17 in the promotion of angiogenesis through induction of VEGF in OA chondrocytes and synovial fibroblasts has been proposed.

IL-18 is produced by macrophages, its receptor shares homology with IL-1RI, and it has effects similar to IL-1β in human chondrocytes, but it stimulates chondrocyte apoptosis. IL-18 deficiency or blockade with IL-18–neutralizing antibody or IL-18–binding protein reduces cartilage destruction and inflammation, and IL-18 gene transfer promotes IL-1β–driven cartilage destruction in a TNF-independent manner. Of the other members of the IL-1 family identified by DNA database searches, IL-1F8 seems to be capable of stimulating IL-6, IL-8, and nitric oxide production by human chondrocytes, but at 100-fold to 1000-fold higher concentrations than IL-1β.

IL-32, a more recently discovered cytokine that induces TNF, IL-1β, IL-6, and chemokines and is expressed in the synovia of patients with RA, contributes to TNF-dependent inflammation and cartilage proteoglycan loss.

**Inhibitory Cytokines**

IL-4, IL-10, IL-13, and the naturally occurring IL-1Ra are classified as inhibitory cytokines because they decrease the production and activities of catabolic and proinflammatory cytokines in chondrocytes in vitro and suppress cartilage destruction in vivo (see Table 3-4). IL-4 and IL-10 inhibit cartilage-degrading proteinases and reverse some effects of catabolic cytokines in vitro; together they produce synergistic suppression of cartilage destruction in vivo. The efficacy of IL-4, IL-10, and IL-13 in retarding cartilage damage may be related in part to their stimulatory effects on IL-1Ra production, and their therapeutic application has been proposed as a means of restoring the cytokine balance in RA. IL-1Ra is capable of blocking the actions of IL-1 if added at sufficiently high concentrations in vitro and is among the first agents to be developed for anticytokine therapy. IL-1Ra can be produced by the same cells that secrete IL-1 and exists as at least three isoforms, including an intracellular form.

Despite the capacity of IL-4 to inhibit the effects of proinflammatory cytokines on chondrocyte function, differential effects have been observed in mice depending on the model. Gene transfer of IL-10 in combination with IL-1Ra inhibits cartilage destruction through a mechanism involving activin, a TGF-β family member. IL-10 is part of the response induced by immunomodulatory neuropetides that have been shown to inhibit inflammation and cartilage and bone destruction by downregulating the Th1-driven immune response and upregulating IL-10/TGF-β–producing T regulatory lymphocytes. IL-13 decreases the breakdown of collagen and proteoglycans by inhibiting MMP-3 and MMP-13 expression induced by a combination of IL-1α and OSM and by upregulating TIMP-1, but it increases MMP-1. Local gene transfer of IL-13 inhibits chondrocyte death and MMP-mediated cartilage degradation despite enhanced inflammation in the immune complex arthritis model. The inhibitory cytokines may have direct effects on cartilage metabolism and indirect effects by mediating the production and actions of catabolic cytokines.

**Other Mediators**

The balance of mediators determining normal homeostasis is complex, and modulation of their activities may produce positive or negative effects on chondrocyte function. In addition to inducing the synthesis of MMPs and other proteinases by chondrocytes, IL-1β and TNF upregulate the production of nitric oxide via iNOS (or NOS2) and prostaglandin E2 (PGE2) by stimulating the expression or activities of COX-2, mPGES-1, and soluble phospholipase A2. In the production of prostaglandins, mPGES-1, which is induced by IL-1β in chondrocytes, is a major player. Although PGE2 and nitric oxide have been well characterized as proinflammatory mediators, evidence suggests that they also may be protective and may play roles in chondrocyte survival and responses to mechanical stress. COX-2 is involved in the chondrocyte response to high shear stress, which is associated with reduced antioxidant capacity and increased apoptosis. The mechanisms of cross-talk between prostaglandins and nitric oxide in chondrocytes have been reviewed.

Another regulator is peroxisome proliferator-activated receptor γ (PPARγ), which is activated by the endogenous ligand 15-deoxy-Δ12,14-prostaglandin J2 (PGJ2). PPARγ activation opposes the induction of COX-2, iNOS, MMPs, and mPGES-1 and the suppression of aggrecan synthesis by IL-1. PPARα agonists may also protect chondrocytes against IL-1–induced responses by increasing the expression of IL-1Ra. The IL-1β–induced COX-2 response depends on the differentiated phenotype of chondrocytes, and PGE2...
opposes the effects of IL-1β by stimulating type II collagen and inhibiting type I collagen gene expression.\textsuperscript{391,392}

Roles for nitric oxide as a mediator of other IL-1–induced responses, including inhibition of aggrecan synthesis, enhancement of MMP activity, and reduction of IL-1Ra synthesis, also have been suggested.\textsuperscript{393} Nitric oxide may increase chondrocyte susceptibility to injury by other oxidants and thus may contribute to resistance to anabolic effects of IGF-I. Nitric oxide has been implicated as an important mediator in chondrocyte apoptosis. PGE\textsubscript{2} may mediate directly the induction of apoptosis by nitric oxide or may sensitize chondrocytes to nitric oxide–induced apoptosis.\textsuperscript{384} However, evidence indicates that nitric oxide may inhibit cytokine production or activity in chondrocytes. IL-1 seems to protect chondrocytes from CD95–induced apoptosis through a mechanism that is independent of IL-1–induced nitric oxide.

Novel mediators that affect chondrocyte metabolism include the IL-1–induced suppressor of cytokine signaling 3 (SOCS3), which acts as a negative feedback regulator during IGF-I desensitization in the absence of nitric oxide by inhibiting insulin receptor substrate-1 phosphorylation.\textsuperscript{247} The receptor for advanced glycation end products (RAGE) interacts preferentially with S100A4, a member of the S-100 family of calcium-binding proteins, in chondrocytes and stimulates MMP-13 production.\textsuperscript{385} Recent studies indicate that S100A8, S100A9, and S100A11 have critical roles in inflammatory arthritis and OA.\textsuperscript{396–398} The fibroblast activation protein α, a membrane serine proteinase, which co-localizes in synovium with MMP-1 and MMP-13 and is induced by IL-1β and OSM in chondrocytes, may play a role in collagen degradation.\textsuperscript{399,392} Many of these proteins may be activated during the chondrocyte response to abnormal stimuli, thereby serving as endogenous mediators of cellular responses to stress and inflammation.

Adipokines, which were identified originally as products of adipocytes, also have roles in cartilage metabolism.\textsuperscript{391} White adipose tissue is a major source of proinflammatory and anti-inflammatory cytokines, including IL-1Ra and IL-10,\textsuperscript{392} and the dysregulated balance between leptin and other adipokines, such as adiponectin, promotes destructive processes during inflammation.\textsuperscript{393} Leptin expression is enhanced during acute inflammation, correlating negatively with inflammatory markers in RA sera,\textsuperscript{394} and may serve as a link between the neuroendocrine and immune systems.\textsuperscript{395} Elevated expression of leptin in OA cartilage and in osteophytes and its capacity to stimulate IGF-I and TGF-β1 synthesis suggest a role for this adipokine in anabolic responses of chondrocytes.\textsuperscript{396} Leptin synergizes with IL-1 or interferon-γ to increase nitric oxide production in chondrocytes,\textsuperscript{397} and leptin deficiency attenuates inflammatory processes in experimental arthritis.\textsuperscript{398}

Human articular chondrocytes are capable of expressing receptors of the IL-1R/Toll-like receptor (TLR) superfamily, including TLR1, TLR2, and TLR4. IL-1 and TNF induce the expression of TLRs in chondrocytes and peptidoglycans, and lipopolysaccharide and fibronecrtin fragments act as TLR ligands. TLR activation is associated with increased production of MMPs, nitric oxide, PGE\textsubscript{2}, and VEGF.\textsuperscript{399–401} Cartilage proteoglycan loss in streptococcal cell wall–induced arthritis is predominantly dependent on TLR2 signaling.\textsuperscript{402} In immune complex–mediated arthritis, TLR4 regulates cartilage destruction by IL-10–mediated upregulation of Fcγ receptor expression and enhanced cytokine production.\textsuperscript{403} A recent study showed that two endogenous shared TLR-2/TLR-4 ligands—low-molecular-weight hyaluronan (LMW-HA) and high-mobility group box chromosomal protein 1 (HMGB-1)—which are increased in OA joints, promote MMP-13–mediated ECM remodeling and chondrocyte differentiation toward hypertropy by engaging TLR-dependent, MyD88 (myeloid differentiation factor 88) signaling, leading to the activation of NFκB–dependent genes, including MMP-13.\textsuperscript{404}

### Chemokines

Chemokines, which are small heparin-binding cytokines identified originally as chemotactic factors, are classified as C, CX3C, or CC molecules, indicating the presence of distinct N-terminal cysteine (C) residues. Chondrocytes, when activated by IL-1 and TNF, express several chemokines, including IL-8, monocyte chemoattractant protein (MCP)-1, MCP-4, macrophage inhibitory protein (MIP)-1α, MIP-1β, RANTES (regulated on activation normal T cell expressed and secreted), and growth-related oncogene (GRO)α, and the receptors that enable responses to some of these chemokines and may feedback-regulate synovial cell responses (Table 3-5).\textsuperscript{405–407} The first report of expression of functional CC and CXC chemokine receptors (CCR and CXCR) on chondrocytes showed that interaction of these receptors with their corresponding ligands—MCP-1, RANTES, and GROα—resulted in upregulation of MMP-3.\textsuperscript{408}

Normal and OA chondrocytes express the CC chemokines, MCP-1, MIP-1α, MIP-1β, and RANTES. RANTES increases expression of its own receptor, CCR5. MCP-1 and RANTES increase MMP-3 expression, inhibit proteoglycan synthesis, and enhance proteoglycan release from the chondrocytes. The RANTES receptors CCR3 and CCR5, but not CCR1, are expressed in normal cartilage, whereas all three receptors are expressed in OA cartilage or after stimulation of normal chondrocytes by IL-1β. RANTES induces the expression of iNOS, IL-6, and MMP-1.

<table>
<thead>
<tr>
<th>Functional Name</th>
<th>Systematic Name</th>
<th>Chemokine Receptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROα</td>
<td>CXCL1</td>
<td>CCR1, CCR2</td>
</tr>
<tr>
<td>IL-8</td>
<td>CXCL8</td>
<td>CCR1, CCR2</td>
</tr>
<tr>
<td>MCP-1</td>
<td>CCL2</td>
<td>CCR2</td>
</tr>
<tr>
<td>MIP-1α</td>
<td>CCL3</td>
<td>CCR1, CCR5</td>
</tr>
<tr>
<td>MIP-1β</td>
<td>CCL4</td>
<td>CCR5</td>
</tr>
<tr>
<td>RANTES</td>
<td>CCL5</td>
<td>CCR1, CCR3, CCR5</td>
</tr>
<tr>
<td>SDF-1</td>
<td>CXCL12</td>
<td>CXC4</td>
</tr>
</tbody>
</table>

*Chemokines are classified according to the positions of the first two cysteines (C) of the four conserved N-terminal cysteine (C) residues: CC, chemokine ligand (CCL), first two cysteines are adjacent; CX, chemokine ligand (CXCL), first two cysteines are separated by amino acid X other than cysteine; CCR, CC chemokine receptor; CXCR, CXC chemokine receptor.*

GROα, growth-related oncogene α; IL-8, interleukin-8; MCP-1, monocyte chemoattractant protein-1; MIP-1α, macrophage inhibitory protein-1; RANTES, regulated on activation, normal T cell expressed and secreted; SDF-1, stromal-derived growth factor-1.
High levels of stromal cell–derived factor 1 (SDF-1) are detected in RA synovial fluids, and its receptor, CXCR4, is expressed by chondrocytes, but not by synovial fibroblasts, suggesting a direct influence of this chemokine on cartilage damage. Microarray studies have elucidated many chemokines that are inducible in chondrocytes by fibronectin fragments and cytokines. SDF-1 and several other chemokines also increase the synthesis of S100A, N-acetyl-β-glucuronidase, cathepsin B, and MMPs by chondrocytes and DNA synthesis, cell proliferation, and PGE2 production. OA chondrocytes in contact with autologous T lymphocytes produce enhanced levels of MMP-1, MMP-3, MMP-13, and RANTES. In addition to recruiting leukocytes to sites of inflammation in arthritic joints and mediating synovial fibroblast responses and actions, chemokines are capable of modulating chondrocyte functions that are associated with cartilage degradation.

**Cytokine Signaling Pathways Involved in Cartilage Metabolism**

Although the receptors for IL-1 and TNF and associated adapter molecules are distinct, they share the capacity to activate some of the same signaling pathways (Figure 3-8). The major pathways induced by catabolic cytokines involve signal transduction by the stress-activated protein kinases, JNKs and p38 kinases, and the NFκB and PI-3K pathways. The JAK/STAT signaling pathway is important for signaling by gp130 cytokines, including IL-6 and OSM. Specific adapter molecules involved in the pathways induced by TNF receptors, which are members of the TNF receptor superfamily, are different from the adapter molecules used by IL-1 signaling pathways. The TNF receptor pathway uses TNF receptor–associated factor 2 (TRAF2), TRAF6, and the receptor interacting protein kinase, whereas the IL-1 receptor pathway uses TRAF6, IL-1 receptor–associated kinase (IRAK), and evolutionarily conserved signaling intermediate in Toll pathways (ECHSIT) as adapter molecules. Signaling through TNFRI associated with TNF receptor–associated death domain (TRADD) activates apoptosis, whereas TNF-RII signaling through TRAF2 activates JNK and NFκB.

In contrast to ERK1 and ERK2, p38 and JNK signaling pathways are weakly activated by growth factors. Studies in chondrocytes in vitro have shown that the p38 and JNK cascades mediate the induction of proteinases and proinflammatory genes by IL-1 and TNF. These pathways also may be activated in chondrocytes by mechanical stress and cartilage matrix products via integrins and other receptor-mediated events. Upregulation of IL-1 and TNF expression via mechanotransduction pathways suggests their involvement as secondary mediators in a feedback loop that enhances cartilage degradation.

**Figure 3-8** Intracellular signaling pathways activated by interleukin-1 (IL-1) in chondrocytes. Binding of IL-1 to the type I IL-1 receptor (IL-1R1) leads to recruitment of the IL-1R accessory protein (IL-1RAP). Cytoplasmic Toll/IL-1 receptor (TIR) domains of the receptor recruit MyD88 via its TIR, and the MyD88 death domain (DD) recruits IL-1 receptor–associated kinases (IRAK and IRAK2) to the receptor complex before being rapidly phosphorylated and degraded. The IRAKs mediate tumor necrosis factor receptor–associated factor 6 (TRAF6) oligomerization, initiating various protein kinase cascades, the major ones of which involve (1) the stress-activated protein kinases, p38 mitogen-activated protein kinase (MAPK), and c-Jun N-terminal kinase (JNK), which lead to activation of activator protein-1 (AP-1) (c-Fos/c-Jun), activating transcription factor-2 (ATF-2) and E26 transformation-specific (ETS) factors, among other transcription factors; and (2) inhibitor of κB (IκB) kinases α and β (IKK-α and IKK-β), which lead to activation of nuclear factor κB (NFκB). TNF also stimulates these pathways, but mainly via TRAF2. Other signaling pathways may influence the target gene responses, such as growth factor–induced or chemokine–induced phosphatidylinositol 3-kinase (PI-3K) via the serine/threonine kinase, Akt/protein kinase B, and gp130 cytokine–induced janus activating kinase (JAK)/signal transducer and activator of transcription (STAT) pathway. Responses of the target genes depend on the presence of DNA sequences within the respective promoters that bind to various transcription factors.
mechanism. At least four isoforms of p38 MAPK exist with different substrate specificities and differential effects on essential chondrocyte functions. JNKs are serine threonine protein kinases that phosphorylate Jun family members, components of AP-1 transcription factors, and they exist in humans as three JNK isoforms: JNK1, JNK2, and JNK3. A potent JNK1/2 inhibitor, SP600125, which blocks inflammation and joint damage in animal models of RA, and other JNK isof orm-specific inhibitors are useful tools for analyzing chondrocyte function in vitro and in vivo. Activated JNK is detected in OA, but not in normal cartilage, and JNK inhibition attenuates cytokine-induced chondrocyte responses. 419,420

NFκB is released from inhibitory 1κB by the catalytic activities of the IKKα and IKKβ subunits of the IKK signalosome complex, permitting translocation of active NFκB to the nucleus. 420,421 In response to a host of proinflammatory stimuli, IKKβ is the dominant 1κB kinase in vivo, whose activation is essential for the nuclear entry of canonical NFκB heterodimers, 420,421 thereby ablation of IKKβ in adult chondrocytes could protect them from the stress and inflammatory responses of OA disease onset or progression. The latter would also be in keeping with the beneficial anti-inflammatory effects of physiologic biomechanical signals, which have been reported to inhibit IKK-mediated NFκB activation in chondrocytes. 422

NFκB mediates the expression of cytokines and chemokines induced by fibronectin fragments, 439 and inhibition of DNA-binding activity of NFκB by agents that deplete polyamine blocks IL-1 and TNF without promoting chondrocyte apoptosis. 423 Chondrocyte phenotype appears to be subject to differential control by IKKα-driven noncanonical and IKKβ-driven canonical NFκB. 419 In addition to NFκB, transcription factors that are members of the C/EBP, ETS, and AP-1 families are important for the regulation of gene expression by IL-1 and TNF in chondrocytes. 435 Activation of protein kinase C ζ (PKCζ) appears to be necessary upstream of NFκB in OA articular chondrocytes. 414 It is interesting to note that PKCζ-mediated NFκB RelA/p65 Ser311 phosphorylation has been shown to maintain RelA/p65 activity through a novel mechanism involving suppression of SETD6-mediated lysine 310 methylation. 423

ROLE OF THE CHONDROCYTE IN CARTILAGE REPAIR

Aging of Articular Cartilage

It is important, but often difficult, to distinguish among the effects of aging itself and diseases such as OA that become more common with increasing age. 427,426 In both cases, biochemical alterations in matrix composition are reflected in changes in cartilage structure. 427 The thickness of articular cartilage, as shown by magnetic resonance imaging, decreases with increasing age. 423 Heterogeneous depletion of glycosaminoglycans at the cartilage surface and fatigue fracture of superficial collagen bundles may contribute to the mild splitting and fraying of superficial cartilage, which is termed fibrillation. If fibrillation progresses into deeper layers of cartilage, abnormal multicellular clusters of chondrocytes that stain intensely for glycosaminoglycans are found at the base of clefts. These changes include decreased size and aggregation of aggrecan and increased collagen denaturation, resulting in loss of compressive stiffness and tensile strength.

Zonal differences in tensile strength and compressive resistance are related to differences in matrix composition and can be observed to change during the aging of adult articular cartilage and in response to traumatic damage. The territorial, or pericellular, matrix and the interterritorial matrix differ in the amounts and types of matrix proteins. The chondrocytes are normally surrounded by a 2 ㎛ pericellular matrix, composed of a highly branched filamentous network of collagen VI tetrmeric, which serves as a scaffold for decorin, biglycan, perlecan, and chondroadherin, which predominate in this region, and hyaluronan, fibrillin-1, and PRELP. The interterritorial region, in contrast, contains primarily a collagen II/XI fibrillar network, which binds decorin, fibromodulin, collagen IX, and COMP, and large numbers of intact aggrecan molecules attached via link protein to long chains of hyaluronic acid. In the deep zone, the interterritorial region most remote from the cells contains a larger number of degraded aggrecan molecules that lack the G3 domain.

Proteoglycans in aged cartilage have a wide range of sizes, with small forms resulting from low substitution of glycosaminoglycan residues and shorter lengths compared with glycosaminoglycans in young articular cartilage. 8 Unsubstituted proteoglycan core proteins of aggrecan and biglycan are detectable in articular cartilage from elderly subjects. Hyaluronan content increases in aged cartilage, but with reduced mean chain length, and link protein seems to be fragmented. Collagen fibrils become thinner with age and are less densely packed. Nonenzymatic glycation results in the formation and accumulation of the advanced glycation end product pentosidine in long-lived proteins, including cartilage collagen and aggrecan. 429 Such biochemical changes may result in part from changes in chondrocyte synthetic function and from increased susceptibility of the matrix to degradation. 410 Increased and more extensive collagen degradation can be observed in cartilage from older, healthy individuals, and, similar to cartilage in early OA, the damage is concentrated closer to the articular surface and co-localizes with MMP-13 activity. 111 (See Chapter 98 for a detailed discussion of the pathogenesis of OA.)

Aging Chondrocyte

Chondrocyte function, including mitotic and synthetic activity, deteriorates with age. Degradative changes are generally due to the actions of proteinases and are, at least in part, the cumulative consequences of adverse conditions, such as mechanical insults or inflammation, to which the chondrocyte is exposed throughout life. Deficiencies in cartilage matrix proteins also may disrupt chondrocyte-matrix interactions that are important to cell survival. The decline in chondrocyte number may be attributed to increased cell death with age. Although programmed cell death, or apoptosis, increases with age in adult rats and mice, this may be due to skeletal growth that occurs throughout life in these animals. In human adult cartilage, apoptotic cell removal does not seem to be common, however. 112 Replicative senescence, detected as β-galactosidase activity and decreased
Markers of Cartilage Matrix Degradation and Turnover

With increasing knowledge of the composition of the cartilage matrix, molecular markers in body fluids have been identified for monitoring changes in cartilage metabolism and for assessing joint damage in arthritis.\(^{32,33,434}\) Molecules originating from the articular cartilage, including aggrecan fragments containing chondroitin sulfate and keratan sulfate; type II collagen fragments; collagen pyridinoline cross-links; and COMP, usually are released as degradation products as a result of catabolic processes. Specific monoclonal antibodies have been developed for analyzing OA and RA body fluids for products of proteoglycan or collagen degradation (catabolic epitopes) or synthesis of newly synthesized matrix components (anabolic neoepitopes), which represent attempts to repair the damaged matrix. Different monoclonal antibodies can distinguish subtle biochemical differences in chondroitin sulfate and keratan sulfate chains that result from degraded versus newly synthesized proteoglycans. Such epitopes can be detected in the synovial fluids and sera of patients with OA and RA, and the synovial fluid-to-serum ratio has been suggested as a potential diagnostic indicator. The degradation of aggrecan in cartilage has been characterized using antibodies 846, 3B3(−), and 7D4, which detect chondroitin sulfate neoepitopes; 5D4, which detects keratan sulfate epitopes; and the VIDIPEN and NITEGE antibodies, which recognize aggreganase and MMP cleavage sites within the interglobular G1 domain of aggrecan (see Chapter 8).\(^{299,300}\)

Similarly, the synthesis of type II collagen can be monitored by measuring serum and synovial fluid levels of the carboxyl-terminal propeptide, and urinary excretion of hydroxylysyl pyridinoline cross-links may indicate collagen degradation.\(^{32,341}\) Specific antibodies that recognize epitopes on denatured type II collagen at the collagenase cleavage site are promising diagnostic reagents. These include the C12C antibody (previously known as Col2-3/4C Long mono) that has been used to detect cleavage of the triple helix of type II collagen in experimental models and in OA and RA cartilage.\(^{311,343}\) The ratios of these markers to the synthetic marker, CPII, are associated with a greater likelihood of radiologic progression in OA patients.\(^{34}\) These biomarker assays have been used as research tools and are currently being developed and validated as diagnostic tools for monitoring cartilage degradation or repair in OA and RA patient populations and for assessment of treatment (see Chapters 71 and 100). Although a single marker may be insufficient, it may be possible eventually to identify a combination of biomarkers that discriminate between different stages of OA in different populations.

Repair of Articular Cartilage

Articular cartilage has a poor capacity for regeneration, and pharmacologic enhancement of cartilage repair would have considerable potential in the treatment of arthritides and intra-articular fractures. The extent of intrinsic repair of a cartilage defect depends on the depth of the lesion and whether the defect penetrates the subchondral bone plate.\(^{335}\) Repair of superficial defects occurs if the chondrocytes remain viable. Owing to the avascularity of cartilage, it differs from most other tissues in its response to injury. The vascular-dependent inflammatory and reparative phases of the classic healing response are unavailable. Partial defects generally do not regenerate because resident chondrocytes cannot migrate into the defect, and there is no vascular access for progenitor cells. Deep cartilage defects with disruption of the subchondral bone plate initiate vascular responses, however, including bleeding, fibrin clot formation, and inflammation, which permit cell invasion from the blood or underlying bone marrow. The lesion becomes filled by granulation tissue, which eventually is replaced by fibrocartilage, but rarely by true hyaline cartilage. Current procedures for cartilage repair include joint lavage, tissue debridement, microfracture of the subchondral bone, and transplantation of autologous or allogeneic osteochondral grafts, in addition to the ultimate therapy of total joint replacement.\(^{335}\) These procedures may lead to the formation of fibrous tissue, chondrocyte death, and further cartilage degeneration and have variable success rates.

Transplantation of cultured autologous chondrocytes has been used successfully to repair small, full-thickness lesions in knee cartilage in young adults with sports injuries.\(^{336}\) Evidence of successful repair has been shown by turnover and remodeling of the initial fibrocartilaginous matrix formed by transplanted chondrocytes as the result of enzyme degradation and new synthesis of type II collagen.\(^{337}\) The donor site, although not load bearing, may undergo significant morbidity and osteoarthritic changes. A randomized, controlled trial suggests little difference in efficacy compared with microfracture of subchondral bone.\(^{338}\)

Major challenges for cartilage repair include restoration of the three-dimensional collagen structure and integration of the newly synthesized matrix with resident tissue.\(^{335,349}\) Novel approaches using autologous chondrocytes genetically engineered ex vivo to express anabolic factors have been explored to promote differentiation before implantation in the defect.\(^{332,340}\) Induction of the synthesis of IGF-I, TGF-β, BMP-2, and BMP-7 by gene transfer increases the synthesis of cartilage proteoglycans and collagens in cultured chondrocytes.\(^{341}\) Many of these factors, including BMP-2, BMP-4, BMP-6, BMP-7, BMP-9, TGF-β, and CDMP-1, are able to induce chondrogenic differentiation of mesenchymal progenitor cells in vitro.\(^{442,443}\) BMP-2 and BMP-7 (OP-1) are approved for multiple indications in the area of bone fracture repair and spinal fusion and can promote cartilage repair in various models of focal cartilage

TGF-β, FGFs, IGF-I, and other anabolic factors that support cartilage matrix biosynthesis are expressed at declining levels with aging, or their activities are downregulated.\(^{237}\) The capacity of BMP-6 to stimulate proteoglycan synthesis and the production of BMP-7 (OP-1) decline with age.\(^{230,277}\) Chondrocytes also show an age-related decline in the anabolic response to IGF-I, possibly owing to increased synthesis of IGFBP-3, which is itself antiproliferative. Chondrocytes from elderly donors depend strongly on IGF-I and IGF-II for survival.\(^{235}\) It has been proposed that the reduction in TGF-β signaling in aging chondrocytes may be a factor in their reduced capacity to repair cartilage.\(^{433}\)
SUMMARY AND CONCLUSION

As the single cellular component in adult articular cartilage, chondrocytes are responsible for maintaining the ECM components in a low-turnover state. The composition and organization of matrix macromolecules, unique to this tissue, are determined during chondrocyte differentiation in embryonic and postnatal development of cartilage. Adult chondrocytes exist in a hypoxic environment within articular cartilage. They are inactive metabolically, partially as a result of the absence of blood vessels and nerves, and display a rounded morphology that reflects their quiescent state. Chondrocyte culture models have been developed with the aim of maintaining differentiated phenotypes, characterized by the major collagen and proteoglycan constituents, type II collagen and aggrecan. Chondrocytes interact with specific ECM components via cell surface receptors, including integrins, annexins, syndecans, DDR2, and CD44.

Studies in vitro and in vivo have shown that adult articular chondrocytes are capable of responding to biologic and mechanical stimuli that are anabolic or catabolic. Anabolic factors include members of the TGF-β/BMP family and IGF-I. Catabolic factors include proinflammatory cytokines, such as IL-1, TNF, IL-17, and IL-18, which stimulate the synthesis of matrix-degrading proteinases, such as MMPs and aggrecanases, and inhibit the synthesis of cartilage matrix proteins. Many of the signaling pathways and transcription factors that mediate the responses of chondrocytes to these factors have been elucidated, but how they orchestrate specific chondrocyte functions is complex and is not fully understood. Under physiologic conditions, the adult articular chondrocyte maintains a stable equilibrium of low-turnover replacement of matrix components. Adult chondrocytes have a poor capacity for mediating effective repair of extensive cartilage lesions, and this capacity declines with age. Age-related changes in chondrocyte function decrease the ability of cells to maintain the tissue, including decreased synthetic activity, synthesis of smaller and less uniform aggrecan molecules and less functional link proteins, and decreased responsiveness to anabolic growth factors. Further understanding of how the adult articular chondrocyte functions within its unique environment would aid in the development of rational strategies for maintaining homeostasis and protecting against cartilage damage.

Selected References


68. Tamamura Y, Otani T, Kanatani N, et al: Developmental regulation of Wnt/beta-catenin signals is required for growth plate assembly,

69. Day TF, Guo X, Garrett-Beal L, Yang Y: Wnt/beta-catenin signaling in mesenchymal progenitors controls osteoblast and chondrocyte differentiation during vertebrate skeletogenesis, *Dev Cell* 8:739–750, 2005.


82. Day TF, Guo X, Garrett-Beal L, Yang Y: Wnt/beta-catenin signaling in mesenchymal progenitors controls osteoblast and chondrocyte differentiation during vertebrate skeletogenesis, *Dev Cell* 8:739–750, 2005.


91. Tchetina EV, Squires G, Poole AR: Increased type II collagen degradation and very early focal cartilage degeneration is associated with transcriptional activation of Sox9, Bone 37:313–322, 2005.


98. Day TF, Guo X, Garrett-Beal L, Yang Y: Wnt/beta-catenin signaling in mesenchymal progenitors controls osteoblast and chondrocyte differentiation during vertebrate skeletogenesis, *Dev Cell* 8:739–750, 2005.


*Full references for this chapter can be found on www.expertconsult.com.*


64. Stockwell RA, Meachim G: The chondrocytes. In Freeman MAR, editor: Adult articular cartilage, Tunbridge Wells, United Kingdom, 1979, Pitman Medical, pp 69–144.


CHAPTER 3 | CARTILAGE AND CHONDROCYTES


Biology, Physiology, and Morphology of Bone

GEORG SCHETT

KEY POINTS

Intramembranous or endochondral ossification generates the bone tissue.

Bones consist of a dense cortical shell and sponge-like trabecular network (cancellous bone).

Bone formation depends on metabolically active osteoblasts synthesizing matrix proteins.

Resorption of bone is mediated by multinucleated hematopoietic cells, the osteoclasts.

The most abundant cell type in bone is the osteocyte.

Bone is continuously rebuilt, a process known as bone remodeling.

The immune system, in particular T lymphocytes, influences bone remodeling. Neuroendocrine systems exert systemic control on bone remodeling.

STRUCTURE AND COMPOSITION OF BONE

Bone is a specialized connective tissue that serves (1) locomotion by providing the insertion site of the muscles, (2) protection of the internal organs and the bone marrow as well, and (3) metabolic function such as storage and provision of calcium to the body. Bone consists of cells and the extracellular matrix, which is composed of type I collagen fibers and a number of noncollagenous proteins. The specific composition of the bone matrix allows its mineralization, which is a specific feature of bone.

Two major types of bones exist: flat bones, which are built by intramembranous ossification, and long bones, which emerge from endochondral ossification. Intramembranous bone formation is based on the condensation of mesenchymal stem cells, which directly differentiate into bone-forming osteoblasts. In contrast, during endochondral ossification of the long bones, the mesenchymal stem cells first differentiate into chondrocytes that will further be replaced by osteoblasts. Long bones consist of the (1) epiphyses, which are protrusions at the ends of the long bones; (2) the diaphysis constituting its shaft; and (3) the metaphyses, which are located between the epiphysis and the diaphysis (Figure 4-1). The metaphysis is separated from the epiphysis by the growth plate, a proliferative cartilage layer, which is essential for the longitudinal growth of bones. After finishing growth, this cartilage layer is entirely remodeled into bone. The external shape of bones is formed by a dense cortical shell (cortical or compact bone), which is particularly strong along the diaphysis, where the bone marrow is located. The cortical bone shell becomes progressively thinner toward the metaphyses and epiphyses, where most of the trabecular bone is located. Trabecular bone (also called cancellous bone) is a sponge-like network consisting of myriads of highly interconnected bony trabeculae. The outer and the inner surfaces of cortical bone are covered by layers of osteogenic cells, termed the periosteum and the endosteum, which are involved in the growth of width by bone apposition at the periosteal and bone resorption at the endosteal sites.

Although cortical and trabecular bone is composed of the same cells and same matrix components, there is a substantial difference between these two forms of skeletal tissue. Cortical bone almost exclusively consists of mineralized tissue (up to 90%), allowing it to fulfill its mechanical requirements. In contrast, only 20% of trabecular bone is mineralized tissue, whereas the bone marrow, blood vessels, and a network of mesenchymal stem cells cover the rest. As a consequence, trabecular bone shares a vast surface with the nonmineralized tissue, which is the basis for the metabolic function of bone, necessitating a high level of communication between the bone surface and the nonmineralized tissue.

Bone Matrix

The key protein component of bone is collagen type I. Collagen fibers follow specific directions forming the basis for the lamellar structure of bone. This lamellar structure, which can be visualized when examining bone in the polarized light, allows dense packaging, resulting in optimal resistance to mechanical load. The lamellar collagen structures can be assembled in parallel (e.g., along the cortical bone surfaces and inside the bony trabeculae) or concentrically around blood vessels embedded in the Haversian channels of the cortical bone. In case of rapid deposition of new bone such as during fracture healing, this lamellar structure is missing and the bone is then called woven bone. Woven bone is consecutively remodeled into lamellar bone, which is also considered “mature” bone. The composition of the collagen backbone also assists the deposition of spindle- or plate-shaped hydroxyapatite crystals, which contain calcium phosphate, allowing the calcification of the bone matrix.

In addition to collagen type I, other so-called noncollagenous proteins also exist in bone. Some of them such as osteocalcin, osteopontin, and fetuin are mineralization inhibitors, which serve to balance the degree of mineralization of the skeletal tissue. Aside from their intrinsic function in bone, noncollagenous proteins also exert important metabolic functions such as the control of energy metabolism by osteocalcin.
Bone Cells: Osteoblasts

Osteoblasts are the bone-forming cells that derive from the mesenchymal stem cells of the bone marrow, which also form chondrocytes, myocytes, and adipocytes. Osteoblasts are cuboid-shaped cells that form clusters covering the bone surface. They are metabolically highly active, synthesizing the collagenous and noncollagenous bone matrix proteins, which are excreted and then deposited between the osteoblasts and the bone surface. This newly built matrix, which is not yet calcified, is termed the osteoid. The lag phase between osteoid deposition and its mineralization is approximately 10 days. Osteoblast differentiation depends on the expression of two key transcription factors, Runx2 and its target Osterix 1, which confer the differentiation of these cells into osteoblasts in response to external stimuli.\(^1\) Prostaglandin E\(_2\) (PGE\(_2\)), insulin-like growth factor (IGF)-1, parathyroid hormone (PTH), bone morphogenic proteins (BMPs), and Wingless and Int-1 (Wnt) proteins are key stimuli for osteoblast differentiation.\(^2,3\) Prostaglandin E2, for instance, is an important anabolic factor for bone and induces the expression of bone sialoprotein and alkaline phosphatase in mesenchymal cells. Bone morphogenic proteins (BMPs) and transforming growth factor (TGF)-\(\beta\), which shares structural similarities with BMPs, foster osteoblast differentiation by activating intracellular Smad proteins. Finally, Wnt proteins, a family of highly conserved signaling molecules, are potent stimulators of osteoblast differentiation. Wnt proteins bind to surface receptors on mesenchymal cells such as Frizzled and LRP5, eliciting activation and nuclear translocation of the transcription factor \(\beta\)-catenin, which induces the transcription of genes involved in osteoblast differentiation. Wnts thereby act not only in close synergy with BMPs but also cross-talk to the receptor activator of nuclear factor \(\kappa\)B ligand (RANKL)-osteoprotegerin (OPG) system, which is involved in the differentiation and function of bone-resorbing osteoclasts.

Bone Cells: Osteocytes

Osteocytes are by far the most abundant cell type within bone. One cubic millimeter of bone contains up to 25,000 osteocytes, which are well connected with each other and the bone surface by small tubes (canaliculi) constituting a large and dense communication network inside the bone, which has similarities to the nervous system. The surface of this network of lacunae containing the osteocytes and the canaliculi containing the interconnecting filaments of the osteocytes covers an area of 1000 to 4000 square meters. Osteocytes are derived from osteoblasts, which are subsequently entrapped in the bone matrix. Osteocytes however, also start to express genes that are specific for these cells and not found in other cells such as osteoblasts. One of the most interesting products of the osteocyte is sclerostin, a secreted molecule that binds LRPs and blocks Wnt-stimulated bone formation.\(^4,5\) Consistent with its function as an inhibitor of bone formation, overexpression of sclerostin leads to low bone mass, whereas deletion of sclerostin leads to increased bone density and strength. Loss-of-function mutations in the human SOST gene encoding sclerostin ential increased bone mass, a disease termed sclerosteosis. Several local and systemic factors have been suggested as possible regulators of sclerostin expression by osteocytes. For instance, intermittent administrations of parathyroid hormone (PTH), which are associated with strong anabolic effects on the bone, potently inhibit sclerostin expression.
Bone Cells: Osteoclasts

Osteoclasts are multinucleated cells containing up to 20 nuclei and are unique in their ability to resorb bone.\(^6,7\) They are directly attached to the bone surface and build resorption lacunae (Howship’s lacunae). Apart from their multiple nuclei, another characteristic of the osteoclast is the ruffled border, a highly folded plasma membrane facing the bone matrix and designed to secrete and resorb proteins and ions into the space between the osteoclast and bone surface (Figure 4-2). The space between this ruffled border and the bone surface is the place where bone resorption occurs. It is sealed by a ring of contractile proteins and tight junctions because it represents one of the few regions of the human body, where a highly acidic milieu is found. Bone degradation by osteoclasts comprises two major steps: first, demineralization of inorganic bone components, and second, removal of organic bone matrix. To demineralize bone, osteoclasts secrete hydrochloric acid through proton pumps into the resorption lacunae. This proton pump requires energy, which is provided by an ATPase allowing the enrichment of protons in the resorption compartment, which, in fact, represents an extracellular lysosome. In addition to protons and chloride, osteoclasts release matrix-degrading enzymes including tartrate resistant acid phosphatase (TRAP), lysosomal cathepsin K, and other cathepsins. Cathepsin K can effectively degrade collagens and other bone matrix proteins. Consequently, inhibitors of cathepsin K block osteoclast function and slow down bone resorption.

Osteoclasts originate from hematopoietic monocytic precursor cells and, upon the influence of specific signals, undergo a series of differentiation steps to become mature osteoclasts. Essential signals for osteoclast differentiation are macrophage colony stimulating factor (M-CSF) and RANKL. During this differentiation and maturation process, osteoclasts acquire specific markers such as TRAP, fuse to multinucleated giant cells, and polarize on contact to bone. Osteoclastogenesis depends on an adequate microenvironment, which provides essential signals such as M-CSF and RANKL and also certain cytokines like tumor necrosis factor (TNF), which further enhance osteoclast differentiation. Mesenchymal cells such as preosteoblasts express M-CSF and RANKL and can induce osteoclast formation highlighting the close interaction between bone formation and bone resorption.

RANKL, a TNF superfamily member, is a surface molecule expressed by a large set of different cell types including preosteoblasts and activated T cells.\(^8\) Under steady-state conditions its expression is induced in cells of the osteoblastic lineage in response to osteotropic factors such as vitamin D, parathyroid hormone, and prostaglandins. Moreover, inflammatory cytokines such as TNF, interleukin (IL)-1, and IL-17 can induce RANKL expression.\(^9-12\) RANKL is essential for the final differentiation steps of osteoclasts, as well as for their bone-resorbing capacity by engaging its receptor RANK on monocytic osteoclast-precursor cells. The interaction of RANKL with its receptor RANK is modulated by OPG, a secreted glycoprotein, which was identified as a soluble factor that strongly suppresses osteoclast differentiation both in vitro and in vivo. Interestingly, OPG expression is induced by estrogens, which explains the increase in osteoclast numbers and enhanced bone resorption during menopause. In accordance, RANKL-deficient mice display severe osteopetrosis due to the lack of osteoclasts. Regarding the central role of the RANKL-RANK-OPG signaling system in bone resorption, researchers are increasingly interested in the therapeutic targeting of this system.

**Figure 4-2** Osteoclasts form from mononuclear precursors (colony-forming unit macrophages [CFU-M]), which differentiate into mononuclear osteoclast precursors. These cells fuse and build a polykaryon, forming a preosteoclast. Final differentiation is characterized by cell polarization and acquisition of a ruffled border. M-CSF, macrophage colony-stimulating factor; RANKL, receptor activator of nuclear factor κB ligand; TRAP, tartrate resistant acid phosphatase.
system in human disease and recent clinical trials on postmenopausal osteoporosis have revealed the potent antiresorptive effect of a neutralizing RANKL antibody (denosumab). Beyond RANKL-RANK interactions, other important pro-osteoclastogenic signaling pathways are based on the triggering receptor expressed on myeloid cells (TREM) 2, which interacts with the tyrosine kinase DAP12 and the osteoclast-associated immunoglobulin-like receptor (OSCAR). Both molecules are strong enhancers of osteoclastogenesis.

Bone Remodeling Process

Developmental bone growth, postdevelopmental maintenance and repair of bone, and provision of calcium from the bone depend on a dynamic process called bone remodeling (Figure 4-3). Yet unknown factors, which may likely be of mechanical nature and sensed by the osteocytes, initiate bone remodeling at a specific site. Death of osteocytes and the resulting metabolic changes leading to a lack of silencers for bone turnover, like sclerostin, may govern this activation process. It is followed by a resorptive phase dominated by osteoclast-mediated degradation of the bone matrix resulting in a resorption lacuna. The naked bone surface inside this lacuna is subsequently populated by mesenchymal cells immigrating from the neighboring bone surface, which start differentiating into osteoblasts and produce the new bone matrix (also termed osteoid). This matrix then mineralizes, and the bone returns to its resting state again. This entire bone remodeling process takes about 3 to 6 months. Adults continuously remodel their skeleton, a process that runs even faster in childhood and adolescence. In adults, it takes 7 to 10 years to remodel the entire skeleton, indicating that we fully replace it several times during our lifetime. Most of the bone remodeling happens in the trabecular bone, which promotes the building of an optimal inner microstructure adapted to the individual mechanical demands. Trabecular bone is the leading structure in the vertebral bodies (up to two thirds of the bone substance) and in long bones such as the femurs (about 50% of the bone substance).

Normal physiologic circumstances ensure a balance between bone formation and bone resorption to maintain skeletal homeostasis. This bone remodeling process requires a tight mutual regulation of bone resorption by osteoclasts and bone formation by osteoblasts, a phenomenon called coupling. Coupling is regulated at least at three different levels: (1) by a direct interaction between osteoblasts and osteoclasts, (2) by local interactions between the immune system and bone cells, and (3) by neuroendocrine systemic control of bone metabolism.

Direct Interactions between Osteoblasts and Osteoclasts

A proper coupling between bone formation and bone destruction is essential to maintain bone integrity (Figure 4-4). This coupling process involves two main mechanisms. The first one is the expression of the essential pro-osteoclastogenic cytokines by the osteoblast lineage, and the second one involves the ephrin ligand/ephrin receptor bidirectional signaling. Preosteoblasts are the main pro-osteoclastogenic cells in normal physiologic conditions providing the first level of coupling between bone formation and bone resorption. In response to Wnt signaling, osteoblasts slowly lose their supportive activity for osteoclasts when maturing toward more mineralizing cells and then becoming the bone-embedded osteocytes. They will then secrete antiosteoclastogenic molecules such as OPG and the Wnt inhibitors sclerostin, dickkopf-1, and secreted frizzled-related protein 1 (SFRP1), which either block osteoclast differentiation (OPG) or also inhibit the further differentiation of osteoblasts. A second level of coupling involves the expression of ephrin ligands on the surface of osteoclast progenitors that can bind to ephrin receptors and activate their tyrosine kinase activity. Two ephrin ligands regulate

![Figure 4-3](image-url)
CHAPTER 4 | BIOLOGY, PHYSIOLOGY, AND MORPHOLOGY OF BONE

Bone remodeling. The first, ephrin-B2, binds to the receptor EphB4 on osteoblast progenitors, increasing their differentiation and stimulating bone formation. The second, ephrin-A2, acts in an autocrine manner by binding to the EphA2 receptor on the osteoclasts, promoting their differentiation in a paracrine manner on the osteoblasts, inhibiting their differentiation.

**Bone Remodeling by the Immune System**

The reciprocal regulation of osteoblast and osteoclast differentiation, as well as local immune cells, regulate bone remodeling. This interaction, which experts in the field of osteoimmunology have recently examined extensively, is highly relevant for pathologic bone remodeling as observed in postmenopausal osteoporosis and inflammatory diseases. RANKL, which is also expressed by activated T cells and promotes dendritic cells’ survival, was found on the various subsets of proliferative T cells (CD8 and CD4, Th1 and Th2), as well as on FoxP3-expressing regulatory T cells (Tregs).

Despite consistent expression of RANKL on various T cell lineages, individual T cells exert different function on the osteoclast, which depend on their cytokine and surface molecule expression pattern. For instance, Th17 cells stimulate osteoclast differentiation via the production of IL-17, which stimulates the synthesis of pro-osteoclastogenic molecules such as RANKL on mesenchymal cells. However, T cells can also secrete or express strong inhibitors of osteoclastogenesis such as OPG, interferon (IFN)-γ, IL-4, or CTLA-4. Indeed, Th1 cells repress rather than stimulate osteoclastogenesis by the expression of IFN-γ. Moreover, regulatory T cells effectively block osteoclast differentiation by the expression of CTLA-4, a molecule that blocks T cell co-stimulation.

**Systemic Control of Bone Remodeling by Neuroendocrine Mechanisms**

Bone remodeling is regulated not only locally but also by various systemic hormonal pathways including the sex hormone and the growth hormone (GH)/insulin-like growth factor (IGF) axes. In addition, two major systemic neuroendocrine regulators of bone homeostasis co-regulate bone, fat, and energy metabolism. The two central players of this systemic loop seem to be osteocalcin and leptin. Osteocalcin is a hormone produced by mature osteoblasts and acts on the β cells of the pancreas to stimulate proliferation and thus insulin production in response to leptin. In addition, osteocalcin can directly stimulate adipocytes to regulate insulin sensitivity. Leptin is a peptide hormone produced by adipocytes of the white adipose tissues. Its deficiency is causing obesity and increased bone mass. Although the increased fat mass is certainly linked to the role of leptin in controlling appetite, the effect of leptin on bone and fat can be dissociated. Indeed, bone formation is negatively regulated by leptin through a hypothalamic pathway: the β-adrenergic sympathetic nervous system that mediates decreased osteoblast proliferation via the induction of clock genes in osteoblasts. How leptin regulates bone formation is debated. However, two potential hypothalamic relays have been identified. The first one is the NPY peptide, a repressor of bone formation, that is silenced by leptin. The second downstream modulator of leptin is...
neuromedin U, which inhibits clock gene expression and increases osteoblast proliferation. These observations suggest that disturbance of the metabolic loops linking adipogenesis, osteogenesis, and insulin production have profound consequences on bone homeostasis.

CONCLUSION

Bone is continuously remodeled by bone-resorbing osteoclasts and bone-forming osteoblasts. This remodeling process allows the optimal adaptation of the bone architecture to individual demands and tight control of calcium homeostasis. Local factors control the bone remodeling process on the basis of osteoclast-osteoblast interactions, as well as by systemic immune and neuroendocrine factors controlling the bone-resorbing and bone-forming cells.

References


The references for this chapter can also be found on www.expertconsult.com.
Muscle: Anatomy, Physiology, and Biochemistry

MARK S. MILLER • BRADLEY M. PALMER • MICHAEL J. TOTH • DAVID M. WARSHAW

STRUCTURE

Muscle Tissue

Parallel, aligned bundles of skeletal muscle fibers make up approximately 85% of muscle tissue and consist of a variety of signaling and contractile proteins (Table 5-1). Nerves, blood supply, and connective tissue structures that provide support, elasticity, and force transmission to the skeleton (see later discussion) constitute the remaining volume. Muscle fibers range in length from a few millimeters to 30 cm and in diameter from 10 to 500 µm, with a typical length of 3 cm and diameter of 100 µm. This elongated shape is determined by the organization of the contractile proteins that occupy most of the sarcoplasm. Each muscle has a limited range of shortening that is amplified into large motions by lever systems of the skeleton, usually operating at a mechanical disadvantage. Variations in geometric arrangements of the fibers—parallel, convergent (fan-shaped), pennate (feather-like), sphincter (circular), or fusiform (thick in the middle with tapered ends)—determine some of the mechanical properties. For example, a muscle with fibers aligned parallel to the force-generating axis will have more basic contractile units (i.e., sarcomeres as discussed later) in series than a similarly sized pennate muscle, allowing the parallel muscle to contract quicker, but with less force, than the pennate muscle. Muscles designed for strength (e.g., gastrocnemius) are typically pennate, whereas those designed for speed (e.g., biceps) tend to have parallel fibers. Muscles are commonly arranged around joints as antagonistic pairs facilitating bidirectional motion. When one muscle (the agonist) contracts, another (its antagonist) is relaxed and passively extended. Their roles reverse to actively generate the opposite motion, unless it occurs passively by the force of gravity.

An extensive network of areolar connective tissue, forming the endomysium, surrounds each muscle fiber. Fine nerve branches and small capillaries, necessary for the exchange of nutrients and metabolic waste products, penetrate this layer. The endomysium is continuous with the perimysium, a connective tissue network that ensheathes small parallel bundles of muscle fibers known as fasciculi, intrafusal fibers, larger nerves, and blood vessels. The epimysium encompasses the whole muscle. All three layers of connective tissue contain collagen, mostly types I, III, IV, and V, with types IV and V predominating in the basement membranes surrounding each skeletal muscle fiber. The α1,α2-chain composition of the collagen IV isoform is the most prevalent and provides the mechanical stability and flexibility of the basal lamina (see Chapter 2). The perimysium and endomysium merge at the junction between muscle fibers and tendons, aponeuroses, and fasciae. These

Approximately 660 skeletal muscles support and move the body under the control of the central nervous system. They constitute up to 40% of adult human body mass. Most skeletal muscles are fastened by collagenous tendons across joints in the skeleton. The transduction of chemical energy into mechanical work by muscle cells leads to muscle shortening and consequent movement. A high degree of specialization in this tissue is evident from the intricate architecture and kinetics of intracellular membrane systems, the contractile proteins, and the molecular components that transmit force extracellularly to the basement membrane and tendons. Muscle cells normally exhibit wide variations in activity level and are able to adapt in size, isoenzyme composition, membrane organization, and energetics. In pathologic states, they often become deconditioned. These examples of plasticity can be surprisingly swift and extensive. This chapter outlines the structure and function of muscle and its relationship to associated connective tissue. It also introduces the basis for the highly adaptive response to altered functional demands and diseases. Two excellent Web sites can be accessed for further information in these areas.

KEY POINTS

The structure and function of skeletal muscle and its neural recruitment pattern can change rapidly in response to activity level (i.e., plasticity).

The smallest functional unit of muscle, the sarcomere, is composed of an almost crystalline array of filamentous proteins that convert metabolic energy into force and movement.

Muscles are connected to the skeleton through collagenous tendons.

Skeletal muscle contraction is controlled by the central nervous system through depolarization of specific efferent neurons called motor neurons.

Motor neurons innervate and depolarize muscle fibers through cholinergic synapses called neuromuscular junctions.

Afferent neurons provide the central nervous system with sensory information required for effective control of movement and posture.

Force is transmitted to the exterior through two sets of protein cell adhesion complexes: integrins and dystroglycans.
layers give the attachment sites great tensile strength and distribute axial force into shear forces over a larger surface area.

Fiber Types
Muscles adapt to their specific functions. In any given muscle, part of this adaptation arises from its composition and organization of fiber types. Human skeletal muscle fibers can be classified according to their myosin heavy chain (MHC) isoform (I, IIA, or IIX). MHC molecules break down adenosine triphosphate (ATP) to produce the energy necessary for muscle contraction. The rate of ATP breakdown, or adenosine triphosphatase (ATPase) rate, of the MHC isoforms is I < IIA < IIX, leading to contractions that are relatively slow in MHC I fibers, fast in MHC IIA fibers and very fast in MHC IIX fibers. ATP synthesis primarily occurs through aerobic respiration (oxygen required) in MHC I or slow oxidative fibers; this is aided by their having more mitochondria, an increased capillary blood supply, and more myoglobin compared with MHC IIX, or fast glycolytic fibers that use anaerobic respiration (oxygen not required) to restore their ATP levels. MHC I fibers produce less power than MHC IIX fibers but are more resistant to fatigue. MHC IIA or fast oxidative-glycolytic fibers can use aerobic and anaerobic respiration and fall in between I and IIA fibers in terms of mitochondria, blood supply, myoglobin, power output, and fatigability. Although most human skeletal muscles contain a mixture of fiber types, MHC I fibers are more prominent in muscles used for posture and endurance, and MHC IIX fibers dominate in muscles used for quick movements of short duration. Notably, individual fibers can contain a mixture of MHC isoforms, leading to six different fiber types in humans (“pure”: MHC I, IIA, IIX; “mixed”: MHC I/IIA, IIA/IIX, I/IIX/IIA), which allow for a wide range of contractile properties. A list of various attributes of MHC isoforms can be found in Table 5-2.

During development, fiber-type specificity may be partially determined before innervation. Although the biologic events and signals responsible for designating functional specialization in muscle fibers are not fully understood, classic cross-innervation experiments demonstrated that innervation can dynamically specify and modify the type of muscle fiber. After cross-innervation, the functional and histologic properties listed in Table 5-2 shift toward the target fiber type over a few weeks’ time, indicating the ability of muscles to adapt and remodel in accordance with the pattern of neuronal activity.

EVENTS DURING MUSCLE CONTRACTION

Neural Control
Voluntary control of muscle activation is a complex process. Afferent neurons emanating from sensory organs, such as cutaneous mechanoreceptors and thermoreceptors, pain receptors, joint receptors, and tendon organ and muscle spindles, provide the central nervous system (CNS) with stimuli in the form of action potentials that, with or without additional stimuli from the brain, provide the necessary information for feedback control of effector organs via efferent neurons. Efferent neurons are specifically called motor neurons if their axons innervate muscle. Many times, more afferent than efferent neurons afford effective feedback control of movement and posture. The afferent and efferent neurons are accompanied by Schwann cells, which are glial
cells residing in the peripheral nervous system. Neurons are termed myelinated if Schwann cells wrap around the axon at regularly spaced intervals. The points of bare axon between these Schwann cells are called nodes of Ranvier. Myelination enhances the velocity of action potential propagation by compelling a saltatory conduction of the action potential between neighboring nodes. Schwann cells may also fully or nearly fully cover the axon, thus rendering the neuron unmyelinated and relatively slow in action potential propagation. Three groups of myelinated motor neurons (α, β, γ) are distinguished by diameter, propagation velocity, and target fiber type. Skeletal muscle fibers typically are innervated at several neuromuscular junctions along their length by branches of an α-motoneuron (largest and fastest) or β-motoneuron (Figure 5-1). Muscle spindles are innervated by β- or γ-motoneurons, in addition to the afferent system, for sensing muscle length and force. A single motor neuron and the muscle fibers it innervates constitute a motor unit. When a motor neuron is excited, all fibers in the motor unit are triggered to contract simultaneously. A motor unit responsible for fine movement contains few muscle fibers, but motor units for gross movement generally contain many fibers. The level of muscle activation is controlled from the CNS by the number of motor units recruited and the stimulus rate. Stimulus rate can be so infrequent as to elicit single muscle twitches, such as occur with the monosynaptic stretch reflex involving patellar tendon stretch and quadriceps activation, or conversely so frequent that individual twitches effectively fuse, causing nearly continuous activation of muscle force.

**Neuromuscular Transmission**

At the neuromuscular junction, the axon tapers, loses its myelin sheath, and ends as a presynaptic terminal crowded with vesicles containing the neurotransmitter.
acetylcholine. The postsynaptic membrane of the muscle is indented into folds that increase its surface area and the number of nicotinic acetylcholine-receptors bound therein (see Figure 5-1). The junctional cleft is a 20- to 40-nm-wide space between the presynaptic and postsynaptic membranes. When the motor neuron action potential reaches the presynaptic terminal, local voltage-gated Ca\(^{2+}\) channels open and extracellular Ca\(^{2+}\) streams into the terminal. Within milliseconds of Ca\(^{2+}\) influx, the acetylcholine-loaded vesicles fuse with the presynaptic membrane. Exocytosed acetylcholine rapidly diffuses across the junctional cleft and binds to the nicotinic acetylcholine receptors, which in turn open Na\(^{+}\) and K\(^{+}\) channels of the postsynaptic membrane. The membrane is locally depolarized. An action potential is initiated and propagates along the muscle membrane (sarcolemma) at velocities up to 5 m/sec.

Excitation-Contraction Coupling

A network of tubes invaginate the sarcolemma and run deep into the muscle fiber. This transverse tubule network (T-tubules) pervades the fiber at regular intervals coinciding with sarcomere boundaries along the length of the muscle and surrounds the contractile apparatus with connected longitudinal and lateral segments (Figure 5-2). The lumen of this network is open to the extracellular space, and it contains high Na\(^{+}\) and low K\(^{+}\) concentrations of interstitial fluid. Action potentials at the surface membrane invade the entire T-tubular system. A specialized type of endoplasmic reticulum forms an entirely intracellular membrane system termed the sarcoplasmic reticulum (SR). Prevalent structures containing a T-tubule flanked by two terminal cisternae of the SR to form junctional complexes are termed triads (see Figure 5-2). Terminal cisternae contain oligomers of the Ca\(^{2+}\)-binding protein calsequestrin that provide the fiber with an internal reservoir of calcium ions. Ca\(^{2+}\) channels, termed dihydropyridine receptors (DHPRs), are localized in the T-tubule membranes facing the cytoplasmic domain of SR Ca\(^{2+}\) release channels, also called ryanodine receptors (RyRs), in the terminal cisternae membranes. These membrane proteins are further characterized in Table 5-1.

When an action potential depolarizes the T-tubular membrane, the DHPRs, primarily voltage sensors in skeletal muscle, transfer a signal from the T-tubules to the RyRs by direct interprotein coupling. Ca\(^{2+}\) is then released cooperatively through the RyRs from the SR into the myoplasm, where it activates the contractile machinery. This sequence of events is termed excitation-contraction coupling.

Mutations in the α-subunit of the DHPR in dysgenic mice lead to paralysis, because in these mutants, depolarization of the skeletal muscle membrane does not initiate release of Ca\(^{2+}\) from the SR. Excitation-contraction coupling can be restored in cultured cells from these mice by transfection with complementary DNA encoding for the DHPR, and transfections using chimeric constructs have pinpointed the domain within the DHPR that specifies skeletal- or cardiac-type excitation-contraction coupling. Isoforms of the RyRs also help determine the characteristics of coupling between T-tubules and the SR. Channelopathies in human skeletal and heart muscle have been linked to DHPR mutations. Human malignant hyperthermia occurs in individuals with mutant RyRs that become trapped in the open state after exposure to halothane anesthetic agents.

Contractile Apparatus

The specific locations and functions of the contractile proteins are listed in Table 5-1. Myofilaments (Figure 5-3D) are long, 1-µm-diameter cylindrical organelles that contain the contractile protein arrays responsible for work production, force generation, and shortening. Each myofibril is a column of sarcomeres, the basic contractile units, which are approximately 2.5 µm in length and are delimited by Z lines (Figure 5-3D and E) containing the densely packed structural protein α-actinin. The contractile and structural proteins within each sarcomere form a highly ordered, nearly crystalline lattice of interdigitating thick and thin myofilaments (Figure 5-3E, I, and J). Myofilaments are remarkably uniform in both length and lateral registration, even during contraction, resulting in the cross-striated histologic appearance of skeletal and cardiac muscles. This

Figure 5-2  Membrane systems that relay the excitation signal from the sarcolemma to the cell interior. In the electron micrograph, two T-tubules are cut in cross-section. Electron densities spanning the gap between T-tubules and sarcoplasmic reticulum membranes are the ryanodine receptors, channels that release calcium into the myoplasm. (From Alberts B, Bray D, Lewis J, et al: Molecular biology of the cell, ed 2, New York, 1989, Garland Publishers. Micrograph courtesy Dr. Clara Franzini-Armstrong, University of Pennsylvania, Philadelphia.)
highly periodic organization has facilitated biophysical studies of muscle by sophisticated structural and spectroscopic techniques.\textsuperscript{25,26}

Thick filaments (1.6-µm long) containing the motor protein myosin are located in the center of the sarcomere in the optically anisotropic A band (Figure 5-3D). These thick filaments are organized into a hexagonal lattice stabilized by M protein\textsuperscript{27} and muscle-specific creatine phosphokinase\textsuperscript{28} in the M line (Figure 5-3D and E). Myosin (Figure 5-3K) is a highly asymmetric 470-kD protein containing two 120-kD globular NH\textsubscript{2}-terminal heads, termed cross-bridges or subfragment-1 (S1) (Figure 5-3L), and an α-helical coiled-coil rod, light meromyosin (Figure 5-3K). Two light chains, essential and regulatory, ranging from 15

---

\textsuperscript{25} Juanqueira LC, Carneiro J, Long JA: Basic histology, ed 5, Norwalk, Conn, 1986, Appleton-Lange.


to 22 kD, are associated with the heavy chain in each S1 (Figure 5-3L). The rod portions of approximately 300 myosin molecules polymerize in a three-stranded helix to form the backbone of each thick filament (Figure 5-3). The cross-bridges, protruding from these backbones, contain ATPase and actin-binding sites responsible for the conversion of chemical energy into mechanical work. Besides their role in muscle contraction, at least 20 classes of nonmuscle myosins accomplish diverse tasks in cell motility such as chemotaxis, cytokinesis, pinocytosis, targeted vesicle transport, and signal transduction. Thus myosin is the target for mutations leading to a number of inherited muscle and neurologic diseases.

Thin filaments (Figure 5-3I) are double-stranded helical polymers of actin that extend 1.1 µm from each side of the Z line and occupy the optically isotropic I band (Figure 5-3D and E). A regulatory complex containing one tropomyosin molecule and three troponin subunits (TnC, TnT, and Tnl) is associated with each successive group of seven actin monomers along the thin filament (Figure 5-3I). In the region where the thick and thin filaments overlap, the thin filaments are positioned within the hexagonal lattice equidistant from three thick filaments (Figure 5-3F). Both sets of filaments are polarized. In an active muscle, an interaction between the two filaments causes a concerted translation of the thin filaments toward the M line that shortens the sarcomere, and thus the muscle fiber and whole muscle (Figure 5-3A through D). Actin is ubiquitous in the cytoskeleton of eukaryotic cells and, like myosin, fulfills many roles in determining cell shapes and motions. Control of the actin cytoskeleton and diseases due to mutations in actin-binding proteins are being intensively investigated.

Two of the largest identified muscle proteins, titin and nebulin, function in assembly and maintenance of the sarcomeric structure. Individual titin molecules (≈3000 kD) are associated with the thick filament and extend from the M line to the Z line. Titin contains repeating fibronectin-like immunoglobulin and unusual proline-rich domains that confer molecular elasticity on the resting sarcomere. Nebulin (≈800 kD) is associated with the Z line and thin filaments. Protein connections from the contractile apparatus through the sarcolemma to the extracellular matrix are described later in this chapter. The cytoskeleton of muscle fibers also contains cytoplasmic actin, microtubules, and intermediate filaments.

**Force Generation and Shortening**

At rest, the thin filament regulatory proteins, troponin and tropomyosin, inhibit contraction (Figure 5-3I). During a twitch, Ca²⁺ released from the SR binds to TnC, relieving this inhibition and thus allowing cross-bridges to attach to actin. A contraction results from cyclic interaction between actin and myosin (the cross-bridge cycle) that produces a relative sliding force between thin and thick filaments. The energy source is the hydrolysis of adenosine triphosphate (ATP) to adenosine diphosphate (ADP) and orthophosphate (P). A simplified model of the chemomechanical events in the cross-bridge cycle is illustrated in Figure 5-4. Motor

---

**Figure 5-4** The actomyosin cross-bridge cycle. Myosin molecules normally have two globular head regions (cross-bridges), but for clarity only one is shown. The closing within the globular domain of myosin represents a hinge point with maximum flexibility. Each head binds with two actin monomers. The sequence of reactions consists of attachment (a); the force-generating transition (b); P releases (c); force generation and filament sliding (d); ADP release (e); ATP binding and detachment (f); and ATP hydrolysis (g). The shadowed heads near detached and force-generating myosin heads indicate high mobility of cross-bridges in these states. ADP, adenosine diphosphate; ATP, adenosine triphosphate; P, inorganic phosphate.
proteins, including myosin, can now be studied by single molecule biophysical techniques, which provide unprecedented detail on their dynamics. When Ca\(^{2+}\) is present, a complex of myosin, ADP, and P\(_i\) attaches to the thin filament (step a), and structural change within the myosin S1 initiates force production and the release of P\(_i\) (steps b and c). The conformational change in the cross-bridge that leads to force generation is a tilting motion of the light chain region. Filament sliding leading to shortening of the sarcomere occurs during a strain-dependent transition between two ADP states (step d). After ADP is released (step e), ATP binds to the active site and dissociates myosin from actin (step f). Myosin then hydrolyzes ATP (step g) to form the ternary myosin–ADP-P\(_i\) complex, which can react with actin for the next cycle.

If the mechanical load on the muscle is high, the contractile apparatus produces a force without changing length (an isometric contraction). If the load is moderate, the thin filaments slide actively toward the center of the sarcomere, resulting in shortening of the whole muscle. The width of the muscle increases during shortening, so the volume stays constant. Work production (concomitant force and sliding) is associated with an increase in the ATPase rate. Thermodynamic efficiency (mechanical power divided by energy liberated by ATPase activity) approaches 50%—a remarkable figure given that manufactured combustion engines seldom achieve efficiencies greater than 20%.

**Relaxation**

The twitch is terminated by reversal of all steps in the activation. Ca\(^{2+}\) released from the SR is taken up again by Ca\(^{2+}\)-ATPase pumps located in longitudinal membranes of the SR. The myoplasmic Ca\(^{2+}\) concentration then decreases, and Ca\(^{2+}\) dissociates from TnC, deactivating the thin filament. When the number of attached cross-bridges declines below a certain threshold, tropomyosin inhibits further cross-bridge attachment, and tension declines to the resting level. Ca\(^{2+}\) diffuses within the longitudinal SR to calsequestrin sites in the terminal cisterna, ready to be released in the next twitch. Myosin continues to hydrolyze ATP at a low rate in relaxed muscle, accounting for a sizable proportion of basal metabolism.

**TRANSMISSION OF FORCE TO THE EXTERIOR**

**Cell-Matrix Adhesions**

The muscle cell is closely connected to the basal lamina along its entire surface. Several transmembrane macromolecular complexes link the myofibrils and actin cytoskeleton to laminins and collagen in the extracellular matrix. Attachment complexes for muscle, analogous to the focal adhesions of motile and epithelial cells and to the adhesion plaques and intercalated disks in cardiac muscle, contain filamentous actin, vinculin, talin, and integrin (primarily the α7β1 isoform), which is the transmembrane link to laminin (Figure 5-5). In muscle, the main laminin isoforms are laminin-2 (α2β1γ1) and laminin-4 (α2β2γ1), which collectively are termed merosin. In addition to providing mechanical coupling between the cytoskeleton and the extracellular matrix, the laminin-integrin system may provide a signaling pathway to regulate localized protein expression. Defects in the expression of many of the cytoskeletal proteins lead to various forms of muscular dystrophy as summarized in Table 5-3.

A specialized linkage between the cytoskeleton and the basal lamina in muscle, complementary to the integrin focal adhesion system, is the dystrophin-glycoprotein complex (see Figure 5-5). Dystrophin, a 427-kD peripheral cytoskeletal protein, has been postulated to function as a mechanical link between the cytoskeleton and the cell membrane or as a shock absorber, or to contribute mechanical strength to the membrane. Its absence or truncation causes Duchenne and Becker muscular dystrophies. The N-terminus of dystrophin binds to actin via a region with sequence homology to the actin-binding domain of α-actinin. This end of the dystrophin molecule may be linked to the basal lamina via the same proteins described previously for focal adhesion-like complexes (see Figure 5-5). The C-terminus binds to a transmembrane dystroglycan-sarcoglycan complex, which in turn binds to the laminins. Muscular dystrophies of varied severity are associated with loss of these components (see Table 5-3). Dystroglycans are also required for early embryonic development of muscle, possibly organizing laminin localization and assembly. Utrophin, a smaller dystrophin-related protein (395 kD), may also link the actin cytoskeleton to the dystroglycans, especially near the neuromuscular junction and in nonmuscle cells. Overexpression of this protein or truncated dystrophin constructs are promising avenues for gene therapy in Duchenne muscular dystrophy. The unusual intricacy of cell-matrix connection systems in muscle may relate to the high forces generated during contraction.

**Myotendinous Junction**

The force of muscle contraction is transmitted to the skeleton via the tendons, which are composed of collagens type I and III, blood vessels, lymphatic ducts, and fibroblasts. At the ends of the muscle fibers, myofibrils are separated by invaginations of sarcolemma filled with long bundles of collagen arising from the tendon. These membrane folds increase the surface area for bearing the mechanical load by approximately 30-fold. Instead of terminating in the Z disks, actin filaments insert into a subsarcolemmal matrix containing α-actinin, vinculin, talin, and integrin. The force is transmitted through laminin to the collagen of the tendon.

**ENERGETICS**

Metabolic pathways in muscle cells are specialized for the variable, at times extreme, rates of ATP splitting by the contractile apparatus and membrane ionic pumps. Among the dozens of metabolic enzymes present, only the most important ones with regard to normal muscle function are mentioned here.

**Buffering of Adenosine Triphosphate Concentration**

The ATP content (5 mM) is sufficient for only a few seconds of contraction, so rapid and effective buffering of
dehydrogenase, which has tissue-specific isoforms. The lactate-dehydrogenase reaction also produces nicotinamide adenine dinucleotide (NAD$^+$), which is necessary for glycolysis, but otherwise lactate is not useful within the muscle. Lactate is freely permeable through the sarcolemma, and a local increase in the extracellular lactate concentration or acidification produces exertional pain (“the burn”). Lactate is transported through the blood to the liver, where it is converted back to pyruvate and then to glucose, which is released into the blood for use by other tissues, such as muscle and brain. This sequence of steps, termed the Cori cycle, transfers some of the high metabolic load to the liver and “buys time” until oxidative metabolism is available.

Oxidative Phosphorylation

In aerobic conditions, pyruvate enters the mitochondria, where it is converted to acetyl-coenzyme A (CoA). Acetyl-CoA can enter the tricarboxylic acid cycle, where it is oxidized to CO$_2$ and H$_2$O, generating reduced NAD (NADH). Fatty acids also contribute to the mitochondrial acetyl-CoA pool through the process of beta-oxidation. The reducing equivalents NADH and FADH$_2$ are subsequently oxidized by the electron transport chain, and an H$^+$ gradient is established across the mitochondrial membrane. This gradient is used to catalyze the phosphorylation of ADP to ATP by the mitochondrial ATP synthase. When glycolysis and oxidative phosphorylation are combined, for example, up to 38 ATP molecules can be generated by the oxidation of each molecule of glucose. This process is energetically much more favorable than the production of lactate, but it can

ATP during contraction is essential for the maintenance of activity. ADP formed by the hydrolysis of ATP is rephosphorylated by transfer of a phosphate group from creatine phosphate (20 mM in a resting cell) by creatine phosphokinase located within the M line of the sarcomere, in the myoplasm, and between the inner and outer membranes of the mitochondria. Adenylate kinase, known in muscle as myokinase, catalyzes the transfer of a phosphate group between two ADP molecules, forming ATP and adenosine monophosphate (AMP). The by-products of the rapid enzymatic reactions that maintain ATP concentration are therefore creatine, P$_i$, and AMP. Some of the AMP is converted to inosine monophosphate by adenylate deaminase. A creatine phosphate shuttle has been proposed to enhance energy flux. According to this hypothesis, creatine phosphate is split within the contractile apparatus and creatine is predominantly rephosphorylated in the mitochondria.

Glycolysis

Muscles use some combination of glucose and fatty acids (and ketone bodies in some situations) as fuel, depending on their metabolic state (fasted vs. fed) and activity status (rested vs. exercise). The muscle compartment contains most of the body storage of glycogen, which is converted to glucose-6-phosphate for local use. Muscle fibers lack glucose-6-phosphatase and thus do not export glucose. During intense activity, especially in anaerobic conditions, the rate of glycolysis and the production of pyruvate exceed the rate of pyruvate consumption by the citric acid cycle. Excess pyruvate is reduced to lactate by lactate dehydrogenase, which has tissue-specific isoforms. The lactate-dehydrogenase reaction also produces nicotinamide adenine dinucleotide (NAD$^+$), which is necessary for glycolysis, but otherwise lactate is not useful within the muscle. Lactate is freely permeable through the sarcolemma, and a local increase in the extracellular lactate concentration or acidification produces exertional pain (“the burn”). Lactate is transported through the blood to the liver, where it is converted back to pyruvate and then to glucose, which is released into the blood for use by other tissues, such as muscle and brain. This sequence of steps, termed the Cori cycle, transfers some of the high metabolic load to the liver and “buys time” until oxidative metabolism is available.

Oxidative Phosphorylation

In aerobic conditions, pyruvate enters the mitochondria, where it is converted to acetyl-coenzyme A (CoA). Acetyl-CoA can enter the tricarboxylic acid cycle, where it is oxidized to CO$_2$ and H$_2$O, generating reduced NAD (NADH). Fatty acids also contribute to the mitochondrial acetyl-CoA pool through the process of beta-oxidation. The reducing equivalents NADH and FADH$_2$ are subsequently oxidized by the electron transport chain, and an H$^+$ gradient is established across the mitochondrial membrane. This gradient is used to catalyze the phosphorylation of ADP to ATP by the mitochondrial ATP synthase. When glycolysis and oxidative phosphorylation are combined, for example, up to 38 ATP molecules can be generated by the oxidation of each molecule of glucose. This process is energetically much more favorable than the production of lactate, but it can
Table 5-3 Classification of Muscular Dystrophies

<table>
<thead>
<tr>
<th>Disease</th>
<th>Genetic Locus</th>
<th>Inheritance</th>
<th>Protein</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duchenne/Becker</td>
<td>XR</td>
<td>AD</td>
<td>Dystrophin</td>
<td>Lethal</td>
</tr>
<tr>
<td>Emery-Dreifuss</td>
<td>XR</td>
<td>AD</td>
<td>Emerin, laminins A and C</td>
<td>40% lethality</td>
</tr>
<tr>
<td>Limb-girdleMuscular Dystrophies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LGMD 1A</td>
<td>5q31</td>
<td>AD</td>
<td>Myotilin</td>
<td></td>
</tr>
<tr>
<td>LGMD 1B</td>
<td>1q11-q21</td>
<td>AD</td>
<td>Laminin A/C</td>
<td>With LGMD, less-severe forms</td>
</tr>
<tr>
<td>LGMD 1C</td>
<td>3p35</td>
<td>AD</td>
<td>Caveolin</td>
<td>can emerge during the first</td>
</tr>
<tr>
<td>LGMD 1D</td>
<td>6q23</td>
<td>AD</td>
<td></td>
<td>three decades, leading to</td>
</tr>
<tr>
<td>LGMD 1E</td>
<td>7q</td>
<td>AD</td>
<td></td>
<td>loss of ambulation after 30</td>
</tr>
<tr>
<td>LGMD 1F</td>
<td>7q22</td>
<td>AD</td>
<td></td>
<td>yr of age. The most severe</td>
</tr>
<tr>
<td>LGMD 1G</td>
<td>4p21</td>
<td>AD</td>
<td></td>
<td>forms start at 3-5 yr of age</td>
</tr>
<tr>
<td>LGMD 2A</td>
<td>15q15.1-q21.1</td>
<td>AR</td>
<td>Calpain 3</td>
<td></td>
</tr>
<tr>
<td>LGMD 2B</td>
<td>2p13</td>
<td>AR</td>
<td>Dysferlin</td>
<td></td>
</tr>
<tr>
<td>LGMD 2C</td>
<td>13q12</td>
<td>AR</td>
<td>γ-Sarcoglycan</td>
<td></td>
</tr>
<tr>
<td>LGMD 2D</td>
<td>17q12-q21.33</td>
<td>AR</td>
<td>α-Sarcoglycan</td>
<td></td>
</tr>
<tr>
<td>LGMD 2E</td>
<td>4q12</td>
<td>AR</td>
<td>β-Sarcoglycan</td>
<td></td>
</tr>
<tr>
<td>LGMD 2F</td>
<td>5q33-q34</td>
<td>AR</td>
<td>δ-Sarcoglycan</td>
<td></td>
</tr>
<tr>
<td>LGMD 2G</td>
<td>17q11-q12</td>
<td>AR</td>
<td>Telethonin</td>
<td></td>
</tr>
<tr>
<td>LGMD 2H</td>
<td>9q31-q34.1</td>
<td>AR</td>
<td>E3-Ubiquitin ligase (TRIM32)</td>
<td></td>
</tr>
<tr>
<td>LGMD 2I</td>
<td>19q13.3</td>
<td>AR</td>
<td>Fukutin-related protein</td>
<td></td>
</tr>
<tr>
<td>LGMD 2J</td>
<td>2q24.3</td>
<td>AR</td>
<td>Titin</td>
<td></td>
</tr>
<tr>
<td>LGMD 2K</td>
<td>9q34</td>
<td>AR</td>
<td>Protein O-mannosyltransferase</td>
<td></td>
</tr>
</tbody>
</table>

CMDs with CNS Involvement

<table>
<thead>
<tr>
<th>Disease</th>
<th>Locus</th>
<th>Inheritance</th>
<th>Protein</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fukuyama CMD</td>
<td>9q31</td>
<td>AR</td>
<td>Fukutin</td>
<td>LE, 11-16 yr</td>
</tr>
<tr>
<td>Walker-Warburg CMD</td>
<td>1p32</td>
<td>AR</td>
<td>O-Mannosyltransferase</td>
<td>LE, &lt;3 yr</td>
</tr>
<tr>
<td>Muscle-eye-brain CMD</td>
<td>1p32-34</td>
<td>AR</td>
<td>O-MNAGAT</td>
<td>LE, 10-30 yr</td>
</tr>
</tbody>
</table>

CMDs without CNS Involvement

<table>
<thead>
<tr>
<th>Disease</th>
<th>Locus</th>
<th>Inheritance</th>
<th>Protein</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merosin-deficient classic type</td>
<td>6q2</td>
<td>AR</td>
<td>Merosin (laminin A2)</td>
<td>Many patients never walk; others have an LGMD pattern</td>
</tr>
<tr>
<td>Merosin-positive classic type</td>
<td>4p16.3</td>
<td>AR</td>
<td>Selenoprotein N1, collagen VI α2</td>
<td>Course stabilizes in late childhood; many continue to walk into adulthood</td>
</tr>
<tr>
<td>Integrin-deficient CMD</td>
<td>12q13</td>
<td>AR</td>
<td>Integrin α7</td>
<td>Presents early in infancy with hypotonia and delayed milestones</td>
</tr>
</tbody>
</table>

Other Dystrophies

<table>
<thead>
<tr>
<th>Disease</th>
<th>Locus</th>
<th>Inheritance</th>
<th>Protein</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facioscapulohumeral</td>
<td>4q35</td>
<td>AD</td>
<td></td>
<td>20% wheelchair bound</td>
</tr>
<tr>
<td>Oculopharyngeal</td>
<td>14q11.2-q13</td>
<td>AD/AR</td>
<td>Polyadenylate binding protein nuclear 1</td>
<td>Onset: ≈48 yr, 100% symptomatic by age 70</td>
</tr>
<tr>
<td>Myotonic dystrophy</td>
<td>19q13.3</td>
<td>AD</td>
<td>DMPK, CCHC-type zinc finger and CNBP</td>
<td>Onset: 50% show signs by age 20; variable severity</td>
</tr>
</tbody>
</table>

AD, autosomal dominant; AR, autosomal recessive; CCHC, cysteine and histidine amino acid sequence in this class of zinc finger; CMD, congenital muscular dystrophy; CNBP, cellular nucleic acid–binding protein; CNS, central nervous system; DMPK, dystrophia myotonica-protein kinase; LE, life expectancy; LGMD, limb-girdle muscular dystrophy; O-MNAGAT, O-mannose β-1,2-N-acetylglucosaminyl transferase; XR, X chromosome related.

occur only when molecular oxygen is available. Myoglobin is an iron-heme complex protein that facilitates oxygen transport within muscle cells. Tissue hydrostatic pressure in a contracting muscle often exceeds arterial perfusion pressure, so the strongest contractions are anaerobic. The content of oxidative enzymes, myoglobin, and mitochondria determines the predominant type of energy metabolism and varies in different muscle types, as was discussed previously (see Table 5-2).

**FATIGUE AND RECOVERY**

During intense or prolonged activity, muscle fatigue is caused by alterations of metabolite levels that suppress force generation at the contractile apparatus, during excitation-contraction coupling, or both. Markedly increased myoplasmic P_i and H+ concentrations and decreased creatine phosphate levels have been detected by magnetic resonance spectroscopy. When the creatine phosphate level declines, maintained activity depends on glycogenolysis until glycogen stores are depleted. During prolonged intense activity, the respiratory and circulatory systems are unable to meet the oxygen demands of the tissue. Force production declines well before the ATP concentration is compromised.

The chemomechanical link between P_i release and force generation (see Figure 5-4) implies that increased myoplasmic P_i in fatigued muscle reduces the magnitude of force simply by mass action. Decreased pH in the muscle, caused in part by lactate accumulation, and insufficient availability of acetylcholine at the neuromuscular junction, leading to failure of synaptic transmission, also contribute
to decreased work production. Because the respiratory and circulatory systems do not supply sufficient oxygen to support metabolism during intense activity, an oxygen debt is incurred. Blood flow and oxygen uptake continue at an enhanced level after the period of exercise to reclaim this energy. Rephosphorylation of creatine can take place within a few minutes, but glycogen resynthesis requires several hours. Recovery processes also involve restoration of ionic gradients across membrane-bound compartments and require consumption of further energy.

**PLASTICITY**

The strength and endurance of a muscle are altered dramatically within weeks after changes in the demands for its use, its mobility, or its hormonal or metabolic environment. The effects of this adaptive response should be considered in any clinical situation that causes a substantial shift in these factors and with regard to the long-term life quality of the patient.

**Adaptation to Muscle Use/Disuse**

Muscle is a use-dependent tissue, meaning that its functional characteristics are closely tied to the number and type of activity patterns that it experiences. Increased muscle use through physical exercise leads to adaptations in muscle fibers that include alterations in specific contractile, regulatory, structural, and metabolic proteins, as well as changes in neural recruitment patterns. The type (aerobic vs. resistance), frequency, intensity, and duration of a training stimulus and the external load influence the adaptive response.

Resistance training causes cross-sectional hypertrophy of primarily fast type II fibers (see Table 5-2) by increasing the size of existing fibers without inducing hyperplasia. This hypertrophy is driven primarily by increased myofilament protein, which occupies the vast majority of fiber volume (=80%), but other myocellular components (e.g., mitochondria) are also increased so as to maintain the relative volume fraction of each cellular component constant with hypertrophy. Aerobic training, on the other hand, enhances the oxidative capacity and volume density of mitochondria in oxidative type I and IIA fibers, and generally does not induce hypertrophy. In this context, resistance training can be seen as altering function by increasing the overall quantity of the muscle, whereas aerobic training alters the functional quality of the muscle to elicit greater endurance to repetitive contraction.

When physical activity is reduced, for instance with bed rest during hospitalization, the cross-section of the fibers decreases, which causes muscle weakness, and endurance is reduced. Of note, these changes are worsened when accompanied by acute and chronic disease, to the extent that up to 10% of muscle protein content can be lost per week.

After significant periods of muscle disuse, loss of muscle strength and endurance can progress to the point where patients are unable to accomplish simple activities of daily living. Although younger, healthy individuals tend to regain muscle size and function readily with exercise rehabilitation, the extent to which older individuals and those with chronic disease can recover from muscle disuse appears to be impaired.

**Hormonal Control**

Hormones, acting in an endocrine or paracrine-autocrine fashion, also have the capacity to alter muscle structure and function. The most prominent hormonal controller of muscle is insulin, which regulates muscle anabolism in response to feeding by suppressing protein breakdown and facilitating protein synthesis via stimulation of amino acid uptake into muscle. In addition to insulin, insulin-like growth factor-I (IGF-I), which largely mediates the effects of growth hormone on muscle, stimulates the hypertrophy of existing muscle fibers by stimulating muscle protein synthesis and inhibiting protein breakdown, and may contribute to muscle growth/regeneration through effects on muscle satellite cells. In men, testosterone has clear anabolic effects on muscle, and conditions that reduce circulating testosterone levels lead to muscle atrophy and weakness. The role of testosterone in women is less certain. Estrogen in women is widely believed to have effects on muscle size and strength, although experimental evidence to support this contention is lacking; and the effects of estrogen on other metabolic processes in muscle appear to be relatively minor at concentrations observed during normal menstrual cycles and in the postmenopausal period.

In many acute and chronic illnesses, changes in the hormonal milieu can affect muscle. The most predominant factors thought to elicit these changes are cytokines and other inflammatory modulators, although alterations in classic stress hormones, such as cortisol and glucagon, also contribute to muscle catabolism. In most cases, the effect of these catabolic hormones is to direct amino acid substrates toward the liver to necessarily support the acute phase protein response. In addition, the type of persistent low-grade inflammation that occurs with aging and in many chronic diseases may have similar detrimental effects on muscle when compounded over extended periods of time. Whether modulation of these catabolic hormonal factors has beneficial effects in maintaining muscle size and functionality under physiologic and pathophysiologic conditions, however, is not certain.

**AGING**

Sarcopenia, defined as loss of skeletal muscle mass and function with age, manifests as an inability to perform simple tasks of everyday life and contributes to disability, a greater risk for falls and fractures, an increase in all-cause mortality, and, in general, a poor quality of life. Normal individuals exhibit an approximately 30% decrease in total muscle mass between the ages of 30 and 80. Although it is universally accepted that whole muscle force production is reduced with age, there is disagreement about whether this is due solely to reduction in muscle mass, or if loss of force production per unit muscle size is also a factor. Most single fiber studies that have measured isometric force production per unit fiber cross-sectional area (i.e., this accounts for age-related reductions in muscle size) show that aging decreases force-producing capabilities. In addition, whole muscle and single fiber studies typically find decreases in contractile velocity with age, and this leads to further reduction in the muscle performance of the elderly, especially for movements requiring high-velocity
contractions. Altogether, these results suggest that aging alters the fundamental contractile properties of skeletal muscle fibers, which decrease, at least in part, whole muscle performance. Age-related changes in motor unit recruitment, activation of agonist and antagonist muscles, and fibrosis can exacerbate single fiber changes and further reduce whole muscle performance.

**SUMMARY**

The complex functional capacity of muscle to produce finely tuned and coordinated movements is ultimately expressed as the transduction of chemical to mechanical energy by actomyosin. A twitch is initiated by an action potential propagated from the central nervous system along an α-motoneuron, neuromuscular chemical transmission, direct protein-protein communication at the T-tubule–sarcoplasmic reticulum junction, Ca\(^{2+}\) diffusion in the myoplasm, and Ca\(^{2+}\) binding to thin filament regulatory proteins. Because the central nervous system controls activity through recruitment of motor units, gradation and coordination of movement depend critically on the pattern of connections between α-motoneurons and muscle fibers and on variations of properties among motor units. Development, maintenance, and aging of the muscular system involve a complex series of genetic programs and cellular interactions that are beginning to be understood at the molecular level. Adaptation of motor unit properties is evident not only in training regimens but also in reduced activity caused by pain or joint immobilization and in compromised metabolic, hormonal, or nutritional conditions. Hence, the plasticity of muscle influences the clinical course of many diseases. In addition to its importance in pathophysiology, muscle serves as an excellent substrate for understanding the molecular basis of cell development, protein structure-function relationships, cell signaling, and energy transduction processes.

**References**


The references for this chapter can also be found on www.expertconsult.com.
Biomechanics combines the field of engineering mechanics with the fields of biology and physiology. Biomechanics applies mechanical principles to the human body. Knowledge of biomechanics is essential in order to understand the intricate interrelationship between mechanical and non-mechanical influences on the musculoskeletal system. An understanding of the loads and properties of the musculoskeletal system is necessary in order to understand the mechanical influences on bone and joint health. Forces that load the joints are generated by muscles and transmitted by tendons. Bones must withstand these forces. Developments in the field of biomechanics have improved our understanding of normal and pathologic gait, mechanics of neuromuscular control, and mechanics of growth and form. This knowledge has contributed to the development of medical diagnostic and treatment procedures. It has provided the basis for designing and manufacturing medical implants, orthotic devices, and rehabilitation therapy. Biomechanics has also been used for improving human performance in the workplace and in athletic competition.

Mechanics is a branch of physics that is concerned with the motion and deformation of bodies that are acted on by mechanical forces. Mechanics is one of the oldest physical sciences dating back to Aristotle (384-322 BCE) with his organized analysis of animal movement. Leonardo da Vinci (1452-1519) worked on the mechanics of the human body, and his detailed anatomic sketches represent the birth of anatomy as a discipline and mechanics as the science governing human motion. Although da Vinci wrote extensively on body mechanics, the man generally credited to be the father of modern biomechanics is Giovanni Alphonso Borelli (1608-1679). His book De motu Animalian provided a quantitative graphical solution to a musculoskeletal biomechanics problem (Figure 6-1).

Engineering mechanics is the discipline devoted to the solution of mechanical problems through the integrated application of mathematic, scientific, and engineering principles. With roots in physics and mathematics, engineering mechanics is the basis of all engineering mechanical sciences. Engineering mechanics is an applied mechanics branch of the physical sciences. The broad field of applied mechanics can be further divided into three main parts: rigid body mechanics, deformable body mechanics, and fluid mechanics. In general, a material can be characterized as either a solid or fluid. Solid materials can then be considered to be rigid or deformable. A rigid body is one that cannot be deformed. In reality, every object undergoes deformation to some extent when acted on by external forces. However, this is a definition of convenience that is used to simplify complex problems. For example, during the study of movement in gait analysis, the bones are considered to be rigid bodies when compared with the soft tissues joining the bones. External loads applied to a rigid body result in internal loads, stresses, and deformations. The mechanics of deformable bodies deals with the relationships between externally applied loads and their internal effects. The mechanics of deformable bodies has strong ties with the field of materials science and is more complex as compared with the analyses required in rigid body mechanics. The purpose of this chapter is to give you only a brief idea of biomechanics. Our focus is on rigid body mechanics.

Basic biomechanics relies heavily on Newtonian mechanics. These laws were introduced by Sir Isaac Newton and form the basis for analyses in statics and dynamics. Statics analyzes the forces that occur in rigid bodies that are in static equilibrium. Dynamics is the study of bodies in motion. The general field of dynamics consists of two major areas: kinematics and kinetics. Ultimately, proper joint constraint and stability enable the function of limbs in characteristic ways.
KINEMATICS

Kinematics is the study of the geometric and time-dependent aspects of motion without analyzing the forces causing the motion. Kinematic analysis is used to relate displacement, velocity, acceleration, and time. To study kinematics in an organized manner, it is common to classify the motion as translational, rotational, or general. Translational motion occurs when a straight line drawn between two points on the body remains in the same direction during the motion. Translational motion can be either rectilinear motion (if the paths are straight lines) or curvilinear (if the paths are curved lines). Rotational motion occurs when the points on the body move in a circular path around an axis of rotation. The angular motion occurs about a central line known as the axis of rotation, which lies perpendicular to the plane of motion. The third class of motion is called general motion, or displacement, which occurs if a body undergoes both translational and rotational motion simultaneously.

Kinematics can be analyzed in two-dimensional (2-D) or three-dimensional (3-D) space. When all points of a rigid body move parallel to a plane, the motion is referred to as planar motion. This motion can be thought of as 2-D motion. 3-D motion is the more general type of rigid body motion. This motion requires six independent parameters to describe the general type of motion. These parameters are called “degrees of freedom” or the number of independent coordinates in a coordinate system that is required to completely specify the position of an object in space. A rigid body in space has a maximum of six degrees of freedom: three translations (expressed by linear coordinates) and three rotations (expressed by angular coordinates). The general movement of an object is defined by a vector quantity that is a combination of both linear and angular displacement. Velocity is the time-related change of displacement. Linear velocity is expressed in units of length per time (m/sec). Angular velocity is expressed in units of angular measure per time (rad/sec). Because velocities reflect vector quantities, both magnitude and direction must be specified. Acceleration is the time rate of change of velocity. Linear acceleration is expressed in units of length per time squared (m/sec²). Angular acceleration is the time rate of change in angular velocity (rad/sec²). Accelerations are also vector quantities, and both magnitude and direction must be specified.

Kinematic techniques have been used to study body movements in both 2-D and 3-D space. The human body is typically modeled as a number of interconnected rigid body segments (Figure 6-2). A coordinate system is affixed to each rigid body segment in order to establish an anatomic coordinate system. External markers are used to define orthogonal coordinate systems whose axes define the position of these body segments. Joint motion is then described as the relative motion of the distal body segment with respect to the proximal body segment. Limb segments are assumed to undergo angular displacement during human movement. However, more sophisticated analyses will also quantify the linear displacement that limb segments may undergo. These measures of relative segmental angles have been used to describe human walking and other activities of daily living.26

As an example, knee motion in the sagittal plane can be characterized throughout the gait cycle (Figure 6-3). At
heel strike, the knee is nearly fully extended (knee flexion of 5 degrees). During midstance, the knee flexes to about 15 degrees, which occurs at 15% of the gait cycle. The knee joint is brought back into extension by midstance. At 50% of the gait cycle, opposite foot contact occurs. The weight is shifted to the opposite limb, and the knee begins flexing. Toe-off occurs at 60% of the gait cycle. Peak knee flexion of 60 degrees occurs during the swing phase. The knee motion can be described as two flexion waves, each starting in relative extension, progressing into flexion, then returning again to extension. The first flexion wave, or stance phase knee flexion, acts as a shock absorber to aid weight acceptance. This curve peaks in early stance at opposite foot-off. The mechanical source for this shock absorber is the eccentrically contracting quadriceps muscles. The second flexion wave is necessary in order to clear the foot in early swing phase. The knee is rapidly flexed beginning just after heel rise to a maximum in swing phase just as the swinging foot passes the opposite limb.

The complexity of kinematic analysis increases substantially when going from planar analysis to 3-D analysis. The complexity of the analysis arises from the technical difficulty that large rigid body rotations cannot be treated as vectors and, hence, do not obey the vectorial principles of transformation, independence, and interchangeability of operations. For finite spatial rotation, the sequence of rotations is extremely important and must be specified for a unique description of joint motion. For the same amount of rotation, different final orientations will result from different sequences of rotation (Figure 6-4). However, with proper selection and definition of the axes of rotation between two bony segments, it is possible to make finite rotation sequence independent or commutative. The concept of Eulerian angles has been adopted in the field of orthopedic biomechanics to unify the definition of finite spatial rotation. In the selection of reference axes, one axis is fixed to the stationary segment and another axis is fixed to the moving segment (Figure 6-5). In the knee joint, for example, the flexion/extension angle, $\Phi$, occurs about a medial-laterally directed axis defined by a line connecting the medial and lateral femoral condyles. The axial rotation angle, $\Psi$, is measured about an axis defined by the line along the shaft of the tibia. The third axis, also defined as the floating axis, is orthogonal to the other two axes and defines abduction/adduction, $\Theta$. These rotations match the Eulerian angle description and are thought to be performed in such a way as to bring the moving segment from the reference orientation into the current orientation. The advantage of using this system for description of the spatial rotation of anatomic joints is that the angular rotations do not have to be referred back to the neutral position of the joint because the rotation sequence can be independent.
Kinetics is the study of the forces that cause motion of a rigid body. When there are unbalanced forces or moments acting on a rigid body, it is under a nonequilibrium, or dynamic, condition, resulting in motion. Understanding the kinetics of human movement provides a fundamental understanding of the musculoskeletal system. Before one can begin to analyze the forces during human movement, some basic definitions and assumptions must be made.

The key quantities in kinetics are force, moment, and torque. Forces between two bodies can be contact forces (bodies touching each other) or field forces (bodies separated by a distance, such as gravitational, electric, or magnetic forces). Forces are represented by vectors. Vectors are composed of four components: magnitude, direction, sense, and position (also called point of application). According to Newton’s Second Law, force is any action that tends to change the state of rest or state of motion of a body to which it is applied. Forces are represented by vectors. The vector may be resolved into several component forces, usually along specified mutually perpendicular coordinate axes. Conversely, forces can be summed using vectorial addition. A moment represents the turning, twisting, or rotational effect of a force. A moment is a vector. A moment is defined as the product of the force and the perpendicular distance between the line of action of the force and the axis of rotation of the motion that the force produces (Figure 6-6). Its magnitude is the force times the perpendicular distance to the axis of rotation. The direction of the moment is along the axis of rotation (or potential rotation) and thus perpendicular to the plane in which the twisting force is applied. The moment arm, the distance used to calculate the moment, is the shortest distance from the force action line to the actual or potential pivot point of the system, regardless of the state of motion. Skeletal
motions are the result of moments applied by muscles that cross the joints on which they act. Moments of a force about an axis measure the tendency of the force to impart to the body a motion of rotation about a fixed axis. A torque is a special type of moment that results when a pair of forces that have equal magnitude, parallel lines of action, and opposite senses act on a body (Figure 6-7). The magnitude of the torque is \( Fd \), where \( d \) is the perpendicular distance between the two forces. The resultant force is zero because the two forces are equal and oppose each other.

Kinetics can be used to analyze forces affecting the musculoskeletal system. These forces can be classified as either external or internal. External forces represent the action of objects contacting the body, gravitational forces, or force due to inertia of the body. Internal forces are the body’s responses to external forces. Internal forces consist of muscle, ligament, and joint contact forces. In general, the limb segments are assumed to be rigid. This simplifies the analysis because the structure is assumed to not deform under load. Further, joints are assumed to be frictionless hinges.

Statics is the study of forces acting on a body at rest. When performing a force analysis, the body or part of a body at equilibrium may be isolated from the environment and the environment is replaced by forces acting on the system. This is called a free-body diagram. Because both forces and moments are vectors, they must sum to zero in each of the three perpendicular directions (reference system). Consider a person standing quietly (Figure 6-8). The person’s weight (force of gravity) tends to pull the person toward the ground (Figure 6-8A). The person does not move downward because the ground is pushing up with a total force equal in magnitude to the individual’s weight (Newton’s Third Law). If the person has his or her weight distributed symmetrically, then their weight is evenly shared by each lower extremity (see Figure 6-8A). The load under each foot can be represented by resultant ground reaction force (GRF) vectors with one half of the body weight supported by each foot. With one half of body weight supported by each foot, the point of application of the resultant GRF passes approximately midway between the subject’s two feet (Figure 6-8B). There is no motion because the external forces (i.e., body weight and GRF) are balanced—they are equal and opposite in magnitude. When the person leans to one side (Figure 6-8C), the GRF shifts in the direction that the person leans. Thus overall body posture can affect the GRF location. If the person leans more, he or she will become unstable (i.e., the downward body weight vector will fall outside the base of support). In order to remain in static equilibrium, the subject requires additional support (Figure 6-8D). The horizontal force applied to the upper body by the wall is balanced with an equal and opposite horizontal GRF component. Also, the two equal and opposite vertical forces are no longer aligned (i.e., collinear). They form a force couple that would tend to rotate the body in a counterclockwise direction. A second clockwise force couple formed by the two equal and opposite horizontal forces balances the counterclockwise force couple. As a result, the subject remains in both translational and rotational static equilibrium.

When there are unbalanced forces or no one is acting on a rigid body, it is under a nonequilibrium, or dynamic, condition resulting in motion. Newton’s Second Law of Motion links the kinematics of a body to its kinetics. The Second Law states, “If the resultant force acting on a body is not zero, the body will have an acceleration proportional to the magnitude of the resultant and in the direction of this resultant force.” Analysis of the motion of the limb segments requires a set of governing equations and assumptions. The equations assume that each limb segment is a rigid body moving in 3-D space. Therefore six scalar equations of motion define the general 3-D forces and motion of each limb segment:

\[
\sum F = ma \\
\sum M = I\alpha
\]
motion depends on the shape of the joint surfaces. For example, the hip is a congruent ball-and-socket joint, the elbow is a congruent hinge joint, the knee and proximal interphalangeal joints are bicondylar joints, and the carpometacarpal joint of the thumb is a saddle joint. In general, the joint articulating motion can be described in terms of sliding, spinning, and rolling (Figure 6-10). Both sliding and spinning motions involve the relative translation of one surface against the other. Rolling has the least relative motion between articulating surfaces. The translational speeds range from 0.06 m/s between the femoral head surface and acetabular surface during normal walking to 0.6 m/s between the humeral head surface and glenoid surface during baseball pitching.

Although diarthrodial joints experience an enormous amount of loading and motion, the cartilage surfaces undergo little wear and tear during normal conditions. This is due to the special lubrication qualities of synovial fluid and the biphasic structure of the cartilage. Synovial fluid, secreted by the synovium into the joint space, contains mainly hyaluronate. Biomechanically, hyaluronate is a highly viscous liquid in which the shear stress depends on the rate of shear strain applied. As the shear rate increases, the viscosity of synovial fluid decreases. However, in the rheumatoid patient, due to enzymatic degradation of the hyaluronate molecules, the viscosity of the rheumatoid synovial fluid does not have a normal shear rate dependency and is believed to be less effective in lubrication.

JOINT BIOMECHANICS

Diarthrodial joints connect long bones to allow force transmission and joint rotation. The type of joint articulating motion depends on the shape of the joint surfaces. For example, the hip is a congruent ball-and-socket joint, the elbow is a congruent hinge joint, the knee and proximal interphalangeal joints are bicondylar joints, and the carpometacarpal joint of the thumb is a saddle joint. In general, the joint articulating motion can be described in terms of sliding, spinning, and rolling (Figure 6-10). Both sliding and spinning motions involve the relative translation of one surface against the other. Rolling has the least relative motion between articulating surfaces. The translational speeds range from 0.06 m/s between the femoral head surface and acetabular surface during normal walking to 0.6 m/s between the humeral head surface and glenoid surface during baseball pitching.

Although diarthrodial joints experience an enormous amount of loading and motion, the cartilage surfaces undergo little wear and tear during normal conditions. This is due to the special lubrication qualities of synovial fluid and the biphasic structure of the cartilage. Synovial fluid, secreted by the synovium into the joint space, contains mainly hyaluronate. Biomechanically, hyaluronate is a highly viscous liquid in which the shear stress depends on the rate of shear strain applied. As the shear rate increases, the viscosity of synovial fluid decreases. However, in the rheumatoid patient, due to enzymatic degradation of the hyaluronate molecules, the viscosity of the rheumatoid synovial fluid does not have a normal shear rate dependency and is believed to be less effective in lubrication.
Figure 6-9  Schematic of a person in single support during gait. A, External demand. The body weight and vertical ground reaction force (GRF) couple tend to rotate the body clockwise. The medial shear GRF and the corresponding lateral inertial force produce a force-couple that tends to rotate the body counter-clockwise. Thus the body is in dynamic equilibrium. B, Internal response. The resultant GRF passes medial to the knee joint, creating an external knee adduction moment. The external moment is balanced internally in the knee by creating a larger force on the medial tibial plateau (M) than on the lateral tibial plateau (L). Once again, the knee is in dynamic equilibrium.

Figure 6-10  The motions of joint articulation: spinning, rolling, and sliding. (From Morrey BF, Itoi E, An KN: Biomechanics of the shoulder. In Rockwood CA, Matsen F, editors: The shoulder, ed 2, vol 1, Philadelphia, 1998, WB Saunders, pp 233–276.)
The human body can perform complex motions because the joints allow multiple degrees of freedom between articulating bones. Due to the variations in anatomic structures of the joints, various limbs have different movement characteristics and load transmissions. Adequate joint stability that balances motion and loads is necessary for a joint to function properly. Chronic disease or traumatic injury may cause damage to joint tissue, thus compromising the joint constraints and stability. To improve the diagnosis and treatment of such joint disorders, it is essential to understand the basic mechanics of the constraint, which provides stability. The anatomic structure responsible for joint constraint can be divided into passive and active elements. Passive elements, consisting of the capsulo-ligamentous structures and bony articulating surfaces, provide static constraints of the joint. The active elements include muscle-tendon units, which provide dynamic constraints of the joint.

The contribution of capsulo-ligamentous structures to joint stability and constraint is determined by the change in length, the line of action, and the material properties of the tissue. For a given loading and displacement of the joint, the greater the elongation of the soft tissue, the higher the passive tension generated by that tissue. However, the specific contribution of such passive tension to joint constraint further depends in turn on the relative line of action of the capsulo-ligament structures. Therefore the relative locations of the origins and insertions of the ligaments may determine the contribution of the ligamentous tension in resisting joint displacement in three orthogonal directions. Surgical alterations of the soft tissue may cause damage to joint tissue, thus compromising the joint constraints and stability.

Two lubrication mechanisms in the synovial joint provide minimal friction and cartilage wear: fluid film lubrication and boundary lubrication. During joint rotation, the sliding speed of the articulating surfaces and the viscosity of the synovial fluid create a thin film, which is capable of supporting the load, resulting in fluid film lubrication. When the joint is loaded, the two articulating surfaces approach each other and a squeeze film is generated to support the load. The pressure developed in the lubricating fluid carries the load applied to the joint. The thin film of lubricant also produces a greater bearing surface. In addition, the cartilage is a biphasic porous medium. Water is bounded in the intrafibrillar space of the collagen matrix by proteoglycans. Intricate interaction of the mechanical stress, the electric charges, and the hydronamics results in a special fluid efflux pattern known as elastohydrodynamic lubrication. The fluid exudation ahead and imbibition behind the moving contact point of the articulating surface further facilitates the lubricant film.

On occasion, heavy loading exceeds the capacity that the lubricant film can support. In this situation, the cartilage surfaces are in direct contact. Under this condition, lubrication is provided by the mechanism of boundary lubrication. Boundary lubrication is accomplished by a monolayer of glycoprotein called lubricin, a superficial zone protein. In the pathologic condition of osteoarthritis, the structure properties of cartilage such as porosity and permeability are altered. The fluid efflux and the synovial fluid film are diminished, and boundary lubrication is the remaining mechanism of joint lubrication.

**JOINT CONSTRAINT AND STABILITY**

The contribution of capsulo-ligamentous structures to joint stability and constraint is determined by the change in length, the line of action, and the material properties of the tissue. For a given loading and displacement of the joint, the greater the elongation of the soft tissue, the higher the passive tension generated by that tissue. However, the specific contribution of such passive tension to joint constraint further depends in turn on the relative line of action of the capsulo-ligament structures. Therefore the relative locations of the origins and insertions of the ligaments may determine the contribution of the ligamentous tension in resisting joint displacement in three orthogonal directions. Surgical alterations of the soft tissue may cause damage to joint tissue, thus compromising the joint constraints and stability.
attachment to the bone would therefore change the constraints of the joint.

The lengthening and shortening of the capsules-ligamentous structures during joint loading and movement determine the deformation of the tissue. The material properties of the tissue along with the amount of the deformation will determine the passive tension of the structure. For the same amount of deformation, the stiffer the tissue, the higher the tension will be; the softer the tissue, the lower the tension. Alterations of the material properties due to physiologic and pathologic conditions will definitely alter the joint constraint and the articulation. For example, the tight posterior capsules of the baseball pitcher would limit the range of physiologic motion. On the other hand, the softening of capsules-ligamentous structures during pregnancy and in rheumatologic diseases could all potentially lead to abnormal motion and contact stress on the joint surface. These alterations of the soft tissue material properties must be considered as essential factors in initiating the early degeneration of joint cartilage and may lead to the vicious cycle of arthritis.

Experimentally, the significance of various passive anatomic structures to joint stability has been assessed using two methods; the stiffness test and the laxity test. The stiffness test is commonly used to determine the relative contributions of individual anatomic elements to joint constraint. The laxity test is considered to demonstrate the outcome of joint instability when one or more constraining units are compromised. In the stiffness test, the joint is displaced under controlled testing and the corresponding load is monitored. On the basis of the joint load-displacement curves, it is obvious that the load required to displace the joint is higher when all the ligaments are intact than when a ligament is lacerated. The difference in the constraining load between these two curves at the same joint displacement represents the contribution to the joint constraint of that particular sectioned ligament. Because the capsules-ligamentous structures are passive tissues, as long as the joint displacement is controlled and experimentally reproduced, the relative contributions of individual tissues will not be dependent on the sectioning sequence of soft tissue. Contrary to the stiffness test, in which displacement is varied and load measured, in the laxity test, a specific load is applied to the joint and the extent of joint displacement is measured. Joint laxity is documented with changes in displacement as anatomic elements are removed. The joint laxity as observed and measured would provide more of a clinical scenario in that joint instability would be experienced by the patient who has encountered a similar soft tissue injury.

In addition to the soft tissue constraints, the joint articulating surface is another element in providing joint constraint and stability. Theoretically, its significance to joint stability is determined by both the geometric shapes of the joint surfaces and the amount of compressive force encountered between those two surfaces (Figure 6-13A). When two curved surfaces articulate, a vertically applied force will induce either a transverse displacement or a transverse constraining force. The greater the congruency between the curved articulating surfaces, the greater the induced lateral joint constraint force. This mechanism is sometimes referred to as concavity-compression. In the shoulder joint, for example, the relative translations between the glenoid and the humeral head and the forces resisting translation were recorded. The stability ratio, defined as the peak translational force divided by the applied compressive force, was higher in the hanging-arm position than in glenohumeral abduction (Figure 6-13B). The highest stability ratio was detected in the inferior and superior directions. The anterior direction was associated with the lowest stability ratio. Resection of the glenoid labrum resulted in an average decrease in the stability ratio of 10%. In the treatment of traumatic recurrent anterior shoulder instability, patients with a large osseous glenoid defect are at risk for recurrent instability after arthroscopic Bankart repair. Bone-grafting performed in shoulders with a large osseous glenoid defect
to restore the articulating surface provides excellent clinical outcomes.

In order to benefit for joint constraint through the concavity-compression, the compressive or oppositional force across the joint surface is required. This joint compressive force is usually generated by the muscle contraction forces. The relative location and direction of the resultant joint force to the available articulating surfaces will determine the potential stability of the joint. For example, in the glenohumeral joint, if the resultant force is well located at the center of the glenoid articulating surface, the joint is considered stable. When the resultant force is directed outside the available joint surfaces, joint dislocation is most likely. However, when the resultant force is located within the joint surfaces but more toward the rim of the glenoid surface, there is potential for subluxation of the joint. The interaction of muscle force resulting in the resultant joint constraint forces is illustrated later.

Intra-articular or intra-capsular pressure is the other component of the passive and static constraint of the joint (Figure 6-14). The joint is usually surrounded by a capsule and sealed with synovial tissue. When the capsule is intact, any distraction force stretches and deforms the capsuloligamentous structure and will generate negative intra-articular pressure. This negative intra-articular pressure will counteract further displacement. When the joint capsule is vented, such negative pressure can no longer develop and cannot constrain the joint. The negative intracapsular pressure is there for the constraint of all joints. However, it is especially important for those joints under gravitational distraction such as the shoulder joint.

Active muscle contraction is important for joint stability. Proper coordination of muscle contraction could provide dynamic joint constraint. On the other hand, deficiency of muscle innervation or defect of tendon could result in unfavorable joint constraint and lead to joint instability. That is, the muscle force can act not only as a joint stabilizer but possibly a joint dislocator as well. The contribution of muscle/joint constraint is similar to that of passive capsuloligamentous tension, theoretically determined by the line of action along the passive and active tension of the muscle. Based on the relative position of the bony structure and the direction of the joint displacement, the same muscle-tendon structure can contribute differently to joint constraint. When the joint is spanned by multiple muscles, the coordination of muscle contraction would ultimately determine the dynamic stability of the joint. For example, the line of action of the muscle was measured and the contributions of muscle contractions to the articulating surfaces of the shoulder were described in the normal and shear directions (Figure 6-15). The shear component has direct influence on the transverse translation of the bony element at the joint. In contrast, the normal component of the muscle force provides the compressive force to bring the surfaces together. A glenohumeral joint with a lax capsule and loose ligaments might be dynamically stabilized if the glenoid concavity is maintained and the function of the external and internal rotators, efficient stabilizers in the position, is enhanced. On the other hand, in the patient with rotator cuff tear, the pull of the major movers such as deltoids during arm elevation of forward rotation will result in the resultant joint force into the superior and anterior directions, and superior impingement and anterior instability are thus expected.

In general, muscle contraction will generate moments about all three axes of rotation of a given joint. However, when the constraints from other periarticular soft tissues are not available, the muscle will be primarily responsible for maintaining the balance of the joint’s rotation. If a muscle crosses multiple joints, the recruitment of muscle activities depends on the moments of all the joints it spans. The concept of muscle with multiple joints is of extreme clinical importance. For example, when considering rehabilitation after reconstructive surgery of the anterior cruciate ligament, the goal is to strengthen the muscles without excessive tension being placed on the ligament. Quadriceps loading, especially with the knee in a near extended position, would result in anterior joint shear force and displacement of the tibia. Such shear force would induce tension in the anterior cruciate ligament. The concept of kinetic-chain has been used extensively to address this problem. In fact, the principle is simply based on two-joint muscles. During closed-kinetic-chain type exercises such as the squat and the leg press, the foot is fixed and motion at the knee is accompanied by motion at the hip and ankle joints. The GRF causes flexion moments at both the knee and the hip joints. The antagonistic muscle of hamstrings at the knee

Figure 6-14  Effect of negative intra-articular pressure on shoulder joint stability. In the normal situation, traction on the arm causes an increase in negative intra-articular pressure and joint distraction is constrained. On the contrary, if the capsule is vented and air or fluid is introduced into the joint, traction on the arm no longer affects the intra-articular pressure, but instead, the joint will be subluxed inferiorly.
joint is thus recruited to balance the hip flexion moment. Co-contraction of quadriceps and hamstrings at the knee joint reduces the anterior joint shear force and the tension in the anterior cruciate ligament.²¹

References

1. Borelli GA: De motu animalium, Batavis, 1685, Lugduni.

Figure 6-15  Coordination of the muscle contractions to accomplish a given task can be well demonstrated by the situation that the deltoid and rotator cuff muscles work together to elevate the shoulder joint. A, For patients with rotator cuff tear, the shoulder elevation will be accomplished by the deltoid muscle alone. The resultant joint force would be directed superiorly due to the line of action of the deltoid muscle. Under this situation the joint is not stable; therefore superior subluxation and migration would be expected. B, On the other hand, when elevating the shoulder with the rotator cuff muscle alone, the direction joint resultant force would point toward the center of the glenoid surface, which would make the joint stable. However, the magnitude of the joint resultant forces is relatively high, which may not be ideal for the cartilage tissue. C, Ideally, shoulder elevation is accomplished with the combination of the deltoid muscles as the workhorse to provide the strength and power, along with the rotator cuff muscles as the stabilizer in directing the joint resultant force to keep the joint in a more stable condition.
KEY POINTS

An understanding of the mechanisms of postnatal tissue repair provides potential therapeutic targets to enhance intrinsic tissue regeneration.

Because tissue repair mimics the cellular and molecular cascade of embryonic tissue formation, it is anticipated that investigating the role of developmental pathways in joint and skeletal disorders/diseases will identify novel therapeutic targets.

Cellular therapeutics and their related products are complex with regard to mechanism of action and manufacture; this has evolved as a concept into advanced therapy medicinal products (ATMPs) with a specific regulatory path.

Recent advances in regenerative medicine and tissue engineering relevant to rheumatology have entered clinical practice and include the repair of joint surface defects by autologous chondrocyte implantation and bone repair using adult mesenchymal stem cells.

Tissue engineering is adopting the concept of biomimetics of in vivo tissue development. Developmental engineering is the term used to describe this novel method for the rational and accurate design of robust, well-controlled manufacturing processes of "biological spare parts."

With dramatic advances in targeted treatments for arthritic disease, inflammation can be kept under control quite efficiently. In addition, a better understanding of the destructive processes beyond inflammation in arthritic disease has identified new molecular targets, such as metalloproteinases, and cellular targets, such as the osteoclasts. These findings in turn have led to the development of new treatment approaches that effectively contribute to improved control of the destruction of joint and joint-associated tissues. The new challenge in arthritic disease is to position all these therapeutic opportunities properly, ultimately leading to a stratified approach (i.e., giving optimal treatment to the proper patient at the right time). New principles have emerged in the management of arthritic disease, such as early detection, remission induction and maintenance therapy, tight disease control, patients at risk and responders to treatment; these have become a part of daily clinical practice.1-3

These developments and in particular the widespread use of new biologics such as tumor necrosis factor (TNF) inhibitors have highlighted the need to consider other aspects of joint biology, in particular the mechanisms driving tissue response and repair.1 Indeed, to restore the balance between tissue destruction and tissue repair (Figure 7-1), we should be looking arguably at the overview picture (i.e., the "systems biology" of the joint). Introducing regenerative medicine provides a novel opportunity to restore joint homeostasis, thus possibly leading to cure. Targeting repair has entered our discipline, and investigating the potential to activate and enhance joint tissue repair mechanisms has become a prime goal.

Regenerative medicine, encompassing tissue engineering, seeks to repair or regenerate damaged tissues and organs, regardless of the cause of the insult, without leaving scar tissue behind, thereby restoring both structure and function of tissues/organs. Nature demonstrates that this is achievable because successful wound healing and fracture repair typically are processes that happen routinely in the postnatal individual, even until advanced age. Indeed in many patients, it is not possible to detect after the healing process where the skin wound or fractured bone site was. We also know, as demonstrated in fetal surgery, that this scarless repair is partially age and context dependent. Thus it is attractive to envision that with an improved and in-depth understanding of repair processes at the cellular and the molecular level, we may be able to interfere quickly at the time of injury to guide the healing process appropriately, thereby preventing scar formation. Postnatal tissue healing mimics the developmental processes of tissue formation.

As an example of, and of relevance for, skeletal tissues, it appears that the process of rebuilding an adult limb has many similarities with how the limb is formed a priori in the embryo; signaling mechanisms are required to specify the final pattern. Thus limb formation and limb regeneration are likely to employ similar molecular pathways.4 Remarkable advances in developmental biology over past decades have provided the knowledge platform to advance into novel regenerative medicine approaches in postnatal life. These advances include not only improved understanding of the mechanisms of body axis formation and organogenesis, but also impressive progress in stem cell biology, including the regulation of stemness and stem cell niches, lineage specification and cell differentiation, and the critical molecular pathways involved.

In view of this, we are now in a position to enter a new era in regenerative medicine and tissue engineering.5,6 In this chapter, we will briefly review approaches seeking to repair damaged and diseased tissues, in particular diarthrodial joints and skeletal structures.

When we seek to repair tissues, two mechanistic approaches are possible: (1) enhancing intrinsic repair mechanisms through stimulation of cell proliferation, differentiation, and tissue metabolic activity, while recruiting endogenous progenitor populations into the damaged tissue;
and (2) if insufficient intrinsic repair leads to clinical symp-
toms and signs, with loss of function, extrinsic repair needs
to be considered (i.e., tissue engineering approaches gener-
ating cell populations and combination products that can
contribute mostly locally to tissue repair processes). We will
now discuss these approaches in greater depth.

**INTRINSIC TISSUE REPAIR**

As the body strives to maintain homeostasis of postnatal
tissues and thus undergoes continuous tissue remodeling,
many signals counteract destructive processes in tissues.
Disease processes may allow destructive processes to become
dominant, leading to loss of tissue homeostasis and loss of
function. Enhancement of expression/secretion of signals
counteracting the breakdown processes is a straightforward
approach to restore homeostasis. Naturally occurring exam-
ple includes production of soluble receptors or receptor
antagonists (sTNF receptors, interleukin [IL]-1RA), use of
inhibitors of matrix modeling enzymes such as tissue inhibit-
ors of metalloproteinases (TIMPs), and inhibition of osteo-
clast formation caused by blocking the receptor activator
of nuclear factor κB (NFκB) (RANK)/RANK ligand
(RANKL)/osteoprotegerin (OPG) system. Some of these
targets have reached the clinic; others are in early or late
phase clinical trials across a range of diseases, including
rheumatoid arthritis (RA), psoriatic arthritis (PsA), and
osteoarthritis (OA). Note, however, that in some arthritic
diseases such as ankylosing spondylitis, the tissue response
is abnormal, leading to joint ankylosis (for review, see Lories
et al.1)

Because postnatal tissue repair mimics in many aspects
tissue formation during development, strong arguments
have been put forth for the belief that developmental path-
ways involved in formation of the joint and joint-associated
tissues may be interesting targets for enhancing tissue repair.

Typical developmental pathways critically involved in the
formation of skeletal tissues and structures such as the syno-
vial joints are contained in the transforming growth factor
(TGF)-β superfamily, including bone morphogenetic pro-
teins (BMPs) and growth differentiation factors (GDFs),
the Wnt family of proteins, fibroblast growth factors (FGFs),
the hedgehog (Hh) proteins, parathyroid hormone (PTH)/
PTH-related protein (PTHrP), and notch signaling. It is not
the intention of this chapter to discuss in detail these path-
ways, but rather to select a few examples of how these sig-
naling pathways are involved in joint formation and to
discuss their role in postnatal joint repair. Specific focus is
placed on some of those pathways that have led to advances
in early clinical development of arthritic disease.

**TGF-β/BMP Signaling**

Recent reviews have highlighted the relevance and critical
role of members of the TGF-β superfamily (TGF-β, BMPs, 
GDFs) in the biology of articular cartilage, bone, joints, and
joint-associated tissues, during development, in postnatal
joint tissue homeostasis and repair, and in tissue response on
injury and aging. TGF-β has been shown to be involved in
the maintenance and aging of articular cartilage and in
osteoarthritis.2 BMPs have been reported to play major roles
in articular cartilage metabolism, with BMP7/OP1 being of
particular interest.3 In addition, data indicate that modu-
lation of TGF-β/BMP downstream receptor-Smad signaling
plays an essential role in both the regulation of chondrocyte
differentiation and the development and progression of
osteoarthritis.4 It is therefore not surprising that because of
overwhelming evidence of the regenerative potential of this
family of growth and differentiation factors in preclinical
models, their therapeutic potential is being explored. In
particular, several ongoing phase I studies on osteoarthrosis
of the knee are seeking to target OP1/BMP7 (clinicaltrials.
gov). In addition, a search is ongoing, through peptide tech-
nology (peptidomimetics) or small molecule screens, to
identify modulators of the TGF-β/BMP receptor/Smad sig-
naling pathways.5 Because the nature of the synovial joint
allows for local treatment, it is expected that some of the
newly identified compounds will first be tested in local
application such as joint surface repair. Indeed because this
family of proteins, in particular the BMPs, plays a role in
many systemic processes, BMP technology may be required
to focus in the first place on local treatment, as exemplified
in the use of BMP devices already in the clinic for orthope-
dic applications such as spine fusion and healing of non-
union.6,7 Wider systemic morbidity upon targeting BMP
pathways is as yet unclear.

**FGF Signaling**

Extensive investigations have identified FGFR3 signaling as
a key regulator of chondrocyte and osteoblast function both
during development and postnatally. In particular, absence
of signaling through FGFR3 in the joints of Fgfr3-/- mice
leads to premature cartilage degeneration and early arthri-
tis.8 These degenerative changes were accompanied by
increased expression of MMP13 and type X collagen, cel-
lar hypertrophy, and loss of proteoglycan at the articular
surface. One of the key ligands of FGFR3 signaling appears
to be FGF18 (for review, see Haque et al\textsuperscript{17}). In the postnatal joint, FGF18 has significant anabolic effects on cartilage metabolism. The efficacy of FGF18 was tested in a rat meniscal tear model of OA.\textsuperscript{18} Intra-articular injection of FGF18 induced a dose-dependent increase in cartilage formation and a reduction in cartilage degeneration scores in the medial tibial plateau of OA rats. It is important to note that this effect was seen only in OA joints, not in normal rat joints, suggesting that this effect may be a specific response to tissue injury. It is interesting to note that at the molecular level, this joint protective effect may be due in part to its interaction with other signaling pathways such as BMP signaling by repressing noggin, a BMP antagonist.\textsuperscript{19} These findings, combined with enhanced understanding of FGF18 signaling in postnatal cartilage and bone biology, led to the development of local treatment with FGF18 and the design of early-phase clinical trials (clinicaltrials.gov) in patients with cartilage lesions and in those with OA of the knee.

### Wnt Signaling

Recent findings indicate a critical role for Wnt signaling in cartilage and bone biology, with specific relevance to osteoporosis and osteoarthritis (for review, see Luyten et al\textsuperscript{20}). Most available data provide circumstantial and/or direct evidence, both in vivo and in vitro, that activation of Wnt/\β-catenin signaling leads to reprogramming of articular chondrocytes toward catabolism or loss of their stable phenotype, with subsequent loss of articular cartilage tissue structure and function. Frzb\textsuperscript{−/−} mice lacking this Wnt antagonist show increased activity of the Wnt signaling pathway, leading to increased bone stiffness and enhanced cartilage damage.\textsuperscript{21} It is important to note that upon induction of joint changes reminiscent of OA by enzymatic treatment (papain-induced osteoarthritis), by accelerated instability (collagenase-induced ligament and meniscal damage), or by inflammation (methylated bovine serum albumin [mBSA]-induced monoarthritis), Frzb\textsuperscript{−/−} mice showed greater cartilage loss than their wild-type counterparts.\textsuperscript{22} Increased cartilage damage in Frzb\textsuperscript{−/−} mice was associated with higher levels of \β-catenin–dependent Wnt signaling and with higher expression levels of matrix metalloproteinase 3. It was also demonstrated that FRZB can inhibit directly matrix metalloproteinase 3, probably through the netrin domain, indicating the potential complexity of the underlying mechanisms of observed phenomena. The role of Wnt signaling as a key regulatory pathway in postnatal joint biology and joint remodeling in chronic arthritis was further highlighted by the findings of Diarra and co-workers.\textsuperscript{23} Through inhibition of DKK1, an antagonist of the canonical Wnt signaling pathway, in a transgenic mouse model of TNF-driven inflammatory arthritis, part of the bone destructive effect was reversed. In view of the complexity of Wnt signaling and its potential downstream effects, it is important to further study the specific roles of canonical and noncanonical Wnt signaling in cartilage, bone, and the osteochondral junction, allowing Wnt-activated pathways and their components to become separate targets for therapeutic intervention. In view of this, recent findings of specific upregulation of Wnt16 in postinjury joint cartilage and in OA are of interest and open the opportunity for joint-specific Wnt targeting.\textsuperscript{24}

### Other Potential Anabolic Treatments

Other targets besides skeletal developmental pathways are signaling molecules that play a critical role in postnatal joint tissue homeostasis and turnover. These proteins/pathways can be regarded as potential postnatal anabolic agents contributing to the restoration of joint homeostasis. In this regard, the growth hormone (GH)/insulin-like growth factor (IGF) axis appears to be of interest. In particular, IGF-1 has been reported to be critical in the maintenance of homeostasis of articular cartilage explants ex vivo.\textsuperscript{25} Further evidence of its anabolic effect in vivo models has led to the early clinical development of intra-articular treatments with IGF-1 in gonarthrosis, although no recent reports of ongoing clinical trials have been found. Alternatively, successful local delivery via gene therapy has been reported.\textsuperscript{26}

Furthermore, it was reported that systemic administration of GH in horses may be beneficial for joint/articular cartilage biology, as it increases IGF-1 levels in synovial fluid.\textsuperscript{27} Improved formulations of GH have been explored to improve duration and effect size in synovial joints.\textsuperscript{28} A relationship between levels of IGF-1 and osteoarthritis has been suggested, further suggesting a potential benefit of targeting the GH/IGF-1 axis, particularly in OA. However, data so far are not conclusive, and further systematic analysis of the hypothalamic-pituitary axis, including growth hormone, IGF-1, and somatostatin, is required.\textsuperscript{29} Taken together, the potential of GH/IGF-1 as anabolic factors for joint repair and in particular their beneficial effects on articular cartilage metabolism have been recognized for a long time, but data supporting their clinical use and impact on joint biology in patients remain to be further explored.

### EXTRINSIC REPAIR

Cell-based therapies and tissue engineering are starting to enter the clinical arena, and thus are becoming of relevance to clinicians today. Many preclinical and early clinical applications appear promising, including the use of embryonic (ES), induced (iPS), or adult stem cells (MSC) in neurologic, cardiovascular, and musculoskeletal tissue repair and regeneration.

The mechanisms through which cellular therapies and combination products contribute to tissue repair are multiple and involve direct engulfment, proliferation, and differentiation to tissue-specific cell types, but they also include paracrine actions such as secretion of growth and differentiation factors that enhance the local tissue response. Indeed, cellular products typically function as multisignal delivery systems and interact with the microenvironment.\textsuperscript{30} In addition, it appears that adult stem cells display immunomodulatory properties; this has been clinically explored in graft-versus-host disease (GVHD) and in autoimmune diseases such as RA and systemic sclerosis. In view of this, cellular products are now subject to regulatory approval before they enter clinical practice; they belong

We herein shall discuss some applications that have reached the clinic that aim to restore the integrity and function of the locomotor system. Among them and most prominent are treatments for the repair of articular cartilage and bone defects.

**Joint Resurfacing**

**Autologous Chondrocyte Transplantation**

Regeneration or repair of symptomatic articular cartilage defects has been at the forefront of regenerative medicine ever since Brittberg and associates reported in 1994 a remarkably good clinical and structural outcome using a procedure called autologous chondrocyte transplantation/implantation (ACI). Briefly, cell populations were prepared by enzymatic release from a biopsy of articular cartilage taken from an unloaded area in the symptomatic joint. The chondrocytes were subsequently expanded in vitro and after six to eight population doublings were reimplanted in the joint surface defect under a periosteal flap, taken from the tibia of the same patient. This was followed by a long rehabilitation to reach optimal outcome at 18 to 24 months. This high-profile publication attracted quite some interest and triggered a wave of basic, translational, and clinical research in the field. It also evolved quickly in 1997 in the marketing of Carticel, an autologous cell product for the repair of symptomatic condylar defects of the knee, by the Genzyme Corporation in the United States.

Since that time, some progress has been made in studies aiming to improve and standardize the autologous chondrocyte preparation, the development of other delivery systems for chondrocytes, and replacement of the periosteal flap by a membrane of diverse composition. Most clinical studies were open label in design with diverse outcomes emerging (for review, see Van Osch et al). The study by Knutsen and colleagues, the first multicenter prospective randomized study, failed to show ACI to be superior to microfracture, a technique mostly considered as the standard of care for small (measuring less than 2 to 3 cm²) symptomatic joint surface defects. Microfracture, a bone marrow stimulation procedure, is based on puncturing of the subchondral bone plate into the bone marrow, the generation of a blood clot containing precursor cell populations derived from subchondral bone marrow, and spontaneous transformation of the repair clot into a fibrocartilaginous repair tissue. Microfracture is clearly associated with clinical benefit and with filling of the joint surface defect with repair tissue, but it is generally accepted that this repair is not durable, resulting in a consistent decline in clinical outcome over the long term. It is the aim of ACI to repair the cartilage defect with high-quality tissue matching the characteristics of neighboring tissue, resulting in improved long-term outcomes (Figure 7-2). Five-year follow-up data from this Scandinavian study did not indicate the clinical superiority of ACI over microfracture, although some indications suggest that good structural repair was associated with durable clinical outcome.

In Europe, an autologous chondrocyte cell product called ChondroCelect emerged as the first cellular product with central registration by the European Medicines Agency (EMA) under the new ATMP regulation. In a prospective randomized multicenter study, ChondroCelect, defined and characterized as a cell population capable of making ectopic stable cartilage in vivo, was shown to result in structurally superior repair tissue at 12 months after implantation as compared with microfracture. At 36 months, a strong tendency toward clinically superior outcome was noted, and at 5 years, it was confirmed (Van Lauwe, Saris et al, abstracts, ICRS 2010, personal communication).

Although still open for debate with respect to long-term clinical durability, the consensus is that ACI can be considered a regenerative treatment with good structural and clinical outcomes. It is important to note that the data show outcomes superior to standard care if “appropriate” patients are treated. Positive predictors of good outcome include early intervention (less than 3 years of symptoms), good quality of chondrocytes, well-trained surgeons, adherence to rehabilitation protocols, and no signs of osteoarthritis (as defined by Kellgren II). Cost-effectiveness will certainly improve if the treatment is shown to prevent progression toward OA, but these data are not yet available.

Long-term clinical outcomes have been reported and indicate good durability of the treatment in a substantial number of patients over periods exceeding 10 years.

**Additional Developments**

The next wave of developments is related to the use of autologous and/or allogeneic progenitor cell populations for the repair of joint surface defects. Adult bone marrow stromal cells, also called mesenchymal stem cells (MSCs), are lead candidates for their use in the repair of cartilage defects and other skeletal tissue such as bone defects. Other tissue sources have been explored for preparation of MSC
populations, including adipose tissue, synovium, and periosteum. Promising data using synovially derived progenitors have been reported in vivo, but it is unclear at this point what the long-term outcomes of these approaches will be. In particular, the safety and durability of an allogeneic approach are not known. Prerequisites for successful outcomes include improved understanding of differentiation protocols toward the proper stable phenotype and short- and long-term immune tolerance of allogeneic cells. Also, longer-term studies will provide necessary insight into the value of these innovative treatment approaches. Finally, a better understanding of the natural course of joint surface defects is needed, as is knowledge of what predisposes toward clinical deterioration and patients at risk; as indicated previously, additional insight is needed into factors contributing to successful repair such as the proper microenvironment in the joint.

**Bone Regeneration**

The most promising means of fully optimizing regenerative medicine approaches in musculoskeletal medicine is bone regeneration. Bone already has a remarkable potential to heal postnatally; this indicates that all necessary tools are available in principle in the postnatal individual to obtain full tissue repair, with no tissue scarring, and thus also to achieve tissue integration and remodeling. It is critical now to understand how nature heals bone fractures postnatally and subsequently for optimal mimicking of these repair processes when bone healing goes wrong, as in delayed union and nonunion, or in the repair of large bone defects such as avascular necrosis, or after resection of bone tumor.

The standard of care in the field is the use of autologous bone grafts, typically obtained from the iliac crest. However, important developments that have moved the boundaries in the field of bone engineering include development and improvement of biomaterials and new technologies in the design and production/manufacturing of these new biomaterials. The spectrum of “smart” biomaterials for bone engineering is vast, and recent progress has resulted in the production of resorbable osteoconductive surfaces; some even display osteoinductive properties (for review, see Mieszawska and Kaplan). These last characteristics in particular have been attained by coating biomaterials with bioactive factors, including growth factors such as bone morphogenetic proteins. BMP devices not only have led to more robust and predictable outcomes in spine fusion, but have demonstrated that recapitulation of embryonic tissue formation processes can lead to successful bone healing.

Despite all this progress, there is still considerable unmet medical need for bone healing. Indeed, we still are not able to heal larger bone defects. In view of this, we and others believe that a combined implant incorporating progenitor cell populations will be required (for a recent review, see Gómez-Barrena et al). Failure of bone healing typically is associated with lack of vascularization and lack of availability of proper precursor cell populations. Therefore the generation of tissue-engineered viable implants will need to come to full fruition to reach this goal. So far, the results of so-called combination products (scaffold and cells, eventually enriched with growth factors) have been somewhat disappointing. Lack of an underlying scientific basis for the design and manufacture of these implants is probably responsible for a large part of this. Therefore a biomimetic model called developmental engineering was proposed and extensively described. A few publications have indicated that this approach indeed may be more successful than the traditional cells-on-scaffold approach.

**REGENERATIVE MEDICINE IN ARTHRITIS**

As we start to understand intrinsic tissue response mechanisms and attempts to counteract destructive processes by enhancing tissue repair, new therapeutic targets have been identified and targeted with “classical” pharmaceutical approaches such as the development of protein therapies and small molecules interfering with critical repair mechanisms.

However, these approaches still may be far from sufficient; therefore more comprehensive approaches may be required. Increasingly, cellular therapies and combination products have been explored. Indeed, cellular therapies will affect local processes in many ways as cell populations can be manufactured so that they deliver a vast array of secreted signals, also called their secretome, which may influence local disease processes. Conversely, it becomes clear that the microenvironment will affect cellular products and will influence what they secrete and how they interact with the environment, including their engraftment, proliferation, differentiation, and tissue integration and remodeling.

Therefore if regenerative medicine and cellular therapies are to be successful, particularly in disease, we need to be able to assess and quantify the microenvironment and their mutual interactions. This translates into the clinic as identifying patient responders, which points toward the importance of personalized medicine. Indeed for regenerative medicine approaches to become cost-effective and successful, identification of the patient at risk and predicting responders to treatment will be of utmost importance.

Different cellular therapeutic approaches have been investigated in chronic arthropathies; most available data relate to the use of adult MSCs. Principles have been established for distinct indications. For inflammatory diseases such as RA and other autoimmune systemic diseases, immune suppressive and immune modulatory effects of cellular therapies can contribute to control of disease activity. Cellular therapies to enhance tissue repair have been explored extensively in preclinical animal models of post-traumatic osteoarthritis. It has also become evident that the use and role of MSCs in particular have been explored in OA and studied intensively. The mechanisms through which they can influence disease processes in OA appear diverse, as already discussed, and include so-called trophic effects and immune suppressive and anti-inflammatory effects, with cell engraftment contributing to local de novo tissue formation leading to meniscal and cartilage repair.

Because cellular therapies can address all these aspects, it is anticipated that properly manufactured cell products optimized for these positive effects may have a major impact when restoration of joint homeostasis is envisioned—a clinical goal that is probably crucial for interfering with disease progression in OA.
Regenerative medicine approaches, including enhancing and guiding intrinsic repair mechanisms, cellular therapeutics, and tissue engineering, are revolutionizing medicine in general and the field of musculoskeletal disorders and diseases in particular. We as rheumatologists will be witnessing and applying in the near future innovative regenerative and repair treatment approaches, including but not limited to repair of damaged joint surfaces, regeneration of difficult to heal fractured bones, repair of damaged ligamentous structures, and fabrication of a variety of “off the shelf” skeletal tissue structures (also called provisional tissues or tissue intermediates), such as viable pieces of bones, ligaments, menisci, and other joint tissues. These regenerative treatments will be of use for post-traumatic damaged joints and skeletal tissues and for (post) inflammatory and osteoarthritic joints.

Ultimately, and if everything else fails at some point leading to total organ failure (joint degeneration for the rheumatologist), implantation of a biological prosthesis with the potential of full tissue integration and remodeling is the dream, and proofs of principle exist to suggest that this may be within reach. To achieve this in a more robust and predictable fashion, we need to employ a developmental engineering approach.

We proposed the term developmental engineering to describe a method for rational and accurate design of robust, well-controlled manufacturing processes. This method integrates concepts from rapid advances in developmental biology, systems biology, and network science and is based on the design of in vitro processes consisting of sequential subprocesses corresponding to in vivo developmental stages under the control of signal pathway networks that can be modeled mathematically. They follow a gradual and coordinated progression of tissue growth and cell differentiation that leads to organization of cells into intermediate tissue forms with modular behavior. The macroscopic developmental modularity of tissues can be attributed to a corresponding modularity of network topology that describes gene interactions during the developmental process. We propose that identification of gene network modules that control developmental modules in vivo is the central theoretical and practical problem of both tissue engineering and developmental biology. Close interaction between the efforts of biologists and those of engineers can speed up tissue engineering processes.

References
7. Stoick-Cooper CL, Moon RT, Weidinger G: Advances in signaling in vertebrate regeneration as a prelude to regenerative medicine, Genes Dev 21:1312–1315, 2007.


Proteinases and Matrix Degradation

YASUNORI OKADA

KEY POINTS
Proteinases are generally classified into aspartic proteinases, cysteine proteinases, serine proteinases, and metalloproteinases according to catalytic mechanism.

Because of the acidic pH optima and intracellular localization within lysosomes, most of the aspartic proteinases and cysteine proteinases are involved in intracellular degradation of extracellular matrix (ECM) components.

Serine proteinases and metalloproteinases are neutral proteinases and play a central role in extracellular degradation of ECM macromolecules.

ECM-degrading metalloproteinases are composed mainly of the MMP (matrix metalloproteinase) and ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) gene families.

Endogenous proteinase inhibitors are proteinase class specific, whereas α2-macroglobulin inhibits the activities of all proteinases.

The activities of ECM-degrading proteinases at the local tissues are regulated by the balance between the proteinases and their inhibitors, which may be determined by production rates of proteinases and inhibitors, their secretion, activation of proenzymes, and anchoring systems of the activated proteinases to cell surfaces.

ProMMPs (zymogens of MMPs) are activated via the extracellular, intracellular, and pericellular pathways depending on the MMP species.

Aggrecan and type II collagen, two major ECM components, in articular cartilage may be degraded by differential or complementary actions of the MMP and ADAMTS species in arthritides.

In rheumatoid arthritis, articular cartilage is destroyed by proteinases accumulated in synovial fluid, direct contact of proteolytic synovium and pannus tissue, and proteinases derived from chondrocytes, whereas bone is resorbed by osteoclasts mainly by the action of cathepsin K and MMP-9 under acidic and hypercalcemic conditions in subosteoclastic compartments.

In osteoarthritis, chondrocyte-derived metalloproteinases including the MMP and ADAMTS species contribute primarily to the breakdown of articular cartilage.

Extracellular matrix (ECM) plays critical roles in normal development and function of the organism by interacting with cells and supporting tissue structures. The in vivo cellular functions regulated by cell-ECM interaction include proliferation, differentiation, apoptosis, and motility. The proteolytic turnover and remodeling of ECM are transient and highly controlled under physiologic conditions, and excessive degradation of ECM components by proteinases causes tissue destruction in many pathologic conditions. In rheumatoid arthritis and osteoarthritis, ECM-degrading proteinases are elevated without sufficient endogenous inhibitors and they are believed to play central roles in the destruction of articular cartilage and bone on the basis of a local imbalance between proteinases and inhibitors. This chapter provides up-to-date information about the ECM-degrading proteinases and their inhibitors.

EXTRACELLULAR MATRIX DEGRADING PROTEINASES

ECM is degraded by endopeptidases (i.e., proteinases) that act internally on polypeptide chains; little evidence is present for the roles of exopeptidases that cleave one or a few amino acids from the N- or C-terminus. Proteinases comprise aspartic proteinases, cysteine proteinases, serine proteinases, and metalloproteinases, which are classified according to catalytic mechanism. Proteinases from each of the four classes are involved in the degradation of ECM macromolecules.

Aspartic Proteinases
Most aspartic proteinases have two aspartic acid residues in their catalytic sites, where the nucleophile that attacks the scissile peptide bond is an activated water molecule. Mammalian aspartic proteinases include the digestive enzymes (pepsin and chymosin), the intracellular cathepsin D and cathepsin E, and rennin. Among the proteinases belonging to this group, cathepsin D is the major aspartic proteinase involved in ECM degradation (Table 8-1). It exhibits proteolytic activity against most substrates such as aggrecan and collagen telopeptides with pH optima between pH 3.5 and 5. Because of the acidic pH optima and intracellular localization within lysosomes, cathepsin D is probably responsible for intracellular degradation of phagocytosed ECM fragments that previously were degraded in the extracellular spaces. A study on cartilage explant cultures using the aspartic proteinase inhibitor suggests, however, the possibility that cathepsin D secreted extracellularly contributes to the degradation of aggrecan in articular cartilage.

Cysteine Proteinases
Cysteine proteinases are endopeptidases in which the nucleophile of the catalytic site is the sulphydryl group of a cysteine residue. The ECM-degrading cysteine proteinases...
Table 8-1  Proteinases That May Be Involved in Degradation of Extracellular Matrix

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Source</th>
<th>Inhibitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspartic Proteinases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cathepsin D</td>
<td>Lysosome</td>
<td>Pepstatin</td>
</tr>
<tr>
<td>Cysteine Proteinases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cathepsin B</td>
<td>Lysosome</td>
<td>Cystatins</td>
</tr>
<tr>
<td>Cathepsin L</td>
<td>Lysosome</td>
<td>Cystatins</td>
</tr>
<tr>
<td>Cathepsin S</td>
<td>Lysosome</td>
<td>Cystatins</td>
</tr>
<tr>
<td>Cathepsin K</td>
<td>Lysosome</td>
<td>Cystatins</td>
</tr>
<tr>
<td>Calpain</td>
<td>Cytosol</td>
<td>Calpastatin</td>
</tr>
<tr>
<td>Serine Proteinases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutrophil elastase</td>
<td>Neutrophils</td>
<td>α1-PI</td>
</tr>
<tr>
<td>Cathepsin G</td>
<td>Neutrophils</td>
<td>α1-Antichymotrypsin</td>
</tr>
<tr>
<td>Proteinase 3</td>
<td>Neutrophils</td>
<td>α1, α2, α3, α7, α9,</td>
</tr>
<tr>
<td>Plasmin</td>
<td>Plasma</td>
<td>Aprotinin</td>
</tr>
<tr>
<td>Plasma kallikrein</td>
<td>Plasma</td>
<td>Aprotinin</td>
</tr>
<tr>
<td>Tissue kallikrein</td>
<td>Glandular tissues</td>
<td>Aprotinin; kallistatin</td>
</tr>
<tr>
<td>tPA</td>
<td>Endothelial cells;</td>
<td>PAI-1; PAI-2;</td>
</tr>
<tr>
<td></td>
<td>chondrocytes</td>
<td></td>
</tr>
<tr>
<td>Tryptase</td>
<td>Fibroblasts;</td>
<td>PAI-1; PAI-2; PN-1</td>
</tr>
<tr>
<td>Chymase</td>
<td>Mast cells</td>
<td>Tryptstatin</td>
</tr>
<tr>
<td></td>
<td>Mast cells</td>
<td>α1, α2, α7, α9, α9,</td>
</tr>
<tr>
<td>Metalloproteinases*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| MMPs                   | Tissue cells;      | TIMP-1, 2, 3, and 4;
|                        | inflammatory cells | RECK for MMP-2, 7, 9, and 11; |
| ADAMTs                 | Tissue cells;      | TIMP-3; RECK for   |
| ADAMS                  | inflammatory cells | ADAM10             |

*For details of MMPs, ADAMTs, and ADAMS, see Tables 8-2, 8-3, and 8-4. ADAMs, a disintegrin and metalloproteinases; ADAMTs, a disintegrin and metalloproteinases with thrombospondin motifs; MMPs, matrix metalloproteinases; PAI, plasminogen activator inhibitor; PI, proteinase inhibitor; PN, proteinase nexin; RECK, reversion-inducing, cysteine-rich protein with Kazal motifs; TIMP, tissue inhibitor of metalloproteinases; tPA, tissue-type plasmin activator; uPA, urokinase-type plasminogen activator.

include lysosomal cathepsins B, L, S, and K and the calpains (see Table 8-1). Cathepsins B and L digest the telopeptide regions of fibrillar collagen types I and II, the nonhelical regions of collagen types IX and XI, and aggrecan at acidic pH. Cathepsin S has a similar spectrum of substrates within a broad range of pH values. Cathepsin K, also called cathepsin O, O2, or OX, is a collagenolytic cathepsin that efficiently cleaves type I collagen at the triple helical regions at pH values between 4.5 and 6.6. The proteinase also degrades gelatin and osteonectin. On the basis of the data that cathepsin K is highly expressed by human osteoclasts and inactivating mutations or deletion of the gene result in an osteopetrotic phenotype in humans and animals, cathepsin K is believed to play a key role in bone resorption in joint diseases (see later). Because calpains B, L, S, and K are expressed in synovium or articular cartilage or both in rheumatoid arthritis and osteoarthritis, they may also be involved in the cartilage destruction through degradation of the ECM macromolecules.

Calpains are Ca²⁺-dependent, papain-like cysteine proteinases and are ubiquitously distributed among mammalian cells. The best-characterized members of the calpain superfamily are μ-calpain and m-calpain, which also are called conventional (μ-calpain) and classic (m-calpain) calpains.

Calpains are involved in various pathologic conditions such as muscle dystrophy by acting intracellularly. They are present in the extracellular spaces and in osteoarthritic synovial fluid, and they can degrade aggrecan.

Serine Proteinases
Serine proteinases require the hydroxyl group of a serine residue acting as the nucleophile that attacks the peptide bond. They include the largest number of proteinases classified to about 40 families. Most can degrade ECM molecules. The major ECM-degrading serine proteinases in joint tissues are subsequently described (see Table 8-1).

Neutrophil Elastase and Cathepsin G
Neutrophil elastase and cathepsin G are serine proteinases that are synthesized as precursors in promyelocytes in bone marrow and subsequently stored in the azurophil granules of polymorphonuclear leukocytes as active enzymes. Mature leukocytes do not synthesize elastase, but they mobilize azurophil granules to the cell surface and release the proteinases in response to various stimuli. Monocytes have low levels of elastase but lose the enzyme during the differentiation into macrophages. Neutrophil elastase and cathepsin G are basic glycoproteins with isoelectric points larger than 9 (neutrophil elastase) and about 12 (cathepsin G). They can be readily trapped in cartilage matrix that has a negative charge.

Neutrophil elastase and cathepsin G cleave elastin; the telopeptide region of fibrillar collagen types I, II, and III; other collagen types IV, VI, VIII, IX, X, and XI; and other ECM components such as fibronectin, laminin, and aggrecan at neutral pH. These serine proteinases also can be involved indirectly in the breakdown of ECM by activating the zymogen of matrix metalloproteinases (proMMPs) and by inactivating endogenous proteinase inhibitors such as α₁-antiplasmin, α₁-antichymotrypsin, and tissue inhibitors of metalloproteinases (TIMPs).

Mast Cell Chymase and Tryptase
Chymase and tryptase are packaged in secretory granules together with histamine and other mediators in mast cells, which are infiltrated in rheumatoid synovium. Chymase is a chymotrypsin-like proteinase with a broad spectrum of activity against ECM components such as type VI collagen and aggrecan. It also activates proMMPs such as proMMP-1, proMMP-3, and proMMP-9. Although prochymase is activated intracellularly and stored in the granules, the activity in the granules is limited at low pH and becomes fully active when released extracellularly. Tryptase is a trypsin-like proteinase that degrades collagen type VI and fibronectin; it also activates proMMP-3.

Plasmin and Plasminogen Activators
Plasminogen is synthesized in liver and secreted to plasma. It can bind to fibrin and to cells, and after activation by plasminogen activators, plasmin readily digests fibrin. Membrane-bound plasmin also degrades many ECM
components including proteoglycan, fibronectin, type IV collagen, and laminin. Other important functions of plasmin are to initiate the activation of proMMPs, activate latent cell-associated transforming growth factor (TGF)-β, and act as proenzyme convertase. Plasmin is generated through activation of plasminogen mainly by plasminogen activators, principally two serine proteinases, tissue-type plasminogen activator (tPA) and urokinase-type plasminogen activator (uPA).

The tPA is synthesized as a proenzyme of 70 kD and is secreted into the circulating blood primarily by endothelial cells, fibroblasts, chondrocytes, and tumor cells. In addition to being a major activator of plasminogen for fibrinolysis, tPA plays a key role in the clearance of fibrin from the circulation.

The uPA molecule was first purified from urine as a proenzyme of 54 kD. It is converted to the active form of two chains of 30 kD and 24 kD linked by a disulfide bond. Another fully active form of 33 kD is generated by plasmin. Although the expression of uPA is limited to certain cells such as renal tubules and bladder urothelium under physiologic conditions, it is more widely expressed in various cells including invasive cancer cells, migrating keratinocytes, and activated leukocytes in pathologic situations. Pro-uPA and two-chain uPA bind to a specific uPA receptor, a single-chain glycoprotein with a glycosylphosphatidylinositol (GPI) moiety expressed on fibroblasts, macrophages, and tumor cells. Receptor-bound uPA preferentially activates cell membrane-bound plasminogen into plasmin. Cell membrane-bound plasmin can activate receptor-bound pro-uPA. Among its specificities, uPA has a limited action on fibronectin.

**Kallikreins**

Two types of kallikreins, plasma and tissue kallikreins, are known. Plasma kallikrein, with two disulfide-linked chains (36 kD and 52 kD), is generated from prokallikrein of 88 kD by coagulation factor XIIa or by kallikrein itself. It activates kininogens to bradykinin and activates proMMP-1 and proMMP-3. Tissue kallikrein is synthesized in glandular tissues. It releases Lys-bradykinin from kininogen and activates proMMP-8.

**Metalloproteinases**

Similar to aspartic proteinases, metalloproteinases are endopeptidases in which the nucleophilic attack on a peptide bond is mediated by a water molecule. A divalent metal cation, usually zinc, activates the water molecule. Among the metalloproteinases, MMPs (matrix metalloproteinases), which are also designated matrixins (a subfamily of the metzincin superfamily), are key ECM-degrading, zinc-dependent endopeptidases (Table 8-2). Accumulated evidence indicates, however, that some members of the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) family, which is an MMP-related gene family, also are involved in the degradation of ECM, especially cartilage proteoglycan (Table 8-3). Only a few members of the ADAM (a disintegrin and metalloproteinase) family have limited activity to ECM components (Table 8-4).

**Matrix Metalloproteinases**

The human MMP family comprises 23 different members, which have MMP designations (numbered according to a sequential numbering system) and common names coined by the authors of the published reports (see Table 8-2). On the basis of the biochemical properties provided by the domain structures and on their substrate specificity, these family members are classified into two major subgroups: secreted-type MMPs and membrane-anchored MMPs. MMP-4, MMP-5, and MMP-6 are excluded from the list because they are identical to other known MMPs (i.e., MMP-3 and MMP-2). MMP-18 and MMP-22 also are missing in Table 8-2 because they are assigned to *Xenopus* collagenase-4 and chicken MMP.

Most secreted-type MMPs including collagenases, stromelysins, and other MMPs are composed of three basic domains—the propeptide, catalytic, and hemopexin-like domains—that are preceded by hydrophobic signal peptides (Figure 8-1). The N-terminal propeptide domain has one unpaired cysteine in the conserved sequence of PRCGXPD. The cysteine residue in the sequence interacts with the essential zinc atom in the catalytic domain to prevent it from binding the catalytic water molecule, maintaining the proenzyme in an inactive state. The catalytic domain has the zinc-binding motif HEXGHXXGXXH, in which three histidines bind to the catalytic zinc atom. The four-blade, C-terminal hemopexin-like domain, which is connected to the catalytic domain by a proline-rich hinge region, interacts with ECM components and can play a role in determining the substrate specificity in some MMPs. Gelatinases have these domains with additional insertions of collagen-binding type II repeats of fibronectin in the catalytic domain (see Figure 8-1); this provides them with collagen-binding properties. Matrilysins are the shortest, lacking the hemopexin-like domains. Furin-activated MMPs contain insertions of a basic motif with a cleavage site by proprotein convertases including furin at the end of the propeptide domains (see Figure 8-1).

Membrane-anchored MMPs include three different types of MMPs: type I transmembrane-type MMPs, GPI-linked MMPs, and type II transmembrane-type MMP. MMP-14, MMP-15, MMP-16, and MMP-24 (MT1-, MT2-, MT3-, and MT5-MMPs, respectively) have type I transmembrane domains in the C-terminal region, but MMP-17 (MT4-MMP) and MMP-25 (MT6-MMP) contain GPI anchors in the C-terminal region without transmembrane domains (see Figure 8-1). MMP-23 is unique in that it has type II transmembrane domain, a cysteine array, and an immunoglobulin-like domain instead of a hemopexin-like domain (see Figure 8-1).

**Secreted-Type Matrix Metalloproteinases**

**Collagenases** (MMP-1, MMP-8, and MMP-13). The collagenases include MMP-1 (interstitial collagenase, collagenase-1), MMP-8 (neutrophil collagenase, collagenase-2), and MMP-13 (collagenase-3). These MMPs attack triple helical regions of interstitial collagen types I, II, and III at a specific single site after the Gly residue of the partial sequences Gly-(Ile or Leu)-(Ala or Leu), located about three-fourths of the distance from the N-terminus. This cleavage generates fragments approximately three-fourths and one-fourth of the size of the collagen molecules.
Table 8-2  Substrates of Human Matrix Metalloproteinases

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>ECM Substrates</th>
<th>Non-ECM Substrates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Secreted-Type MMPs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Collagenases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP-1 (Interstitial collagenase)</td>
<td>Collagens I, II, III, VII, X; gelatins; aggrecan; link protein; entactin; tenascin; perlecan</td>
<td>α₂-Macroglobulin; α₁-Pi; α₁-antichymotrypsin; IGF-BP-2, -3, -5; pro-IL-1β; CTGF α₁-Pi</td>
</tr>
<tr>
<td>MMP-8 (Neutrophil collagenase)</td>
<td>Collagens I, II, and III; gelatins; aggrecan; link protein</td>
<td>CTGF; pro-TGF-β; α₁-antichymotrypsin</td>
</tr>
<tr>
<td>MMP-13 (Collagenase-3)</td>
<td>Collagens I, II, III, IV, IX, X, XIV; aggrecan; Fn; tenascin</td>
<td></td>
</tr>
<tr>
<td><strong>Gelatinases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP-2 (Gelatinase A)</td>
<td>Gelatins; collagens IV, V, VII, XI; Ln; Fn; elastin; aggrecan; link protein</td>
<td>Pro-TGF-β; FGF receptor I; MCP-3; IGF-BP-5; pro-IL-1β; galectin-3; plasminogen Pro-TGF-β; IL-2 receptor α; Kit-L; IGF-BP-3; pro-IL-1β; α₁-Pi; galectin-3; ICAM-1; plasminogen</td>
</tr>
<tr>
<td>MMP-9 (Gelatinase B)</td>
<td>Gelatins; collagens III, IV, V; aggrecan; elastin; entactin; link protein</td>
<td></td>
</tr>
<tr>
<td><strong>Stromelysins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP-3 (Stromelysin-1)</td>
<td>Aggrecan; decorin; gelatins; Fn; Ln; collagens III, IV, IX, X; tenascin; link protein</td>
<td>IGF-BP-3; pro-IL-1β; HB-EGF; CTGF; E-cadherin; α₁-antichymotrypsin; α₁-Pi; α₂-macroglobulin; plasminogen; uPA; proMMP-1, 7, 8, 9, 13 ProMMP-1, 7, 8, 10</td>
</tr>
<tr>
<td>MMP-10 (Stromelysin-2)</td>
<td>Aggrecan; Fn; Ln; collagens III, IV, V; link protein</td>
<td></td>
</tr>
<tr>
<td><strong>Matrilysins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP-7 (Matrilysin-1)</td>
<td>Aggrecan; gelatins; Fn; Ln; elastin; entactin; collagen IV; tenascin; link protein</td>
<td>Pro-α₁-defensin; Fas-L; β4 integrin; E-cadherin; pro-TNF; CTGF; HB-EGF; RANKL; IGF-BP-3; plasminogen α₁-Pi; proMMP-9</td>
</tr>
<tr>
<td>MMP-26 (Matrilysin-2)</td>
<td>Gelatin; collagen IV; Fn; fibrinogen</td>
<td></td>
</tr>
<tr>
<td><strong>Furin-Activated MMPs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP-11 (Stromelysin-3)</td>
<td>Fn; Ln; aggrecan; gelatins</td>
<td>α₁-Pi; α₂-macroglobulin; IGF-BP-1</td>
</tr>
<tr>
<td>MMP-28 (Epilysin)</td>
<td>Unknown</td>
<td>Casein</td>
</tr>
<tr>
<td><strong>Other Secreted-Type MMPs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP-12 (Metalloelastase)</td>
<td>Elastin; Fn; collagen V; osteonectin; Collagen IV; gelatin; Fn; tenascin; aggrecan; COMP; Ln; nidogen</td>
<td>Plasminogen; apolipoprotein A IGF-BP-3</td>
</tr>
<tr>
<td>MMP-19 (RAS-1)</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>MMP-20 (Enamelysin)</td>
<td>Amelogenin; aggrecan; gelatin; COMP</td>
<td>Unknown</td>
</tr>
<tr>
<td>MMP-21</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>MMP-27</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td><strong>Membrane-anchored MMPs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type I Transmembrane-Type MMPs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP-14 (MT1-MMP)</td>
<td>Collagens I, II, III; gelatins; aggrecan; Fn; Ln; fibrin; Ln-5</td>
<td>ProMMP-2, 13; CD44; tissue transglutaminase</td>
</tr>
<tr>
<td>MMP-15 (MT2-MMP)</td>
<td>Fn; tenascin; nidogen; aggrecan; perlecan; Ln Collagen III; Fn; gelatin</td>
<td>ProMMP-2; tissue transglutaminase</td>
</tr>
<tr>
<td>MMP-24 (MT5-MMP)</td>
<td>PG</td>
<td>ProMMP-2; tissue transglutaminase</td>
</tr>
<tr>
<td><strong>GPI-Linked MMPs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP-17 (MT4-MMP)</td>
<td>Gelatin; fibrinogen</td>
<td>Unknown</td>
</tr>
<tr>
<td>MMP-25 (MT6-MMP)</td>
<td>Gelatin; collagen IV; fibrin; Fn; Ln</td>
<td>ProMMP-2</td>
</tr>
<tr>
<td><strong>Type II Transmembrane-Type MMP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP-23</td>
<td>Gelatin</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

COMP, cartilage oligomeric matrix protein; CTGF, connective tissue growth factor; ECM, extracellular matrix; Fn, fibronectin; GPI, glycosylphosphatidylinositol; HB-EGF, heparin-binding epidermal growth factor; IGF-BP, insulin-like growth factor binding protein; ICAM-1, intercellular adhesion molecule 1; RANKL, receptor activator for nuclear factor κB ligand; IL-1, interleukin-1; Ln, laminin; MCP, monocyte chemoattractant protein; PG, proteoglycan; PI, proteinase inhibitor; TGF, transforming growth factor; TNF, tumor necrosis factor; uPA, urokinase-type plasminogen activator.

A biochemical study has disclosed the molecular mechanism of the cleavage: MMP-1 unwinds the triple helical structure by interacting with the α2(I) chain of type I collagen and cleaves the three α chains in succession.9 MMP-13 is unique in that it cleaves α chains of type II collagen at two sites of the Gly<sup>906</sup>-Leu<sup>907</sup> Gly<sup>909</sup>-Gln<sup>910</sup> bonds.10 All of these collagenases degrade the interstitial collagens, but their specific activities against the collagens are different; MMP-1, MMP-8, and MMP-13 preferentially digest collagen types III, I, and II.10,11 Although rodents such as mice were originally thought to have only two collagenases (MMP-8 and MMP-13) and to lack the MMP-1 gene,
Table 8-3  Members of the ADAMTS Family

<table>
<thead>
<tr>
<th>ADAMTS</th>
<th>Other Names</th>
<th>Demonstration of Proteinase Activity*</th>
<th>Functions and Biochemical Features</th>
<th>Tissue and Cell Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAMTS1</td>
<td>C3-C5; METH1; KIAA1346</td>
<td>+</td>
<td>Digestion of aggrecan and versican; binding to heparin</td>
<td>Kidney; heart; cartilage</td>
</tr>
<tr>
<td>ADAMTS2</td>
<td>Procollagen N-proteinase; hPCPNI; PCINP</td>
<td>+</td>
<td>Processing of collagen I and II N-propeptides</td>
<td>Skin; tendon</td>
</tr>
<tr>
<td>ADAMTS3</td>
<td>KIAA0366</td>
<td>+</td>
<td>Processing of collagen N-propeptides</td>
<td>Brain</td>
</tr>
<tr>
<td>ADAMTS4</td>
<td>KIAA0068; aggrecanase-1; ADMP-1</td>
<td>+</td>
<td>Digestion of aggrecan, brevican, and versican</td>
<td>Brain; heart; cartilage</td>
</tr>
<tr>
<td>ADAMTS5</td>
<td>ADAMTS11; aggrecanase-2; ADMP-2</td>
<td>+</td>
<td>Digestion of aggrecan</td>
<td>Uterus; placenta; cartilage</td>
</tr>
<tr>
<td>ADAMTS6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Placenta</td>
</tr>
<tr>
<td>ADAMTS7</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Various tissues</td>
</tr>
<tr>
<td>ADAMTS8</td>
<td>METH-2</td>
<td>+</td>
<td>Digestion of aggrecan; inhibition of angiogenesis</td>
<td>Lung; heart</td>
</tr>
<tr>
<td>ADAMTS9</td>
<td>KIAA1312</td>
<td>+</td>
<td>Digestion of aggrecan</td>
<td>Cartilage</td>
</tr>
<tr>
<td>ADAMTS10</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Lung (fetus)</td>
</tr>
<tr>
<td>ADAMTS12</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Lung (fetus)</td>
</tr>
<tr>
<td>ADAMTS13</td>
<td>VWFCP; C9orf8</td>
<td>+</td>
<td>Cleavage of von Willebrand factor</td>
<td>Liver; prostate; brain</td>
</tr>
<tr>
<td>ADAMTS14</td>
<td>—</td>
<td>+</td>
<td>Processing of collagen N-propeptides</td>
<td>Brain; uterus</td>
</tr>
<tr>
<td>ADAMTS15</td>
<td>—</td>
<td>+</td>
<td>Digestion of aggrecan</td>
<td>Liver (fetus); kidney (fetus)</td>
</tr>
<tr>
<td>ADAMTS16</td>
<td>—</td>
<td>+</td>
<td>Digestion of aggrecan</td>
<td>Prostate; brain; uterus</td>
</tr>
<tr>
<td>ADAMTS17</td>
<td>FLJ32769; LOC123271</td>
<td>—</td>
<td>—</td>
<td>Prostate; brain; liver</td>
</tr>
<tr>
<td>ADAMTS18</td>
<td>ADAMTS21; HGNC:16662</td>
<td>+</td>
<td>Digestion of aggrecan</td>
<td>Prostate; brain</td>
</tr>
<tr>
<td>ADAMTS19</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Lung (fetus)</td>
</tr>
<tr>
<td>ADAMTS20</td>
<td>—</td>
<td>+</td>
<td>Digestion of versican (and aggrecan)</td>
<td>Brain; testis</td>
</tr>
</tbody>
</table>

*Proteinase activities are shown in 13 members of the ADAMTS family, but not in 6 other members.

rodent homologues of the human MMP-1 gene were cloned and named mouse collagenase A and B (Mcol-A and Mcol-B).12

In addition to the interstitial fibrillar collagens, MMP-1, MMP-8, and MMP-13 degrade many other ECM macromolecules. MMP-1 digests entactin, collagen X, gelatins, perlecans, aggrecans, and cartilage link protein (see Table 8-2). MMP-8 digests aggrecan, gelatins, and cartilage link protein (see Table 8-2). MMP-13 hydrolyzes aggrecan; types IV, IX, X, and XIV collagens; fibronectin; and tenascin. Non-ECM substrates of MMP-1, MMP-8, and MMP-13 include α2-macroglobulin, α2-antiproteinase inhibitor, α2-antichymotrypsin, insulin-like growth factor binding protein (IGF-BP)-2 and IGF-BP-3, connective tissue growth factor (CTGF), and pro-TGF-β (see Table 8-2).

Gelatinases (MMP-2 and MMP-9). MMP-2 (gelatinase A) and MMP-9 (gelatinase B) belong to the gelatinase subgroup. Both MMPs readily digest collagens and cleave collagen types IV and V.13,14 Elastin, aggrecan, and cartilage link protein also are substrates of the gelatinases. Although MMP-2 and MMP-9 share some substrates, they have different activities on several ECM macromolecules. MMP-2, but not MMP-9, digests fibronectin and laminin,15 and type III collagen and α2 chains of type I collagen are degraded only by MMP-9.14 The gelatinases also process directly TGF-β into an active ligand (see Table 8-2). MMP-2 and MMP-9 cleave fibroblast growth factor receptor type I and interleukin (IL)-2 receptor type α (see Table 8-2). MMP-9 also releases soluble Kit-ligand.13 MMP-2 processes monocyte chemoattractant protein (MCP)-3 into an MCP-3 fragment deleting the N-terminal four amino acids, which can bind to CC-chemokine receptors and act as a general chemokine antagonist.16

Stromelysins (MMP-3 and MMP-10). The subgroup of stromelysins consists of MMP-3 (stromelysin-1) and MMP-10 (stromelysin-2). They share 78% identity in amino acid sequence and have similar enzymatic properties.17 The enzymes hydrolyze numerous ECM macromolecules including aggrecan, fibronectin, laminin, and collagen IV (see Table 8-2).18 Collagen types III, IX, X and telopeptides of collagen types I, II, and XI also are digested by MMP-3.19 In addition to the ECM components, MMP-3 is active on IGF-BP-3, IL-1β, heparin-binding epidermal growth factor (HB-EGF), CTGF, E-cadherin, α2-antichymotrypsin, and α2-proteinase inhibitor (see Table 8-2). MMP-3 also activates many proMMPs.8 A similar activator function has been identified for MMP-10.20

Matrilysins (MMP-7 and MMP-26). Matrilysins include MMP-7 (matrilysin-1) and MMP-26 (matrilysin-2), which are the smallest of the MMPs, having only the propeptide and catalytic domains. The substrate specificity of MMP-7 is similar to that of stromelysins, digesting numerous ECM components including aggrecan; gelatins; fibronectin; laminin; elastin; entactin; collagen types III, IV, V, IX, X, and XI; fibrin/fibrinogen; vitronectin; tenascin; and link protein (see Table 8-2). Although these substrates overlap with the substrates of other MMPs, the specific activity of MMP-7 to most substrates is highest among the MMPs.21,22 Non-ECM molecules such as α-defensin, Fas ligand, β4 integrin, E-cadherin, plasminogen, tumor necrosis factor...
Table 8-4 Members of the Human ADAM Family

<table>
<thead>
<tr>
<th>ADAM</th>
<th>Other Names</th>
<th>Proteinase-Type (P) or Non–Proteinase-Type (NP)</th>
<th>Functions and Biochemical Features</th>
<th>Tissue and Cell Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAM2</td>
<td>PH-30β; Fertilin-β</td>
<td>NP</td>
<td>Sperm/egg binding/fusion; binding to integrin αβ1, αβ3, and αβ9; aggregation of αβ1 to αβ9</td>
<td>Sperm</td>
</tr>
<tr>
<td>ADAM7</td>
<td>EAP I; GP-83</td>
<td>NP</td>
<td>Binding to integrin α4β1, α4β3, and α4β7; shedding of CD23</td>
<td>Testis</td>
</tr>
<tr>
<td>ADAM8</td>
<td>MS2 (CD156)</td>
<td>P</td>
<td>Neutrophil infiltration; shedding of CD23</td>
<td>Macrophages; neutrophils</td>
</tr>
<tr>
<td>ADAM9</td>
<td>MDC9; MCMP; Meltrin-γ</td>
<td>P</td>
<td>Shedding of HB-EGF, TNF-p75 receptor; and APP; digestion of fibronectin and gelatin; binding to integrin αβ1, αβ3, αβ4, αβ9, and αβ9; shedding of CD23</td>
<td>Various tissues</td>
</tr>
<tr>
<td>ADAM10</td>
<td>MDAM; Kuzbanian</td>
<td>P</td>
<td>Shedding of TNF, Delta, Delta-like 1, Jagged, N-cadherin, E-cadherin, VE-cadherin, Ephrin A2, Ephrin A5, Fas-1, IL6R, APP, L1, CD44, and HB-EGF; digestion of collagen IV, gelatin, and myelin basic protein; presence of RRKR sequence</td>
<td>Kidney; brain; chondrocytes</td>
</tr>
<tr>
<td>ADAM11</td>
<td>MDC</td>
<td>NP</td>
<td>Muscle formation; presence of RRKR sequence; binding to integrin αβ1 and αβ3; shedding of HB-EGF, and epiregulin; digestion of collagen IV, gelatin, and fibronectin</td>
<td>Brain</td>
</tr>
<tr>
<td>ADAM12</td>
<td>Meltrin-α; MCMP; MLTN; MLTNA</td>
<td>P</td>
<td>Expression in arteriosclerosis; binding to integrin αβ3, αβ5, and αβ9; shedding of CD23</td>
<td>Osteoblasts; muscle cells; chondrocytes; placenta</td>
</tr>
<tr>
<td>ADAM15</td>
<td>Metarginin; MDC15; AD56; CR II-7</td>
<td>P</td>
<td>Expression in arteriosclerosis; binding to integrin αβ3, αβ5, and αβ9; shedding of CD23</td>
<td>Smooth muscle cells; chondrocytes; endothelial cells; osteoclasts</td>
</tr>
<tr>
<td>ADAM17</td>
<td>TACE; cSVP</td>
<td>P</td>
<td>Shedding of TNF, TGF-β, TNF-p75 receptor, RANKL, amphiregulin, epiregulin, HB-EGF, APP, L-selectin, and CD44; presence of RRKR sequence</td>
<td>Macrophages; various tissues; carcinoma tissue</td>
</tr>
<tr>
<td>ADAM18</td>
<td>tMD III</td>
<td>NP</td>
<td>—</td>
<td>Testis</td>
</tr>
<tr>
<td>ADAM19</td>
<td>Meltrin-β; FKG34</td>
<td>P</td>
<td>Formation of neuron; shedding of neuregulin, and RANKL; binding to integrin αβ1 and αβ3</td>
<td>Testis</td>
</tr>
<tr>
<td>ADAM20</td>
<td>—</td>
<td>P</td>
<td>Formation of sperm</td>
<td>Testis</td>
</tr>
<tr>
<td>ADAM21</td>
<td>—</td>
<td>P</td>
<td>—</td>
<td>Testis</td>
</tr>
<tr>
<td>ADAM22</td>
<td>MDC2</td>
<td>NP</td>
<td>—</td>
<td>Brain</td>
</tr>
<tr>
<td>ADAM23</td>
<td>MDC3</td>
<td>NP</td>
<td>Binding to integrin αβ3</td>
<td>Brain; heart</td>
</tr>
<tr>
<td>ADAM28</td>
<td>e-MDC II; MDC-Lm; MDC-Ls</td>
<td>P</td>
<td>Digestion of myelin basic protein, IGF-BP-3, and CTGF; shedding of CD23; binding to integrin αβ1, αβ4β7, and αβ9</td>
<td>Testis; lung; lymphocytes; pancreas; uterus</td>
</tr>
<tr>
<td>ADAM29</td>
<td>svph 1</td>
<td>NP</td>
<td>—</td>
<td>Testis</td>
</tr>
<tr>
<td>ADAM30</td>
<td>svph 4</td>
<td>P</td>
<td>—</td>
<td>Testis</td>
</tr>
<tr>
<td>ADAM32</td>
<td>AJ131563</td>
<td>NP</td>
<td>Mutation in bronchial asthma patients; shedding of APP and KL-1; digestion of insulin B chain; binding to integrin αβ1, αβ3, and αβ9</td>
<td>Testis</td>
</tr>
<tr>
<td>ADAM33</td>
<td>—</td>
<td>P</td>
<td>—</td>
<td>Testis</td>
</tr>
<tr>
<td>ADAMDEC1</td>
<td>—</td>
<td>P</td>
<td>—</td>
<td>Lymphatic system; gastrointestinal system</td>
</tr>
</tbody>
</table>

Note: ADAM11, ADAM-like decysin 1; APP, amyloid precursor protein; CTGF, connective tissue growth factor; HB-EGF, heparin-binding epidermal growth factor; IGF-BP, insulin-like growth factor binding protein; IL6R, interleukin-6 receptor; RANKL, receptor activator of nuclear factor κB ligand; TGF, transforming growth factor; TNF, tumor necrosis factor.

(TNF), and CTGF also are the substrates for MMP-7 (see Table 8-2). MMP-26 degrades gelatin, type IV collagen, fibronectin, fibrinogen, and α1-proteinase inhibitor, but information about other substrates is still limited. Furin-Activated Matrix Metalloproteinases (MMP-11 and MMP-28). MMP-11 (stromelysin-3) and MMP-28 (epilysin) contain an RKRR sequence at the end of the propeptide, which is a unique motif for intracellular processing of proproteins to mature molecules by furin and other proprotease convertases. ProMMP-11 is activated intracellularly by the action of furin. MMP-11 shows only weak proteolytic activity against gelatin, laminin, fibronectin, and aggrecan, but it has respectable catalytic action in digesting α1-proteinase inhibitor, α2-macroglobulin, and...
**CHAPTER 8 | PROTEINASES AND MATRIX DEGRADATION**

103

**Membrane-Anchored Matrix Metalloproteinases**

**Type I Transmembrane-Type MMPs.** These include MMP-14 (MT1-MMP), MMP-15 (MT2-MMP), MMP-16 (MT3-MMP), and MMP-24 (MT5-MMP). All of these MT-MMPs can activate proMMP-2, but MMP-14 may play a major role in the activation of proMMP-2 in various tissues (see later). Besides the activator function, however, MMP-14 digests the triple helical portions of interstitial collagen types I, II, and III and other ECM components including fibronectin, laminin, aggrecan, and gelatin (see Table 8-2).

**MMP-15** also digests fibronectin, tenascin, nidogen, aggrecan, perlecan, and laminin.

**MMP-16** cleaves collagen type III, fibronectin, and gelatins.

**MMP-17 (MT4-MMP)** and **MMP-25 (MT6-MMP)** are GPI-linked MMPs. MMP-17 and MMP-25 can digest gelatin and fibrin/fibrinogen (see Table 8-2).

**MMP-23** (cysteine array-MMP, MIFR) is a type II transmembrane-type MMP; almost identical genes are cloned, called MMP-23A and MMP-23B. MMP-23 digests gelatin, but no information about other substrates is available (see Table 8-2). A unique aspect of MMP-23 is that this MMP is IGF-BP-1 (see Table 8-2).

**Other Secreted-Type Matrix Metalloproteinases (MMP-12, MMP-19, MMP-20, MMP-21, and MMP-27).** MMP-12 (metalloelastase), MMP-19 (RASI-1), MMP-20 (enamelysin), and MMP-21, and MMP-27 have structural characteristics similar to collagenases and stromelysins. These MMPs are not classified into the previously mentioned subgroups, however, because their substrates and other biochemical characters are not fully examined at present. Overall information on the substrate specificity of MMP-12, MMP-19, MMP-20, and MMP-27 suggests that they are stromelysin-like proteinases. MMP-12 digests elastin, fibronectin, collagen V, osteonectin, and plasminogen (see Table 8-2). MMP-19, which was originally reported as MMP-18 but renamed as MMP-19, cleaves type IV collagen, laminin, fibronectin, gelatin, tenascin, entactin, fibrin/fibrinogen, aggrecan, and cartilage oligomeric matrix protein (COMP) (see Table 8-2). MMP-20 also digests amelogenin, aggrecan, and COMP. Substrates of MMP-21 and MMP-27 are unknown, however.
expressed in only female and male reproductive organs such as endometrium, ovary, testis, and prostate, but its functions are not well established.

**ADAM and ADAMTS Families**

Two ADAM (a disintegrin and metalloproteinase) gene families exist: membrane-type ADAM with transmembrane domain (ADAM) and secreted-type ADAM with thrombospondin motifs (ADAMTS) (see Figure 8-1). The active sites in the catalytic domains of most members of both gene families contain a common sequence of HEXGXXGXXH with the "Met-turn," which also is present in MMP members.

The ADAMTS family includes 19 members. Although information about substrates and biologic functions is still limited, ADAMTS1, ADAMTS2, ADAMTS3, ADAMTS4, ADAMTS5, ADAMTS8, ADAMTS9, ADAMTS14, ADAMTS15, ADAMTS16, ADAMTS18, and ADAMTS20 all are ECM-degrading proteinases (see Table 8-4). ADAMTS1, 49 ADAMTS4, 49 ADAMTS5, 51 ADAMTS8, ADAMTS9, and ADAMTS15 can preferentially cleave aggrecan at the five Glu-X bonds including the Glu173-Ala474 bond (the aggrecanase site). Because of aggrecan-degrading activity, ADAMTS4 and ADAMTS5 are named aggrecanase-1 and aggrecanase-2, respectively; versican also is digested by these proteinases, and brevican is cleaved by ADAMTS4 (see Table 8-3). 53 The C-terminus-truncated ADAMTS4 also degrades fibromodulin and decorin. 54 ADAMTS16, ADAMTS18, and ADAMTS20 also appear to have weak aggrecanase activity. ADAMTS2 and ADAMTS3 process the N-terminal propeptides of type I and II collagens and are named procollagen N-proteinase. Activity of procollagen N-proteinase also is known with ADAMTS14. ADAMTS13 is a von Willebrand factor–cleaving proteinase, and its mutation causes thrombotic thrombocytopenic purpura. Proteinase activities of other ADAMTS species are still unknown.

The human genome contains 25 ADAM genes including 4 pseudogenes, and thus the human ADAM family is composed of 21 members (Table 8-4). Among the ADAMs, ADAM8, ADAM9, ADAM10, ADAM12, ADAM15, ADAM17, ADAM19, ADAM20, ADAM21, ADAM28, ADAM30, ADAM33, and ADAMDEC1 exhibit proteolytic activity (proteinase-type ADAMs) (see Table 8-4). Although ADAM10, ADAM12, and ADAM15 degrade type IV collagen, the main substrates of these ADAMs are various membrane proteins, which include precursors of cytokines and growth factors such as TGF-α, HB-EGF, and neuregulin, IGF-BPs, receptors such as p75 TNF receptor, IL-1 receptor II, and other membrane proteins related to development such as Notch ligand and ephrin (see Table 8-4). 55 According to these data, a major function of the ADAMs is shedding of the membrane proteins. ADAM17 cleaves a proform of TNF at the physiologic processing site into the soluble form of TNF and is called TNF-converting enzyme. ADAM17 also is involved in release of L-selectin, TGF-α, and p75 TNF receptor. 55 ADAM9, ADAM12, and ADAM17 can shed HB-EGF from its precursor. ADAM12 and ADAM28 cleave IGF-BP-3 and IGF-BP-5. CD23 is shed by ADAM8, ADAM15, and ADAM28. Other functions of ADAMs include binding to integrins, cell-cell interaction, cell migration, and signal transduction (see Table 8-4). 56

**ENDOGENOUS PROTEINASE INHIBITORS**

Endogenous proteinase inhibitors control the activities of proteinases in vivo. The inhibitors are derived from plasma or cells in the local tissues. Plasma contains several proteinase inhibitors, and about 10% of all the plasma proteins are proteinase inhibitors. Most are proteinase class specific, but α2-macroglobulin inhibits the activities of proteinases from all four groups. Major endogenous inhibitors of the ECM-degrading proteinases are listed in Table 8-5.

**α2-Macroglobulin**

The α2-macroglobulin molecule is a large plasma glycoprotein of 725 kD, which consists of four identical subunits of 185 kD that are linked in pairs by disulfide bonds. The pairs assemble noncovalently. Almost all active proteinases, regardless of the proteinase classes, bind and attack the so-called bait region, located near the center of the subunit. After cleaving within the bait region, the proteinase is physically trapped within the molecule by inducing a conformational change of the inhibitor, resulting in a proteinase/α2-macroglobulin complex. Although the proteinase in the complex remains active against small substrates, it is trapped by the arms of the α2-macroglobulin from degrading larger proteins. Besides the function as a proteinase inhibitor, α2-macroglobulin may act as a carrier protein because it also binds to numerous growth factors and cytokines such as platelet-derived growth factor, basic fibroblast growth factor, TGF-β, insulin, and IL-1β.

The α2-macroglobulin molecule is synthesized mainly in liver, but also locally by macrophages, fibroblasts, and adrenocortical cells. Concentration of the inhibitor in plasma is 250 mg/dL. Because of its large molecular weight, it is not present in noninflammatory synovial fluid. During synovial inflammation, α2-macroglobulin penetrates into the joint cavity. Rheumatoid synovial fluid has about the same concentration of the inhibitor as plasma.

**Inhibitors of Serine Proteinases**

The primary inhibitors of serine proteinases include the members of the serpin (serine protease inhibitor) gene family, Kunitz-type inhibitors, and others (see Table 8-5). The serpins are glycoproteins of 50 to 100 kD and share homology with human α1-protease inhibitor. 64 The major serpins involved in the regulation of ECM-degrading serine proteinases are α1-protease inhibitor, α1-antichymotrypsin, α1-antiproteinase, plasminogen activator inhibitors (PAI-1 and PAI-2), protein C inhibitor (PAI-3), CI-inhibitor, kalliclardin, and protease nexin-1 (PN-1). The main proteinases inhibited by these molecules are listed in Table 8-5. Although PAI-1 and PAI-2 inhibit tPA and uPA, the inhibition by PAI-1 and PAI-2 is more effective to tPA and uPA, respectively.

Kunitz-type inhibitors include aprotinin, trypstatin, and PN-2, which is identical to a β-amyloid protein precursor. Secretory leukocyte proteinase inhibitor, which inhibits...
Table 8-5  Endogenous Inhibitors of Extracellular Matrix–Degrading Proteinases

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Molecular Mass (kD)</th>
<th>Source</th>
<th>Target Enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>α₂-Macroglobulin</td>
<td>725</td>
<td>Plasma (liver); macrophages; fibroblasts</td>
<td>Most proteinases from all classes</td>
</tr>
<tr>
<td><strong>Inhibitors of Serine Proteinase</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serpins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α₁-Proteinase inhibitor</td>
<td>52</td>
<td>Plasma; macrophages</td>
<td>Neutrophil elastase, cathepsin G, proteinase 3</td>
</tr>
<tr>
<td>α₁-Antichymotrypsin</td>
<td>58</td>
<td>Plasma</td>
<td>Cathepsin G; chymotrypsin; chymase; tissue kallikrein</td>
</tr>
<tr>
<td>α₂-Antiplasmin</td>
<td>67</td>
<td>Plasma</td>
<td>Plasmin</td>
</tr>
<tr>
<td>Proteinase nexin-1</td>
<td>45</td>
<td>Fibroblasts</td>
<td>Thrombin; uPA; tPA; plasmin; trypsin; trypsin-like serine proteinase</td>
</tr>
<tr>
<td>PAI-1</td>
<td>45</td>
<td>Endothelial cells; fibroblasts; platelets; plasma</td>
<td>tPA; uPA</td>
</tr>
<tr>
<td>PAI-2</td>
<td>47</td>
<td>Plasma; macrophages</td>
<td>uPA; tPA</td>
</tr>
<tr>
<td>Protein C inhibitor</td>
<td>57</td>
<td>Plasma; urine</td>
<td>Active protein C; tPA; uPA; tissue kallikrein</td>
</tr>
<tr>
<td>C₁-inhibitor</td>
<td>96</td>
<td>Plasma</td>
<td>Plasma kallikrein; C₁ esterase</td>
</tr>
<tr>
<td>Kallistatin</td>
<td>92</td>
<td>Plasma; liver; stomach; kidney; pancreas</td>
<td>Tissue kallikrein</td>
</tr>
<tr>
<td><strong>Kunins</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aprotinin</td>
<td>7</td>
<td>Mast cells</td>
<td>Plasmin; kallikrein</td>
</tr>
<tr>
<td>Trypsatin</td>
<td>6</td>
<td>Mast cells</td>
<td>Tryptase</td>
</tr>
<tr>
<td>Proteinase nexin-2 (β-amyloid protein precursor)</td>
<td>100</td>
<td>Fibroblasts</td>
<td>EGF binding protein, NGF-γ, trypsin, chymotrypsin, factor Xla</td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLPI</td>
<td>15</td>
<td>Bronchial secretions; seminal plasma; cartilage</td>
<td>Neutrophil elastase; cathepsin G; chymotryptsin; trypsin</td>
</tr>
<tr>
<td>Elafin</td>
<td>7</td>
<td>Horny layers of skin</td>
<td>Neutrophil elastase; proteinase 3</td>
</tr>
<tr>
<td><strong>Inhibitors of Cysteine Proteinases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stefins A and B</td>
<td>11</td>
<td>Cytosol</td>
<td>Cysteine proteinases</td>
</tr>
<tr>
<td>Cystatin C</td>
<td>13</td>
<td>Body fluids</td>
<td>Cysteine proteinases</td>
</tr>
<tr>
<td>Cystatin S</td>
<td>13</td>
<td>Seminal plasma; tears; saliva</td>
<td>Cysteine proteinases</td>
</tr>
<tr>
<td>Kininogens</td>
<td>50-78/108-120</td>
<td>Plasma</td>
<td>Cysteine proteinases</td>
</tr>
<tr>
<td>Calpastatin</td>
<td>120</td>
<td>Cytosol</td>
<td>Calpains</td>
</tr>
<tr>
<td><strong>Metalloproteinase Inhibitors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIMP-1</td>
<td>28</td>
<td>Connective tissue cells; macrophages</td>
<td>MMPs</td>
</tr>
<tr>
<td>TIMP-2</td>
<td>22</td>
<td>Connective tissue cells; macrophages</td>
<td>MMPs</td>
</tr>
<tr>
<td>TIMP-3</td>
<td>21/24*</td>
<td>Fibroblasts; synovial cells</td>
<td>MMPs; ADAMs; ADAMTS</td>
</tr>
<tr>
<td>TIMP-4</td>
<td>21</td>
<td>Heart; brain; testis</td>
<td>MMPs</td>
</tr>
<tr>
<td>RECK</td>
<td>110</td>
<td>Many tissue cells; fibroblasts</td>
<td>MMP-2; MMP-7; MMP-14 (MT1-MMP); ADAM10</td>
</tr>
</tbody>
</table>

*Glycosylated form.

EGF, epidermal growth factor; NGF, nerve growth factor; PA, plasminogen activator; PAI, plasminogen activator inhibitor; RECK, reversion-inducing cysteine-rich protein with Kazal motifs; SLPI, secretory leukocyte proteinase inhibitor; TIMP, tissue inhibitor of metalloproteinases; tPA, tissue-type plasminogen activator; uPA, urokinase-type plasminogen activator.

Neutrophil elastase and cathepsin G, is present in many secretory and inflammatory fluids and in cartilage. Elafin is a serine proteinase inhibitor with 38% identity with the second domain of secretory leukocyte proteinase inhibitor; it inhibits neutrophil elastase and proteinase 3.

**Inhibitors of Cysteine Proteinases**

The members of the cystatin superfamily and calpastatin belong to the family of inhibitors of ECM-degrading cysteine proteinases (see Table 8-5). Cystatins capable of inhibiting lysosomal cysteine proteinases consist of three groups. Subgroup 1 comprises stefins A and B. Each has a molecular mass of 11 kD, and the stefins reside within cells. Subgroup 2 comprises cystatin C and S, each with a molecular mass of 13 kD. They occur at relatively high concentrations in cerebrospinal fluid and saliva. Subgroup 3 comprises the kininogens. Kininogens that participate in blood coagulation and inflammation also are inhibitors of cysteine proteinases. Calpains are not inhibited by cystatins but are inhibited by calpastatin (120 kD), which is a cytosolic-specific inhibitor of calpain.
Tissue Inhibitors of Metalloproteinases

TIMPs are a gene family consisting of four different members with approximately 40% to 50% sequence identity (i.e., TIMP-1, TIMP-2, TIMP-3, and TIMP-4), which have molecular masses ranging from 21 to 28 kD in humans.\(^{65-69}\) Virtually all TIMPs inhibit the activities of MMPs by binding in a 1:1 molar ratio to form tight, noncovalent complexes\(^{63}\) except that TIMP-1 does not inhibit efficiently MT-MMPs.\(^{41,42,70}\) TIMPs contain 12 highly conserved cysteine residues that form six intrachain disulfide bonds, which are essential for maintaining the correct ternary structure of the molecule.\(^ {67,71}\) and stable inhibitor activity.\(^ {25}\)

The TIMP molecules have two structurally distinct subdomains: an N-terminal subdomain that consists of loops 1 through 3 and a C-terminal subdomain that consists of loops 4 through 6. The N-terminal subdomain of each TIMP molecule contains the inhibitory activity for MMPs.\(^ {65}\) Studies on the crystal structures of the MMP/TIMP complexes show that the wedge-shaped TIMPs bind with their edge into the entire length of the active-site cleft of their cognate MMPs.\(^ {67}\) High affinity and efficient inhibitor activity of TIMP-2 to MMP-14 (MT1-MMP) are explained by the interaction between a quite long hairpin loop of TIMP-2 and a loop over the rim of the active-site cleft of MMP-14.\(^ {72}\)

TIMP-1 and TIMP-2 are unique in that they make the complexes with proMMP-9 and proMMP-2 (i.e., the proMMP-9/TIMP-1 and proMMP-2/TIMP-2 complexes). Similar complex formation also is known between TIMP-4 and proMMP-2. Because the complexes are made through the interaction between their C-termini,\(^ {45}\) TIMPs in the complexes retain inhibitor activity against MMPs. The activation of proMMP-9 and proMMP-2 is suppressed in the complex forms; the complex formation may be a safety device for these gelatinases.\(^ {14}\) The proMMP-2/TIMP-2 complex is useful for the efficient activation of proMMP-2 by MMP-14 on the cell membranes because MMP-14 captures proMMP-2 to the cell membranes through the trimeric complex formation between the catalytic domain of MMP-14 and the N-terminal domain of TIMP-2 (see later).\(^ {73,74}\)

Besides the inhibition and interactions of TIMPs to MMPs, TIMP-3, among the TIMPs, most efficiently inhibits the activities of ADAM10, ADAM12, ADAM17, ADAM28, and ADAM33,\(^ {75}\) although ADAM8, ADAM9, and ADAM19 are not inhibited by TIMPs. Because TIMP-3 also efficiently inhibits the aggrecan-degrading activity of ADAMTS4 and ADAMTS5, TIMP-3 may be a common tissue inhibitor of the ADAM members. The N-terminal subdomain of TIMP-3 is critical to the inhibition of the activities of ADAM members and MMPs, but the inhibition mechanism seems to be different.\(^ {76}\) TIMPs are multifunctional proteins with more diverse actions than MMP/ADAM inhibitors including growth factor activity, antian- giogenic activity, and regulatory activity of apoptosis.\(^ {69}\)

Another new MMP inhibitor is RECK (reversion-inducing, cysteine-rich protein with Kazal motifs).\(^ {77}\) RECK is a GPI-linked glycoprotein harboring three inhibitor-like domains and inhibits the activities of at least MMP-2, MMP-7, MMP-9, MMP-14, and ADAM10. Although this inhibitor seems to play a key role in the angiogenic processes in vivo, its biochemical mechanism as an inhibitor to the MMPs and ADAM10 and functions in pathologic conditions such as arthritides remain unknown.

REGULATION OF PROTEINASE ACTIVITY

The activities of ECM-degrading proteinases in tissues are regulated by the balance between the proteinases and their inhibitors. The balance at the local tissues depends on several factors including production rates of proteinases and inhibitors, their secretion, activation of proenzyme, and anchoring systems of the activated proteinases to cell surfaces. Production levels of the proteinases and inhibitors within the cells are controlled mainly by their gene expression. Activation processes of proMMPs and membrane anchoring of their activities have been established by extensive experimental work.

Gene Expression of Proteinases and Inhibitors

Matrix Metalloproteinases and Tissue Inhibitors of Metalloproteinases

Normal cells except for inflammatory cells produce only limited levels of MMPs or TIMPs in the tissues under physiologic conditions, but their expression is stimulated by many factors under pathologic conditions. Neutrophils and macrophages synthesize MMP-8 and MMP-9 during the differentiation and store them within the granules of the differentiated cells. Tumor cells express many MMPs such as MMP-1, MMP-7, MMP-9, MMP-10, and MMP-14 (MT1-MMP) and TIMP-1 predominantly by oncogenic transformation. The gene expression of MMPs and TIMPs in the tissue cells other than inflammatory cells and tumor cells is regulated by numerous factors, however, including cytokines, growth factors, and chemical and physical stimuli.

Much information is available for regulators of MMP-1 and MMP-3, which are coordinately expressed in many cell types after stimulation with cytokines and growth factors, factors acting at the cell surface, and chemical agents (Table 8-6). The induced production of MMP-1 and MMP-3 is suppressed by retinoic acid, TGF-β, and glucocorticoid. The gene expression of MMP-7 and MMP-9 is regulated by similar factors, but the regulation is stricter and fewer factors modulate the expression (see Table 8-6). MMP-14 expression is upregulated by phorbomyristate acetate (12-O-tetradecanoylphorbol-13-acetate), concanavalin A, basic fibroblast growth factor, and TNF, and it is downregulated by glucocorticoids in various cells. TNF and IL-1α stimulate osteoarthritic chondrocytes to express the MMP-14 gene.\(^ {78}\) In contrast to these MMPs, MMP-2 and TIMP-2 are unique in that factors capable of enhancing the production of MMP-1, MMP-3, and TIMP-1 are inactive.

TIMP-1 expression is enhanced or suppressed in response to many factors including cytokines, growth factors, and oncogenic transformation (see Table 8-6). Effects of these stimulatory factors are common to the gene expression of MMPs, but they are regulated independently. TGF-β, retinoic acid, progesterone, and estrogen enhance TIMP-1 expression in fibroblasts, but they suppress the expression of
MMP-1 and MMP-3. Although information about stimulating and suppressing factors of MMP-1, MMP-2, and MMP-3 is available (see Table 8-6), factors controlling the gene expression of MMP-4 are not well known. Previous studies have identified the elements in the promoters of MMPs and TIMPs, which are related to or responsible for the stimulation or suppression of the gene expression with various factors. Regulation of the gene expression is generally explained by the structural characteristics of the promoters.

**Serine Proteinases and Their Inhibitors**

Neutrophil elastase, cathepsin G, chymase, and tryptase are stored within the secretory granules and secreted into the extracellular milieu after activation of neutrophils and mast cells. The expression of these serine proteinases is controlled mainly by the cellular differentiation. Precursors of plasmin and plasma kallikrein are constitutively synthesized predominantly in liver, circulate in blood as zymogen forms (i.e., plasminogen and prekallikrein), and reach the inflamed tissues by being released from blood vessels. The proteinase activities in the tissues are controlled mainly through activation of the proenzymes by activators. The uPA and tPA molecules, activators of plasminogen, are synthesized by tissue cells, and their gene expression is regulated by many factors (Table 8-7). The uPA synthesis is upregulated in many normal cell types and in transformed cells by agents that increase intracellular cyclic adenosine monophosphate (cAMP) levels (e.g., calcitonin, vasopressin, cholera toxin, cAMP analogues); growth factors (e.g., EGF, platelet-derived growth factor, TGF, transforming growth factor; TNF, tumor necrosis factor; TPA, 12-O-tetradecanoylphorbol-13-acetate; VEGF, vascular endothelial growth factor).

**Lysosomal Cysteine and Aspartic Proteinases**

The expression of lysosomal cysteine proteinases, cathepsins B, L, and K, is generally constitutive, but cellular transformation is often associated with increased synthesis of cathepsins B and L. Cathepsin B transcription varies with cell type and the state of differentiation of tumor cells; it is increased in chondrocytes by IL-1. Malignant transformation, tumor promoters, and growth factors stimulate the

---

**Table 8-6** Factors That Modulate Synthesis of Matrix Metalloproteinases and Tissue Inhibitors of Metalloproteinases

<table>
<thead>
<tr>
<th>Enzyme or TIMP</th>
<th>Stimulating Factor*</th>
<th>Suppressing Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMP-1</td>
<td><strong>Cytokines and growth factors:</strong> IL-1; TNF; EGF; PDGF; bFGF; VEGF; NGF; TGF-α; IFN-α; IFN-γ; TNF-α; IL-6; IL-11; TPA; TGF-β</td>
<td>Retinoic acids; glucocorticoids; estrogen; progesterone; TGF-β; transmembrane neural cell adhesion molecule; cAMP; IFN-γ; adenovirus E1A</td>
</tr>
<tr>
<td></td>
<td><strong>Factors acting at cell surface:</strong> calcium ionophore A23187; cell fusion; collagen; concanavalin A; integrin receptor antibody; crystals of urate, hydroxyapatite, and calcium pyrophosphate; SPARC (osteonectin/BM 40); iron; extracellular matrix metalloproteinase inducer (EMMPRIN/CD147/basigin/M6 antigen); phagocytosis</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Chemical agents:</strong> cAMP; colchicine; cytochalasins B and D; LPS; pentoxifylline; TPA; calmodulin inhibitors; serotonin; 1,25(OH)2D3; vitamin D3; platelet-activating factor; serum amyloid A; β2-microglobulin</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Physical factors:</strong> heat shock; ultraviolet irradiation</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Others:</strong> viral transformation; oncogenes; autocrine agents; aging of fibroblasts</td>
<td></td>
</tr>
<tr>
<td>MMP-2</td>
<td>TGF-β; concanavalin A; H-ras transformation; extracellular matrix metalloproteinase inducer (EMMPRIN/CD147/basigin/M6 antigen)</td>
<td>Adenovirus E1A</td>
</tr>
<tr>
<td>MMP-3</td>
<td>IL-1; TNF; EGF; concanavalin A; SPARC (osteonectin/BM 40); LPS; TPA; extracellular matrix metalloproteinase inducer (EMMPRIN/CD147/basigin/M6 antigen); viral transformation; oncogenes; integrin receptor antibody; heat shock; calcium ionophore A23187; cytochalasin B</td>
<td>Retinoic acids; glucocorticoids; estrogen; progesterone; TGF-β; adenovirus E1A</td>
</tr>
<tr>
<td>MMP-7</td>
<td>IL-1; TNF; EGF; TPA; LPS</td>
<td>Unknown</td>
</tr>
<tr>
<td>MMP-8</td>
<td>TNF; TPA; IL-1</td>
<td>Unknown</td>
</tr>
<tr>
<td>MMP-9</td>
<td>IL-1; TNF; EGF; TGF-β; TPA; H-ras; v-Src; SPARC (osteonectin/BM40)</td>
<td>Retinoic acids; adenovirus E1A</td>
</tr>
<tr>
<td>MMP-10</td>
<td>TPA; A23187; TGF-β; EGF</td>
<td>Unknown</td>
</tr>
<tr>
<td>MMP-11</td>
<td>Retinoic acids</td>
<td>bFGF</td>
</tr>
<tr>
<td>MMP-13</td>
<td>bFGF; TNF; TGF-β</td>
<td>Unknown</td>
</tr>
<tr>
<td>MMP-14 (MT1-MMP)</td>
<td>Concanavalin A; TPA; bFGF; TNF; IL-1α</td>
<td>Glucocorticoids</td>
</tr>
<tr>
<td>TIMP-1</td>
<td>IL-1; IL-6; IL-11; TPA; TGF-β; TNF; retinoic acids; LPS; progesterone; estrogen; oncogenic transformation; viral infection</td>
<td>Extracellular matrix; cytochalasins</td>
</tr>
<tr>
<td>TIMP-2</td>
<td>Progesterone</td>
<td>TGF-β; LPS</td>
</tr>
<tr>
<td>TIMP-3</td>
<td>EGF; TGF-β; TPA; TNF; glucocorticoids; oncostatin M</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

*Factors regulating gene expression of other MMPs excluded from Table 8-6 are unknown. bFGF, basic fibroblast growth factor; cAMP, cyclic adenosine monophosphate; EGF, epidermal growth factor; IFN, interferon; IL, interleukin; LPS, lipopolysaccharide; NGF, nerve growth factor; PDGF, platelet-derived growth factor; TGF, transforming growth factor; TNF, tumor necrosis factor; TPA, 12-O-tetradecanoylphorbol-13-acetate; VEGF, vascular endothelial growth factor. |
synthesis of cathepsin L. Cathepsin K gene expression in monocyte-macrophage lineage depends on the cellular differentiation to osteoclasts, but all-trans retinoic acid upregulates the expression in rabbit osteoclasts. Lysosomal aspartic proteinase, cathepsin D, is constitutively expressed in almost all cells, although estradiol, calcitriol, and retinoic acid can regulate the expression.

**Activation Mechanisms of the Zymogens of Matrix Metalloproteinases**

All of the MMPs are synthesized as inactive zymogens (proMMPs), and activation of proMMPs is prerequisite to their functioning in vivo. ProMMPs are kept inactive by an interaction between a cysteine-sulfhydryl group in the conserved propeptide sequence PRCGXPD and the zinc ion bound to the catalytic domain, preventing the formation of a water-zinc complex that is essential to the enzymatic reaction. Activation requires proteolytic removal of the propeptide domain. There are three pathways of proMMP activation—extracellular, intracellular, and pericellular (Figure 8-2).

**Extracellular Activation**

Extracellular activation, which is applicable to many secreted MMPs (e.g., proMMP-1, proMMP-3, proMMP-7, proMMP-8, proMMP-9, proMMP-10, proMMP-12, proMMP-13), is initiated through the disruption of the Cys-Zn$^{2+}$ interaction by treatment with nonproteolytic agents or proteinases and completed by autocatalytic processing. Nonproteolytic activators used in vitro include thiol-modifying reagents (e.g., mercurial compounds, iodoacetamide, N-ethylmaleimide, oxidized glutathione), hypochlorous acid, sodium dodecyl sulfate, chaotropic agents, and physical factors (heat and acid exposure). Most of these factors, especially 4-aminophenylmercuric acetate (APMA), enable proMMP molecules to generate a short-lived intermediate, which is formed by removal of a part of propeptide by an intramolecular reaction. The fully activated form is made by an intermolecular autocatalysis cleaving three amino acids downstream from the conserved sequence PRCGXPD and leading to generation of active MMPs starting with Tyr or Phe at the N-terminus. Such a process may not be essential, however, for proMMP-9 activation by APMA because a fully active form retaining the PRCGXPD sequence is generated by the cleavage of the Ala$^{74}$-Met$^{75}$ bond upstream of the conserved sequence.

Concerning proMMP-9 activation during cerebral ischemia in vivo, nitric oxide is reported to activate proMMP-9 by S-nitrosylation.

A similar stepwise activation is proposed for the proteolytic activation of proMMPs. Proteinases initially attack the proteinase-susceptible bait regions in the propeptides and generate proteolytically active intermediates through destabilization of the Cys-Zn$^{2+}$ interaction. In the second step, the final activation site is autolytically catalyzed by the active intermediate instead of the trigger proteinases, and active MMPs without propeptides are made. In many cases, the bait region sequences in the propeptide dictate which proteinases can become an activator of a particular MMP. Potential activators of proMMPs are listed in Table 8-8. Plasmin may play a major role in the activation of proMMP-3 and proMMP-10 in vivo because treatment of these proMMPs with plasmin leads to full activation in vitro. ProMMP-1 activation by plasmin alone results in only about 25% of the potential MMP-1 activity, however, and full activation requires the subsequent cleavage of the Gln$^{6}$-Phe$^{61}$ bond by active MMP-3, MMP-7, or MMP-10. MMP-3 and MMP-10 can directly activate proMMP-7, proMMP-8, proMMP-9, and proMMP-13 into fully active forms. This intermolecular activation cascade of MMPs may be important to in vivo activation of proMMPs.

**Intracellular Activation**

Because proMT-MMPs, proMMP-23, proMMP-11, and proMMP-28 have basic motifs containing an RXKR sequence at the end of the propeptide domains, proprotein convertases such as furin, a processing enzyme in the trans-Golgi apparatus, are considered to activate these proMMPs intracellularly (see Figure 8-2). Intracellular activation of proMMP-11 and proMMP-14 (proMT1-MMP) by furin is shown. After the activation, MMP-11 is secreted from the cells and MMP-14 is expressed on the cell membranes. Because other proMT-MMPs, proMMP-23, and proMMP-28 also have the motif, furin is presumably responsible for the intracellular activation of these proMMPs.

**Pericellular Activation**

ProMMP-2 is unique in that it is activated pericellularly by MMP-14 (MT1-MMP), MMP-15 (MT2-MMP), MMP-16

---

**Table 8-7 Factors That Regulate Expression of Plasminogen Activators and Their Inhibitors**

<table>
<thead>
<tr>
<th>Enzyme or Inhibitor</th>
<th>Stimulatory Factor</th>
<th>Suppressive Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>uPA</td>
<td>TPA; IL-1; IFN-γ; EGF; PDGF; bFGF; VEGF; TGF-β; cholera toxin; cAMP; estrogen; calcitonin; vasopressin; disruption of E-cadherin–dependent cell-cell adhesion</td>
<td>Glucocorticoids; TGF-β</td>
</tr>
<tr>
<td>tPA</td>
<td>TPA; EGF; bFGF; VEGF; retinoic acids; glucocorticoids; cAMP; thrombin; plasmin; follicle-stimulating hormone; luteinizing hormone; gonadotropin-releasing hormone</td>
<td>TNF</td>
</tr>
<tr>
<td>PAI-1</td>
<td>IL-1; TNF; TGF-β; bFGF; VEGF; TPA; glucocorticoids</td>
<td>cAMP</td>
</tr>
<tr>
<td>PAI-2</td>
<td>TPA; LPS; TNF; colony-stimulating factor; cholera toxin; dengue virus</td>
<td>Glucocorticoids</td>
</tr>
<tr>
<td>PN-1</td>
<td>TPA; EGF; thrombin</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

bFGF, fibroblast growth factor; cAMP, cyclic adenosine monophosphate; EGF, epidermal growth factor; IFN, interferon; IL, interleukin; LPS, lipopolysaccharide; PAI, plasminogen activator inhibitor; PDGF, platelet-derived growth factor; PN, proteinase nexin; TGF, transforming growth factor; TNF, tumor necrosis factor; TPA, 12-O-tetradecanoylphorbol-13-acetate; tPA, tissue-type plasminogen activator; uPA, urokinase-type plasminogen activator; VEGF, vascular endothelial growth factor.
N-terminal inhibitor and C-terminal tail domains of TIMP-2 bind to the catalytic domain of MMP-14 and the C-terminal hemopexin-like domain of proMMP-2, forming a trimolecular complex of MMP-14/TIMP-2/proMMP-2 on the cell membranes (Figure 8-2). Capturing proMMP-2 by the trimolecular complex formation on the cell membranes assists proMMP-2 activation by increasing a local concentration of proMMP-2 and presenting it to the near, noninhibited MMP-14.

Dimerization of MMP-14 through an interaction of the C-terminal hemopexin-like domain is required for the efficient activation of proMMP-2 by MMP-14. MMP-14 initiates proMMP-2 activation by attacking a part of the proMMP-2 propeptide, and another already activated MMP-2 finally activates proMMP-2 by removing a residual portion of the propeptide. Integrins such as αvβ3 may be involved in the process as an additional receptor for transferring activated MMP-2 to the integrin.

MMP-15, MMP-16, or MMP-24 can activate proMMP-2 in the transfected cells, but the activation mechanisms for these MT-MMPs are not well understood. MMP-14 also activates proMMP-13 on the cell surface, and this activation does not seem to require TIMP-2.

Pericellular activation of proMMP-7 has been discovered through screening proMMP-7-binding molecules by a yeast two-hybrid system. ProMMP-7 is captured on the cell membrane by the interaction of the proMMP-7 propeptide with the C-terminal extracellular loop of CD151, a member of the transmembrane 4 superfamily, and is pericellularly activated. This new pericellular activation of proMMP-7 requires a substrate of MMP-7. Integrins such as α3β1 and α6β4, α chains of which interact with CD151, also may be involved in the activation. Although the precise molecular mechanisms of this pericellular activation system, including the activator itself, are still unclear, proMMP-7 and CD151

Table 8-8 Activators of Pro-Matrix Metalloproteinases

<table>
<thead>
<tr>
<th>ProMMP</th>
<th>Activator</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProMMP-1</td>
<td>Trypsin (partial); plasmin (partial); plasma kallikrein (partial); chymase (partial); MMP-3; MMP-7; MMP-10; MMP-11</td>
</tr>
<tr>
<td>ProMMP-2</td>
<td>MT1-MMP; MT2-MMP; MT3-MMP; MT5-MMP</td>
</tr>
<tr>
<td>ProMMP-3</td>
<td>Plasmin; plasma kallikrein; trypsin; tryptase; chymase; cathepsin G; chymotrypsin; neutrophil elastase; thermolysin</td>
</tr>
<tr>
<td>ProMMP-7</td>
<td>MMP-3; MMP-10 (partial); trypsin (partial); neutrophil elastase (partial)</td>
</tr>
<tr>
<td>ProMMP-8</td>
<td>MMP-3; MMP-10; tissue kallikrein; neutrophil elastase; cathepsin G; trypsin</td>
</tr>
<tr>
<td>ProMMP-9</td>
<td>MMP-3; MMP-2; MMP-7; MMP-10 (partial); MMP-13; trypsin; chymotrypsin; cathepsin G; tissue kallikrein</td>
</tr>
<tr>
<td>ProMMP-10</td>
<td>Plasmin; trypsin; chymotrypsin</td>
</tr>
<tr>
<td>ProMMP-11</td>
<td>Furin</td>
</tr>
<tr>
<td>ProMMP-13</td>
<td>MMP-2; MMP-3; MT1-MMP; plasmin</td>
</tr>
<tr>
<td>ProMMP-14</td>
<td>Furin</td>
</tr>
</tbody>
</table>
are overexpressed in osteoarthritic chondrocytes, and proMMP-7 is activated by the interaction with CD151 in articular cartilage of osteoarthritis.88

Pericellular Docking of Matrix Metalloproteinases

Discovery of membrane-anchored MMPs (i.e., MT1-, MT2-, MT3-, MT4-, MT5-, MT6-MMPs and MPP-23) and subsequent studies of proMMP-2 activation by MPP-14 (MT1-MMP) have established pericellular actions of these MMPs including MPP-2. Secreted MMPs were originally thought to digest ECM macromolecules extracellularly after the activation, but more recent studies have indicated the possibility that they also may function on the cell membranes through their cell surface docking.74,89,90 Besides the proMMP-2/TIMP-2/MMP-14 system, several secreted MMPs are reported to interact with cell membrane proteins, which include α2 chains of integrin α2β1 and CD147 (EMMPRIN) for MMP-1,91 αβ3 integrin and caveolin-1 for MMP-2,83,92 CD44 heparan sulfate proteoglycan and cholesterol sulfate on the cell surface for MMP-7,93,94 and CD44 for MMP-9.95 Because all of these cell membrane proteins bind to active forms of the MMPs, their proteolytic activities can be used on the cell surfaces to digest ECM and non-ECM molecules located close to the cell membranes. The importance of pericellular docking of secreted MMPs in articular tissues should be shown by further work.

JOINT DESTRUCTION AND PROTEINASES

Degradation of Extracellular Matrix in Articular Cartilage

In most joint diseases such as rheumatoid arthritis and osteoarthritis, articular cartilage is the major target tissue for destruction, and excessive degradation of cartilage ECM by proteinases is a key process in the destruction. Histologically, depletion of proteoglycans from articular cartilage (degradation of proteoglycans) is a common initial change in these joint diseases, and subsequently collagen fibrils are degraded, leading to fibrillation and laceration secondary to destruction of the arcade structures of collagen fibrils in the articular cartilage (Figure 8-3).

Aggrecan, a major proteoglycan in cartilage, is susceptible to degradation by many proteinases including MMPs, ADAMTS species, neutrophil elastase, cathepsin G, and cathepsin B. Because most of these proteinases mainly cleave the peptide bonds located in the interglobular G1-G2 domain, the major glycosaminoglycan-bearing aggrecan fragments are detached from the hyaluronan attachment site (G1 domain) after the cleavage and released from the cartilage matrix. The two major aggrecan fragments with the N-terminal sequences starting from Phe142 or Ala174 of the core protein are detected in joint fluids from patients with various inflammatory arthropathies and osteoarthritis.96 Many MMPs including MMP-1, MMP-2, MMP-3, MMP-7, MMP-8, MMP-9, MMP-13, and MMP-14 preferentially cleave the Asn141-Phe142 bond (the MMP site).97 ADAMTS species including ADAMTS1,98 ADAMTS4,99 ADAMTS5,100 ADAMTS8, ADAMTS9, and ADAMTS15 clip the Glu173-Ala174 bond (the aggrecanase site) in addition to other sites in the G2-G3 domains. Members of the MMP and ADAMTS families may play central roles in the aggrecan degradation in arthritides. On the basis of the data that synthetic MMP inhibitors do not efficiently prevent aggrecan degradation in articular cartilage, the focus has been on ADAMTS species as aggrecan-degrading proteinases in arthritides. It is not settled, however, whether MMP and ADAMTS species play differential or complementary roles in the aggrecan degradation. Decorin, a leucine-rich repeat proteoglycan, also is digested by MMP-2, MMP-3, and MMP-7100 and ADAMTS4.101 Information about proteinases responsible for the degradation of other proteoglycans including fibromodulin, lumican, biglycan, PRELP (arginine-rich leucine-rich repeat protein), chondroadherin, and syndecan present in articular cartilage is limited, although fibromodulin is cleaved by C-terminus-truncated ADAMTS4.101

Fibrillar interstitial collagens (i.e., types I, II, and III collagens) are extremely resistant to most proteinases because of their triple-helical structures. In general, classic collagenases including MMP-1, MMP-8, and MMP-13 are responsible for degradation of these collagens. Among them, MMP-13 may be most important for the degradation of cartilage collagen because it preferentially digests type II collagen.11 MMP-14 degrades types I, II, and III collagens.10 Type III collagen also is susceptible to degradation by MMP-3,10 MMP-9,10 MMP-16,10 and neutrophil elastase. Cleavage
of the telopeptides by the telopeptidase activity of MMP-3, MMP-9, neutrophil elastase, cathepsin G, and cysteine proteinase cathepsins is important for depolymerization of the cross-linked collagens. When the collagen molecules are cleaved, the helical structures are unwound at 37°C (i.e., body temperature) and become denatured into gelsins, which are digested into smaller peptides by gelatinases (MMP-2 and MMP-9) and other nonspecific tissue proteinases. Type V collagen is readily digested by MMP-2 and MMP-9. In contrast, type VI collagen is resistant to most MMPs including MMP-1, MMP-2, MMP-3, MMP-7, MMP-9, MMP-10, and MMP-14 but is susceptible to neutrophil elastase, cathepsin G, chymase, and trypsin. Type IX collagen is degraded by MMP-3 and MMP-9. Type X collagen is susceptible to MMP-1 and MMP-2, and type XI collagen is degraded by MMP-2.

Fibronectin is degraded by many MMPs including MMP-2, MMP-3, MMP-7, MMP-10, MMP-11, MMP-13, MMP-19, MMP-14, MMP-15, MMP-16, and other serine proteinases. Link protein also is susceptible to many proteinases such as MMP-1, MMP-2, MMP-3, MMP-7, MMP-8, MMP-9, MMP-10, neutrophil elastase, and cathepsin G. COMP is digested by MMP-19 and MMP-20. Proteinases capable of digesting cartilage matrix protein and cartilage intermediate layer protein are unknown, however.

**Cartilage Destruction by Proteinases in Rheumatoid Arthritis**

Articular cartilage in rheumatoid arthritis is destroyed by proteinases in three pathways: (1) destruction from surfaces of articular cartilage by proteinases present in synovial fluid, (2) destruction through direct contact of proteolytic synovium or pannus tissue or both to articular cartilage, and (3) intrinsic destruction by proteinases derived from chondrocytes (see Figure 8-3).

Rheumatoid arthritis is characterized by chronic proliferative synovitis, which shows hyperplasia of the synovial lining cells, inflammatory cell infiltration, and angiogenesis in the sublining cell layer. Hyperplastic synovial lining cells overproduce MMP-1, MMP-3, MMP-9, MMP-14, and ADAMTS4 and TIMP-1 and TIMP-3. Sublining fibroblasts produce MMP-2 and TIMP-2. Polymorphonuclear leukocytes infiltrated in the synovium and joint cavity contain MMP-8 in the specific granules and MMP-9 and neutrophil elastase, cathepsin G, and proteinase 3 in the azurophil granules. They are released from cells during phagocytosis of tissue debris and immune complexes. Other inflammatory cells in the synovium include macrophages, lymphocytes, and mast cells. Macrophages produce MMP-1, MMP-9, TIMP-1, and TIMP-2. uPA and cathepsins B, L, and D also are secreted from activated macrophages. T lymphocytes in the synovium synthesize MMP-9. Chymase and tryptase are degranulated from mast cells in response to activation by immune complexes. Endothelial cells express many MMPs including MMP-1, MMP-2, MMP-3, MMP-9, and MMP-14; tPA; and their inhibitors. These proteinases may be involved in tissue remodeling during angiogenesis in the synovium instead of cartilage destruction.

All of these proteinases and inhibitors produced by synovial tissue cells and inflammatory cells seem to be secreted into the synovial fluid and attack the surfaces of articular cartilage when active proteinases overwhelm inhibitors. MMP-1, MMP-2, MMP-3, MMP-8, MMP-9, TIMP-1, and TIMP-2 are detectable in rheumatoid synovial fluids, and the molar ratios of MMPs to TIMPs correlate with metalloproteinase activity, which is detectable in rheumatoid synovial fluids. MMPs are thought to be in favor of proteinase in rheumatoid synovial fluids. The cartilage in the central part of the articular surface shows surface irregularity (fibrillation) and proteoglycan depletion without being covered by pannus tissue even in the early stage of rheumatoid arthritis (Figure 8-4). This cartilage degradation may be ascribed to the proteolytic damage by the action of the proteinases present in synovial fluid (see Figure 8-3). Among the MMPs detected in synovial fluids, MMP-3 has the highest concentration and serum MMP-3 can be used to monitor the activity of rheumatoid synovitis.

Articular cartilage at the margins of the articular surface, to which synovial tissue can directly attach, is progressively degraded even in the early stage (see Figure 8-4). Because rheumatoid synovial lining cells exhibit strong gelatinolytic activity, which is probably generated through activation of proMMP-2 by the action of MMP-14, direct contact of the proteolytically active synovial tissue to articular cartilage is a destruction pathway of the cartilage (see Figure 8-3). Although rheumatoid synovium contains high concentrations of active proteinases, the synovium can avoid the attack by MMPs because type VI collagen, a major component in the lining cell layer, is resistant to the activities of MMP-1, MMP-2, MMP-3, and MMP-14. Pannus tissue is a connective tissue growing from the marginal transitional zone on the surface of the partially degraded articular cartilage. It is unknown whether pannus tissue formation is a sign of active destruction or repair of the articular cartilage, but the balance of data favors the

**Figure 8-4** Destruction of articular cartilage of the proximal interphalangeal joint obtained by autopsy from an early-stage rheumatoid patient. Articular cartilage shows fibrillation and proteoglycan depletion at the central part of the articular surface (small arrows) and marked destruction at the marginal area, which contacts synovium (large arrows). Alcian blue staining for proteoglycans.
former. Immunolocalization of MMP-1 and phagocytosis of collagen fibrils by pannus cells at sites of pannus-cartilage junction may suggest a role of the tissue in the cartilage destruction.

In addition to the extrinsic pathway for the cartilage damage, cartilage may be destroyed by proteinases derived from chondrocytes (see Figure 8-3). Chondrocytes under stimulation are capable of expressing various proteinases including MMP-1, MMP-2, MMP-3, MMP-7, MMP-9, MMP-13, MMP-14, MMP-16, ADAM9, ADAM10, ADAM17, ADAMTS4, and other classes of proteinases. In rheumatoid arthritic cartilage, MMP-1, MMP-2, MMP-3, MMP-7, MMP-9, MMP-13, and MMP-14 are expressed in the chondrocytes located in the proteoglycan-depleted zone. When large areas of the cartilage surface are ulcerated after degradation of cartilage ECM, death of chondrocytes occurs, leading to further progressive cartilage destruction.

Bone Resorption in Rheumatoid Arthritis

Bone is resorbed by osteoclasts even in the early stage of rheumatoid arthritis. This is commonly observed at the bare zone, where pannus-like granulation tissue invades the bone marrow and destroys subchondral bone. Activated osteoclasts attach to only mineralized bone matrix, and this cell-matrix contact is carried out between αvβ3 integrin of osteoclasts and Arg-Gly-Asp (RGD) sequence of osteopontin in the matrix. ECM degradation of the mineralized bone is possible only after demineralization of the bone matrix by proton secreted by osteoclasts because proteinases cannot permeate the matrix components in the mineralized tissues. Matrix degradation by osteoclasts is performed in the subosteoclastic compartments, which have acidic (pH 4 to 5) and hypercalcemic (40 to 50 mM Ca²⁺) conditions.¹⁰⁶

The major component of the ECM proteins in mature bone is insoluble, highly cross-linked type I collagen, although type III and V collagens also are present. Other minor components in bone matrix are leucine-rich repeat proteoglycans (decorin and biglycan) and glycoproteins such as osteopontin, osteoectin (SPARC), osteocalcin (bone Gla-protein), and thrombospondin. Among collagenolytic cysteine proteinases including cathepsins B, K, L, and S, cathepsin K is considered to be most important for bone resorption because of its collagenolytic activity with a broad pH optimum and selective expression in osteoclasts and giant cells of giant cell tumors. Mutations of human cathepsin K are responsible for pyknodysostosis, an autosomal recessive osteochondrodysplasia characterized by osteopetrosis and short stature.¹⁰⁷ Cathepsin K-deficient mice have a similar phenotype. Despite the importance of cathepsin K in bone resorption, osteoclastic bone resorption cannot be completely inhibited by cysteine proteinase inhibitors; it is inhibited to a similar degree by MMP inhibitors.¹⁰⁸ MMP-9 is highly expressed in osteoclasts in normal and rheumatoid bones and giant cells of giant cell tumors. MMP-9 has telopeptidase activity against soluble and insoluble type I collagen and strong gelatinolytic activity.¹⁰⁸,¹⁰⁹ ProMMP-9 is activated by acid exposure, and when activated, it is proteolytically active under acidic and hypercalcemic conditions.¹⁰⁸ MMP-9-deficient mice show a transient disturbance of growth plate development. Cathepsin K and MMP-9 may be involved in bone resorption in rheumatoid arthritis. Although MMP-14 is reportedly expressed in osteoclasts in rheumatoid arthritis,¹⁰⁵ evidence of the direct involvement in osteoclastic bone resorption is limited. A more recent study shows that prostate cancer–induced osteolysis is carried out via the solubilization of receptor activator of nuclear factor κB ligand by the action of MMP-7 expressed by osteoclasts in rodents.¹¹⁶ No data are available, however, for the involvement of MMP-7 in the osteoclastic bone resorption in rheumatoid arthritis.

Cartilage Destruction by Proteinases in Osteoarthritis

In osteoarthritis, no prominent inflammatory changes occur in the synovium during early stages of the disease, but elevated production of enzymes by the chondrocytes themselves contributes to the breakdown of cartilage. Many MMPs including MMP-1,¹¹¹ MMP-2,⁷⁸,¹¹² MMP-3,¹¹¹ MMP-7,¹¹⁴ MMP-8,¹¹¹ MMP-9,¹¹¹ MMP-13,¹⁰¹,¹¹¹ and MMP-14⁷⁸ are expressed in osteoarthritic cartilage. MMP-3, MMP-7, and MMP-14 are immunolocalized to chondrocytes in the proteoglycan-depleted zone of osteoarthritic cartilage, and the levels of their staining correlate directly with the histologic Mankin score.⁷⁸,¹¹³,¹¹⁴ Among the classic collagenolytic MMPs (i.e., MMP-1, MMP-8, and MMP-13), MMP-13 may be most important for degradation of the cartilage collagen because of preferential digestion of type II collagen over type I and III collagens.⁷⁸,¹¹¹ Because MMP-14 efficiently activates proMMP-2 within the osteoarthritic cartilage, and it can activate proMMP-13,⁶⁶ MMP-14 may play a key role in cartilage degradation through activation of proMMP-2 and proMMP-13 and its own proteolytic activity against cartilage ECM. Considering intermolecular activation cascade for proMMPs by active MMPs, another key MMP in osteoarthritic cartilage tissue is MMP-3, which can activate proMMP-1, proMMP-7, proMMP-8, proMMP-9, and proMMP-13. MMP-3 not only digests many cartilage ECM components such as aggrecan, type IX collagen, and link protein but also activates those proMMPs. ProMMP-7 is activated pericellularly after being anchored via the complex formation with CD151 in osteoarthritic chondrocytes.⁸⁸ Because CD151 immunoreactivity directly correlates with the Mankin score and the degree of chondrocyte cloning, and MMP-7 not only digests cartilage ECM but also sheds precursors of growth factors such as HB-EGF, MMP-7 may be involved in cartilage destruction or chondrocyte cloning or both.⁸⁸

Other proteinases implicated in cartilage destruction in osteoarthritis are members of the ADAMTS family. Chondrocytes are known to express ADAMTS1, ADAMTS4, and ADAMTS5, the latter two of which most efficiently degrade aggrecan. On the basis of the data that the cartilage destruction in experimental osteoarthritis is prevented in ADAMTS5 knockout mice, but not ADAMTS4 knockout mice, ADAMTS5 is considered to play a key role in the cartilage destruction in mice.¹¹⁵ In human osteoarthritic cartilage, however, ADAMTS4 is overexpressed, whereas ADAMTS5 expression is constitutive.¹¹⁶ The expression of ADAMTS4, but not ADAMTS5, is stimulated with IL-1 and TNF in human osteoarthritic chondrocytes.¹¹⁷ The differential role of ADAMTS4 and ADAMTS5 in human osteoarthritic cartilage remains elusive. Members of the
ADAM family (i.e., ADAM10, ADAM12, ADAM15, and ADAM17) are expressed in osteoarthritic cartilage. ADAM12 may be involved in cartilage repair because ADAM12 is overexpressed by clustered chondrocytes and plays a role in chondrocyte proliferation by promoting the availability of IGF-I through cleavage of IGF-BP-5. On the basis of the data of ADAM15 knockout mice, ADAM15 is considered to have a chondroprotective role in cartilage. RECK is overexpressed by clustered chondrocytes in human osteoarthritic cartilage and may play a role in chondrocyte migration and promotion of chondrocyte proliferation.

References

Dendritic cells (DCs) are potent antigen-presenting cells (APCs) implicated in the induction of immunity and in the maintenance of tolerance. DCs represent a sparsely distributed population of bone marrow–derived mononuclear cells that are found in most tissues of the body. In the immature state, they are primed to capture antigens through expression of several receptors that enable recognition and acquisition of foreign and self-antigens. Upon encountering pathogen-associated molecular patterns (PAMPs) or danger-associated molecular patterns (DAMPs), DCs undergo a process of “maturation,” involving changes in the DC phenotype, antigen acquisition capacity, migration, and an ability to traffic to draining lymph nodes, where they prime humoral and cellular immune responses. Under steady state conditions, DCs play an active role in maintaining tolerance to self-antigens. Although most autoimmune cells are deleted in the thymus through a process of negative selection, others must be tolerized through active and sustained mechanisms.

This chapter summarizes our current understanding of the functions of DCs, their potential role in autoimmunity, and their use as immunotherapeutic agents for autoimmune diseases such as rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE).

KEY POINTS

Dendritic cells (DCs) are professional antigen-presenting cells, abundant at body surfaces and within tissues, where they sense microbes and sample the environment for antigens.

Upon antigen capture, DCs migrate to lymphoid tissues, where they present processed antigens to naïve T cells and induce immunity or tolerance.

DCs must undergo a process of “maturation,” exemplified by antigen processing, induction of major histocompatibility complex (MHC) molecules, co-stimulatory molecules (CD80/86), and cytokine production to activate T cells.

Depending on the stimuli, maturation of DCs confers them with the ability to differentiate naïve T cells into T helper (Th)1, Th2, or Th17 cells, or regulatory T cells (Tregs). Maturing DCs also express cytokines that enable the activation of B cells and natural killer (NK) cells.

For antigen uptake, DCs express a variety of pattern recognition receptors (PRRs), such as Toll-like receptors (TLRs), C-type lectin receptors (CLR), retinoic acid inducible gene I (RIG-I)-like receptors (RLRs), nucleotide-binding oligomerization domain (NOD)-like receptors (NLRS), apoptotic cell recognition receptors (ACRs), and Fcγ receptors (FcγRs).

Two major subsets of DCs—plasmacytoid DCs (pDCs) (CD123+, CD45RA+) and conventional DC (cDCs) (CD11c+, CD45RO+)—are characterized by distinct origins, receptors, and functions. DCs can be subdivided into additional subsets based on their location and function, such as Langerhans cells.

DCs play a critical role in the maintenance of tolerance in the thymus and in the periphery. Constitutive ablation of DCs breaks self-tolerance of CD4+ T cells and results in autoimmunity.

DENDRITIC CELL SUBSETS

No single cell surface antigen uniquely recognizes all DC subsets, unlike B cells and T cells, which can be recognized by expression of surface immunoglobulin or T cell receptors, respectively. This is due to the heterogeneity of DCs, which include several distinct subpopulations, such as conventional DCs (cDCs) and plasmacytoid DCs (pDCs), producers of interleukin (IL)-12 and interferon (IFN)-α, respectively. This division of DCs into cDC and pDC subsets is likely to present an oversimplified view of DC heterogeneity. The heterogeneous nature of DCs also presents a challenge for identifying developmental lineages; however, their heterogeneity is probably a reflection of the specialized functional role of DC subsets in different tissues and antigenic challenges.
In this chapter, we concentrate on human DCs with little additional reference to murine models. Readers are encouraged to seek additional information in several comprehensive reviews.2,6

Conventional DCs (cDCs)

Conventional DCs, or myeloid DCs, herein referred to as cDCs, comprise different subsets based on location, function, and phenotype (Table 9-1). cDCs are APCs with high phagocytic activity as immature cells and high antigen presentation as mature cells with the ability to secrete large quantities of cytokines.2 They express myeloid markers (CD11c, CD33, CD13) and high levels of major histocompatibility complex (MHC)-I, MHC-II, and co-stimulatory molecules (CD80, CD86). cDCs are present in the circulatory system and can be found in virtually every peripheral tissue, as well as in lymphoid organs. They are highly migratory cells that upon maturation express chemokine receptors, such as CCR7, allowing them to move from tissues to the subcapsular sinus and the T cell zone of lymphoid organs via afferent lymphatic and high endothelial venules. In lymph nodes, cDCs can regulate T cell responses both in steady state (to induce tolerance or anergy) and during infection (Figure 9-1). They are potent producers of IL-12 and prime naïve T cells toward a Th1-like profile.

cDCs can be subcategorized into three components according to their localization: (1) peripheral tissue resident, such as skin; (2) secondary lymphoid organ resident; and (3) circulating blood cDC.5 Here we briefly describe some of the cDC subsets based on their localization. Understanding of DC subsets is critical not only in any future development of vaccines or treatments but also in understanding of autoimmune disease pathogenesis because each subset induces distinct responses.

Tissue Resident cDCs

In skin, three DC subsets can be identified: The epidermis contains only Langerhans cells (LCs), and the dermis contains CD1a+ and CD14+ DCs.

**Langerhans Cells (LCs).** LCs are located in the epidermis and have distinct markers, including CD1a and the C-type lectin receptor (CLR) langerin (CD207); LCs contain large granules called Birbeck granules. One of the pathways utilized by LCs to deliver antigen to T cells runs through langerin, which delivers antigen to Birbeck granules to be presented in a CD1a-restricted manner.7 The CD1 family of molecules is associated with presentation of glycolipids, and CD1a antigen presentation is independent of endosomal localization and acidification. LCs constitutively express E-cadherin, a homotypic adhesion molecule that anchors LCs to neighboring keratinocytes.8 Under steady state conditions, immature LCs migrate to the lymph node and induce T cell tolerance. During inflammation, migratory LCs upregulate MHC class II molecules, co-stimulatory molecules such as CD40, and CCR7, while expression of E-cadherin is downregulated, which facilitates LC disengagement from neighboring keratinocytes.8 In mice, LCs have been shown to be self-renewing in situ and are not replaced by bone marrow transfer in the steady state; data supporting their role as immunogenic APCs are conflicting.8

**Dermal DCs.** The CD14+ dermal DCs express a broad spectrum of C-type lectins, including DEC-205 (see Table 9-1). They have been shown to induced the differentiation of CD40-activated B cells into immunoglobulin (Ig) M-producing plasma cells through cytokine secretion (IL-12 and IL-6) and through direct cell-cell interaction.9 CD14+ DCs promote antibody responses through skewing of CD4+ T cells into T follicular helper (Thf) cells, which induce naive B cells to switch isotype and secrete large quantities of immunoglobulins.9

CD1a+ dermal DCs express blood DC antigen 1 (BDCA-1) and are present in the upper layer of the dermis. Phenotypically they are similar to LCs but lack expression of langerin and E-cadherin.10 On stimulation, CD1a+ DCs produce IL-15 and few proinflammatory cytokines, except IL-8, and chemokines also expressed by dermal CD14+ DCs.7 They can induce Th2 responses and cytotoxic T lymphocyte (CTL) responses, however with less potency than LCs or CD14+ DCs.

**Table 9-1 Dendritic Cell (DC) Subsets**

<table>
<thead>
<tr>
<th>Subset</th>
<th>Langerhans Cells</th>
<th>Dermal DCs</th>
<th>Blood DCs</th>
<th>Plasmacytoid DCs (pDCs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CD14+</td>
<td>CD1a+</td>
<td>BDCA1+</td>
<td>BDCA3+</td>
</tr>
<tr>
<td>TLR</td>
<td>TLR1, 2, 3, 6, 10</td>
<td>TLR2, 4, 5, 6, 8, 10</td>
<td>TLR2, 3, 4, 5</td>
<td>TLR2, 3, 8</td>
</tr>
<tr>
<td>C-type lectins</td>
<td>Langerin</td>
<td>DC-SIGN, DEC-205, LOX-1, CLEC-6, Dectin-1, DCIR</td>
<td>Langerin, DEC-205</td>
<td>CLEC9A (DNGR-1), BDCA-3</td>
</tr>
<tr>
<td>Cytokine production</td>
<td>IL-12, IL-15, IL-23, IL-6, TNF, IL-1β</td>
<td>Priming of antigen-specific CD4+, CD8+ T cells, B cells; NK cell activation via IL-12; differentiation of T helper cells</td>
<td>IFN-α, IFN-β, IL-6, TNF Induction of Tregs; induction of plasma cells; NK cell activation via type I IFN-α</td>
<td></td>
</tr>
<tr>
<td>Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BDCA, blood DC antigen; CLEC, C-type lectin; DCIR, dendritic cell inhibitory receptor; DC-SIGN, DC-specific ICAM-3 grabbing nonintegrin; DEC, dendritic and epithelial cells; DNGR, DC NK lectin group receptor; ICAM, intercellular adhesion molecule; IFN, interferon; IL, interleukin; LOX-1, lectin-like oxidized LDL receptor-1; NK, natural killer; TLR, Toll-like receptor; TNF, tumor necrosis factor.
Lymphoid Tissue DCs

DC subsets can be identified in tonsils, lymph nodes, and spleen. Intestinal DCs can be found in the lamina propria, Peyer's patches, solitary lymphoid tissue, and mesenteric lymph nodes, where they process and sample luminal and self-antigens for presentation. Here we briefly describe BDCA-3+ and CD103+ DCs.

BDCA-3+ DCs. BDCA-3+ (CD141) DCs are considered to be the counterpart of mouse CD8α+ DCs found in lymph nodes, based on functional and phenotypic analyses. BDCA-3+ DCs are found in lymph nodes, tonsils, bone marrow, and spleen, where they localize to T cell areas, but they are the rarest population of cDCs in blood. They express Toll-like receptor (TLR)3 and TLR8, and CLEC9A (DC natural killer [NK] lectin group receptor-1 [DNGR-1]), a C-type lectin that can act as a sensor of necrotic cells. These DCs phagocytose dead cells, process captured antigens, and cross-present them to CD8+ T cells. Because cross-presentation is essential for eradication of cancers, viruses, and other pathogens, BDCA-3+ DC cells present a promising target in vaccination strategies.

The chemokine receptor XCR1 also seems to be selectively expressed by BDCA-3+ DCs. Activated NK cells and CD8+ T cells produce XCL1, a ligand for XCR1, at the site of infection and may serve as a signal for recruitment of BDCA-3+ DCs. On TLR stimulation, BDCA-3+ DCs produce IL-12 and IFN-β.

CD103+ DCs. CD103+ DCs reside in the intestinal lymph node and efficiently induce expression of gut homing receptors CCR9 and α4β7 by responding T cells and B cells through a retinoic acid receptor-dependent mechanism. CD103+ DCs induce generation of regulatory T cells (Tregs) from naïve CD4+ T cells mediated through retinoic acid and transforming growth factor (TGF)-β. In Crohn's disease, CD103+ DCs maintain their ability to induce CCR9 expression by T cells, similar to healthy control DCs in small intestine; however, it remains to be determined whether they maintain their tolerogenic phenotype in Crohn's disease.

Circulating DCs

Blood contains both cDCs and pDCs. cDCs migrate first to inflammatory sites and then into secondary lymphoid tissues through afferent lymph. The physiologic roles of circulating blood cDCs might include surveillance for pathogens in circulation and tissue migration to replenish DCs or recruitment in inflammatory settings. cDCs in blood can be
Further subdivided into BDCA-1 (CD1c), CD16+, and BDCA-3+ populations. Monocytes have been shown to differentiate into dendritic cells in vitro and as well as in vivo in mice.

**BDCA-1+ (CD1c) DCs.** BDCA-1+ DCs originate from bone marrow. They circulate in the blood and migrate to secondary lymphoid organs and peripheral tissues, such as spleen and tonsil, as resting interstitial DCs. They are hypothesized to be similar to mouse CD8α+ DCs. BDCA-1+ DCs have a limited ability to produce IL-12 and CXCL10 (IFN-induced protein 10 [IP-10]) but produce IL-8, a chemotactant for granulocytes, monocytes, and lymphocytes after stimulation with poly(I: C), a synthetic double-stranded RNA. They can also produce IL-1β upon poly(I: C) stimulation, which can be mediated through the retinoic acid inducible gene I (RIG-I), rather than TLR3. Mouse CD8α+ DCs selectively express RIG-I. Similar to their mouse counterpart, they have the capacity to present antigens to CD4+ T cells but limited ability to cross-present.

**Monocyte-Derived Dendritic Cells (moDCs).** Monocytes are precursors to macrophages and can differentiate into DCs (moDCs). In the presence of granulocyte-macrophage colony-stimulating factor (GM-CSF) and IL-4, monocytes acquire dendritic cell morphology, produce inflammatory cytokines, and skew naive T cell responses in vitro. Although moDCs have not been identified in humans or mice in steady state, recent studies show that in mice, lipopolysaccharide (LPS) or LPS-expressing bacteria can induce differentiation of monocytes to moDCs. These moDCs acquire DC morphology, localize to the T cell area via CCR7, and have similar antigen-presentation capacity as classic DCs, including cross-presentation. Hence, blood monocytes can be a reservoir of DCs in response to inflammatory situations.

**Plasmacytoid DCs (pDCs)**
pDCs have high capacity for type I IFN production. pDCs are found in peripheral blood, thymus, and many lymphoid tissues. pDCs display distinct plasma cell morphology, contain abundant endoplasmic reticulum, and express CD4 and high levels of IL-3αR (CD123) but lack myeloid antigens, including CD11c, and most lineage markers. pDCs secrete high amounts of IFN-α upon viral infection but no IL-12, and although they are primed to capture virus (e.g., flu), they generally display poor antigen capture and presentation capacity. Upon activation, pDCs show similar characteristics to cDCs (e.g., dendritic morphology, high expression of MHC class II molecules) but a limited capacity to prime naive T cells. pDCs can prime naive T cells toward the Th1 profile in IFN-α-dependent and IL-12-independent pathways upon viral/TLR stimulation but can also induce Tregs in an indoleamine-2,3-dioxygenase (IDO)-dependent manner, possibly to contain immune activation that is triggered, as in SLE or chronic human immunodeficiency virus (HIV) infection. In contrast to cDCs, pDCs contain “lymphoid” mRNA transcripts for pre-T α chains, germline IgK, and Spi-B.

pDCs have been shown to accumulate in inflamed tissues, as in early stages of psoriasis, through their expression of ChemR23 and CXCR4.

**DENDRITIC CELL DEVELOPMENT**
The developmental origin of pDCs and cDCs is still debated because pDCs and cDCs can be derived from both the common lymphoid progenitor (CLP) and the common myeloid progenitor (CMP), suggesting that pDCs and cDCs may arise during hematopoiesis from progenitors with already distinct and restricted lineage potential. Most studies on the developmental origin of human DC subsets have used in vitro culture systems. DC precursors, similar to other hematopoietic cells, are continuously produced at a steady rate in a pathogen-independent manner from CD34+ hematopoietic stem cells (HSCs) within the bone marrow. DC development arises through several commitment steps in the bone marrow, from hematopoietic precursors, including macrophage/DC precursors (MDPs) and common DC precursors (CDPs). Proliferative CDPs then differentiate into DC precursors (pre-DC) and pDCs. Pre-DCs localize in bone marrow, blood, and lymphoid organs, where they give rise to immature DCs (imDCs) and mature DCs (mDCs). The pre-DC subset expresses several myeloid markers, including CD11b, CD11c, CD13, CD14, and CD33.

DC development is dependent on cytokines and specific transcription factors. Fms-like tyrosine kinase-3 ligand (flt-3-L) is critical for the development of both cDCs and pDCs. cDC differentiation is dependent on the transcription factors Ikaros and PU.1, whereas pDC development is dependent on IRF8, E2-2, and the Ets family transcription factor Spi-B, and probably PU.1.

**Maturation**
In their resting state, imDCs are highly specialized in antigen uptake through a variety of receptors and mechanisms. Upon encountering a pathogen or other “activation stimulus,” DCs undergo phenotypic and functional changes referred to as maturation; this influences their “effector” function and antigen-presenting capability (see Figure 9-1).

**Antigen Recognition**
imDCs take up and recognize and receive signals from antigens through phagocytosis, macropinocytosis, and endocytosis using Fc receptors (Fcγ receptor types I [CD64] and II [CD32]), integrins (αVβ3 or αVβ5), C-type lectins (mannose receptor DEC-205), Toll-like receptors (TLRs), nucleotide-binding oligomerization domain (NOD) proteins, and RIG-I-like receptors (discussed later). Upon antigen recognition, DCs secrete cytokines and express chemokine receptors to facilitate their migration.

**Receptor Upregulation**
Maturation is accompanied by increased expression of chemokine receptors, adhesion molecules, and co-stimulatory molecules that allow DC migration toward secondary lymphoid organs, facilitating interaction with T cells. DCs upregulate CCR7, which directs cells to lymphoid tissue in response to chemokines CCL19 and CCL21 expressed in the T cell areas of lymph nodes. Adhesion molecules such
as intercellular adhesion molecule-1 (ICAM-1) are also upregulated upon maturation and bind lymphocyte function–associated antigen-1 (LFA-1) on T cells. Mature DCs are also characterized by high-level expression of MHC and T cell co-stimulatory molecules with an enhanced ability to present antigen captured in the periphery to T cells. DCs express various members of the B7 superfamily, such as B7-1 (CD80) and B7-2 (CD86), which stimulate or inhibit T cell activation, respectively, upon binding T cell CD28 or CTLA-4. DCs constitutively express CD40, which is upregulated upon maturation and binds CD40L (CD154) expressed by activated CD4+ T cells, resulting in enhanced upregulation of MHC molecules and increased production of cytokines and DC survival.

Cytokine and Chemokine Secretion

cDCs predominantly produce IL-12, and pDCs IFN-α, although BDCA-3+ DCs can also produce IFN-β. DCs also secrete chemokines, such as regulation upon activation normal T cell expressed and presumably excreted (RANTES), macrophage inflammatory protein (MIP)-1α, and IP-10; this enables recruitment of T cells, monocytes, NK cells, and other DCs into the local environment. Concomitantly, DCs reduce their capacity for phagocytosis, macropinocytosis, and antigen processing (discussed later). After migration to secondary lymph nodes, DCs prime naïve CD4+ and CD8+ T cells.

ANTIGEN ACQUISITION AND RECOGNITION

DCs are defined by their high capacity to capture, process, and present antigens—a prerequisite for T cell priming. DCs efficiently sample their environment through pattern recognition receptors (PRRs), which recognize PAMPs. PRRs serve as an important link between innate and adaptive immunity as they directly mature DCs and consequently dictate the production of cytokines and chemokines. PRRs can be divided, among others, into TLRs, CLR, NOD-like receptors (NLRs), and RIG-I-like receptors (RLRs). Another set of receptors important for homeostasis of tissues includes apoptotic cell recognition receptors involved in the clearance of apoptotic cells and Fc receptors, which recognize antigen-antibody complexes.

Toll-Like Receptors (TLRs)

TLRs recognize a variety of PAMPs, such as bacterial, fungal, and nucleic acids (Table 9-2). They can be grouped into two subfamilies: TLRs expressed on cell surface (TLR1, -2, -4, -5, -6) sampling for the presence of bacterial, fungal, and protozoan components; and TLRs localized to specialized endosomal compartments (TLR3, -7, -8, and -9), detecting the presence of nucleic acids (see Chapter 18). Although most TLRs appear to function as homodimers, TLR2 forms heterodimers with TLR1 or TLR6, and TLR4 is complexed with MD-2. DC subsets show a heterogeneous TLR expression pattern (Table 9-2).

Activation of DCs in response to binding of TLR agonists is controlled by signaling through the Toll/IL-1 receptor (TIR) domain present in the cytoplasmic domain of TLR. Syncereption downstream of the TIR domain is mediated through recruitment of myeloid differentiation factor 88 (MyD88) for all TLRs, except for TLR3. MyD88 recruitment is required for induction of nuclear factor κB (NFκB), which induces expression of proinflammatory cytokines such as IL-12, tumor necrosis factor (TNF), and IL-6. In pDCs, stimulation of TLR7 and TLR9 with nucleic acid induces IFN response factor 7 (IRF7) through MyD88 signaling, leading to IFN-α secretion.

In TLR3 signaling, the TIR domain of TLR3 recruits the adapter protein, TIR domain–containing adapter protein, inducing IFN-β (TRIF), which induces IRF3, leading to production of IFN-β.

TLRs play an important role in induction of autoimmunity. Under steady state conditions, DC responses to extracellular self-nucleic acids are prevented by the endosomal seclusion of nucleic acid–recognizing TLRs. However, cellular proteins and chaperones can bind and present self-nucleic acids to DCs, causing inflammation. For example, LL37 (cathelicidin), an antimicrobial peptide produced by neutrophils and keratinocytes, is found in high levels in the skin lesions of psoriasis patients. LL37 forms aggregates with self-DNA and RNA from necrotic cells, protects them from degradation, and transports them into endosomal compartments of DC. In pDCs, RNA-LL37 and DNA-LL37 complexes activate TLR7 and TLR9, respectively, and trigger the secretion of IFN-α. In contrast, RNA-LL37 complexes in cDCs trigger TLR8 and induce secretion of proinflammatory cytokines. Similarly, high mobility group B1 (HMGB1) is released from necrotic cells and stimulates

Table 9-2  TLR Expressed by Dendritic Cells (DCs) and Their Ligands

<table>
<thead>
<tr>
<th>Conventional DCs</th>
<th>Plasmacytoid DCs</th>
<th>Ligands</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLR1</td>
<td>TLR1</td>
<td>Multiple triacyl lipopeptides</td>
</tr>
<tr>
<td>TLR2</td>
<td></td>
<td>Peptidoglycan (Staphylococcus aureus), lipoproteins, and lipopeptides from several bacteria</td>
</tr>
<tr>
<td>TLR3</td>
<td></td>
<td>Glycosphingolipidinostol anchors from Trypanosoma cruzi</td>
</tr>
<tr>
<td>TLR4</td>
<td></td>
<td>Lipoaminomannan from Mycobacterium tuberculosis, zymosan (yeast)</td>
</tr>
<tr>
<td>TLR5</td>
<td></td>
<td>Double-stranded RNA (e.g., poly I:C)</td>
</tr>
<tr>
<td>TLR6</td>
<td>TLR6</td>
<td>Lippolysaccharide + MD-2, taxol, HSP60 (?), heparan sulfate (?), RSV, fibronectin</td>
</tr>
<tr>
<td></td>
<td>TLR7</td>
<td>Flaggelinn (Salmonella typhimiturum, Usteria)</td>
</tr>
<tr>
<td></td>
<td>TLR8</td>
<td>?/Undergoes dimerization with TLR2</td>
</tr>
<tr>
<td></td>
<td>TLR8</td>
<td>Imiquimod (Alda), R848 (resiquimod), single-stranded RNA</td>
</tr>
<tr>
<td></td>
<td>TLR9</td>
<td>Imiquimod (Alda), R848 (resiquimod), single-stranded RNA</td>
</tr>
<tr>
<td></td>
<td>TLR10</td>
<td>CpG ODNs, DNA from bacteria and viruses, chromatin-IgG complexes</td>
</tr>
</tbody>
</table>

CpG ODNs, CpG oligodeoxynucleotides; HSP, heat shock protein; Ig, immunoglobulin; RSV, respiratory syncytial virus; TLR, Toll-like receptor.
TLR2 and TLR4, mediating secretion of proinflammatory cytokines. Heat shock proteins (HSPs), such as HSP60, HSP70, gp96, and HSP22, are also reported to be recognized by TLR2 and TLR4.

C-Type Lectin Receptors (CLRs)

DCs can also recognize, bind, and uptake specific carbohydrate moieties expressed by pathogens or self-tissues through CLRs such as the macrophage mannose receptor (MMR), DEC-205, DC-specific ICAM-3 grabbing nonintegrin (DC-SIGN), Dectin-1, Dectin-2, dendritic cell inhibitory receptor (DCIR), myeloid inhibitory C-type lectin-like receptor (MICL, CLEC12A), CLEC9A, and BDCA-2, all of which can function as endocytic receptors to internalize antigens for degradation or antigen processing and presentation.

CLRs Modulate TLR Signaling

CLRs can positively and negatively modulate TLR responses. DC-SIGN recognizes mannose and fructose on pathogens such as HIV-1, measles virus, Mycobacterium tuberculosis, and Candida albicans. Modulation of TLR signaling by DC-SIGN is dependent on prior activation of NFκB by TLR. Acetylation of NFκB p65 mediated by DC-SIGN prolongs and increases transcriptional activation from the IL-8 and IL-10 promoters, leading to increased cytokine secretion.

On the contrary, binding of Salp15, a protein antigen from the tick Ixodes scapularis, to DC-SIGN decreases TLR1-TLR2-dependent inflammatory cytokine production by decaying IL-6 and TNF mRNA while impairing nucleosome remodeling at the IL12a promoter.

DC-SIGN has an essential role in DC–T cell interaction through ICAM-2 and ICAM-3, which are highly glycosylated structures. In RA synovium, the numbers of cells expressing DC-SIGN on macrophages and DC and ICAM-3 on naïve T cells were increased in the affected joints, with interaction leading to increased secretion of matrix metalloproteinases.

BDCA-2 expressed by pDC pairs with FcεRIγ, which contains an immunoreceptor tyrosine-based activation motif (ITAM). Activation of BDCA-2 by antibody ligation results in downregulation of TLR9-mediated IFN-α, IFN-β, TNF, and IL-6 in pDCs. HIV gp120 was also found to inhibit TLR9-mediated IFN-α through binding of BDCA-2 on pDCs. SLE patients have reduced expression of BDCA-2 on pDCs, and lack of negative regulation through BDCA-2 signaling may account for excessive IFN-α production on TLR stimulation.

DCIR and MICL are the only known CLRs with immunoreceptor tyrosine-based inhibitory motifs (ITIMs) in their cytoplasmic tails, which recruits SH2-domain-containing protein tyrosine phosphatase 1 (SHP-1) or SHP-2 following antibody ligation. So far no ligands have been identified for DCIR and MICL. Activation of DCIR leads to internalization into an endosomal compartment, with TLR8 and TLR9 leading to the downregulation of TLR8-induced TNF and IL-12 by cDCs, or TLR9-induced IFN-α and TNF production by pDCs. Polymorphisms in the gene encoding DCIR have been associated with RA. In mice, DCIR deficiency leads to unrestricted expansion of DC numbers and development of autoimmune sialadenitis, enthesis, and autoantibodies, including rheumatoid factor, anti-Ro and La, and antinuclear antibodies, implicating DCIR in the negative regulation of DC expansion or differentiation from bone marrow precursors.

Similarly, antibody-mediated cross-linking of MICL expressed by cDCs leads to SHP-1 and SHP-2 recruitment and the suppression of TLR-induced IL-12 production.

CLR Signaling Independent of TLR

Dectin-1 recognizes β-1-3 glucan expressed by fungal pathogens. Dectin-1 induces two different signaling pathways: activation by recruitment of spleen tyrosine kinase (SYK)-dependent activation of NFκB, and activation of RAFl, leading to phosphorylation of p65, similar to DC-SIGN and TLR cross-talk. Activation and modification of NFκB by dectin-1 leads to an increase in IL-6, IL-10, IL-1β, and IL-12p40 production and modulation of IL12p35 and IL-23p19 subunits, leading to bioactive IL-12p70 and IL-23, respectively, resulting in generation of Th1 and Th17 cells.

Dectin-2, a CLR, can associate with Fcγ receptors (FcγRs) similar to BDCA-2. Unlike BDCA-2, Dectin-2 induces cytokine production through FcγRs. Dectin-2 recognizes fungal hyphae from C. albicans, Trichophyton rubrum, and Microsporum audouini and recruits SYK to FcγR and induces production of TNF and IL-6.

CLRs Enhance Antigen Presentation

The ability of CLRs to deliver antigen to different compartments for processing has been used for targeting in vaccine studies. It has been shown that antigen linked to DEC-205–specific antibodies resulted in enhanced antigen uptake and presentation by DCs to both CD4+ and CD8+ T cells. CLRs have a specific expression pattern, and some CLRs, such as DC-SIGN and Dectin-1, are expressed by several DC subsets, while expression of other CLRs is restricted to specific DC subsets (see Table 9-1). CLEC9A recognizes necrotic cells and on binding recruits Syk; in mice this activation is required for cross-presentation of necrotic cell–associated antigens by CLEC9A. BDCA-3* cells in humans express CLEC9A and have been shown to cross-present viral antigens to CD8 T cells and could potentially play a part in cross-presentation of self-antigen in autoimmune diseases.

Because CLRs are involved in immune homeostasis, their dysregulation might contribute to the predisposition to autoimmune disease. This is supported by associations of single-nucleotide polymorphisms of CLR and key components of their signaling pathways with a number of autoimmune diseases, including CLEC16A with type 1 diabetes, multiple sclerosis, juvenile rheumatoid arthritis, and inflammatory bowel disease; DCIR with rheumatoid arthritis; and CARD9 with the spondyloarthropathies, ankylosing spondylitis, and inflammatory bowel disease.

Retinoic Acid Inducible Gene 1 (RIG-I)-Like Receptors (RLRs). RLRs are cytoplasmic PRRs, recognizing genomic or replication intermediate viral double-stranded RNA (dsRNA). The RLR family is composed of RIG-I,
melanoma differentiation-associated gene 5 (MDA5), and LGP2. RIG-I and MDA5 recognize short dsRNA with 5’-triphosphate ends and long dsRNA, respectively. LGP2 functions as a positive regulator of both RIG-I and MDA5. The CARDs of RLR activate a signaling cascade, leading to expression of type I IFN genes via TRAF3 and IRF3/7. In addition, IPS-1 induces nuclear translocation of NFκB via cleavage of caspases 8 and 10.

**Nucleotide-Binding Oligomerization Domain (NOD)-Like Receptors (NLRs).** The NLR family of proteins consists of cytosolic, intracellular PRRs that recognize PAMPs and endogenous ligands, which induce a signaling cascade leading to activation of NFκB, or a cytoplasmic multiprotein complex known as the inflammasome, to produce inflammatory cytokine. NLR family proteins have a tri-modular structure and contain the following domains: a central nucleotide-binding NOD domain, essential for self-oligomerization, and a C-terminal domain consisting of tandem leucine-rich repeats (LRRs), essential for sensing and recognition of PRR.

Some proteins contain variable N-terminal domains. NOD1 contains a CARD domain, which activates transcription of proinflammatory mediators, whereas NACHT-LRR-PYD-containing protein 3 (NALP3) contains a pyrin domain (PYD), and the NLR family apoptotic inhibitory protein 5 (NAIP5) contains a baculovirus inhibitor of the apoptosis protein repeat (BIR) domain. PYD and BIR domains are components of the inflammasome that regulate caspase-1 activation, leading to the production of IL-1β.

NOD1 and NOD2 sense various bacterial PAMPs in the cytoplasm. Upon PAMP recognition, they undergo self-oligomerization through the CARD domain and form a complex with CARD-containing IL-1β-converting-enzyme–associated kinase (CARDIAK). This induces inflammatory cytokines through activation of NFκB and mitogen-associated protein kinases (MAPKs).

Pathogen infection of DCs results in the production of various cytokines of the IL-1 family, including IL-1β, IL-18, and IL-33 through TLR, NLR, and RLR. The production of these cytokines is dependent on PRR-mediated translation and transcriptional upregulation of pro-forms of these cytokines. Inflammasomes, activated by caspase-1, process cytosolic pro-IL-1β and pro-IL-18 to bioactive and secretory cytokines. Inflammasomes are categorized into three types: NALP3, IL-1β-converting-enzyme–protease-activating factor (IPAF), and NALP1 inflammasome. They are activated directly or indirectly by various ligands. The NALP3 inflammasome can be activated by exogenous ligands such as viral RNA and by host endogenous ligands, including monosodium urate (MSU) and adenosine triphosphate (ATP). Gout, an acute and chronic inflammation of joints, is associated with the deposition of MSU, which most likely induces inflammation through NALP3 inflammasomes.

NLRs are also involved in various inflammatory and autoimmune diseases. NOD2 signaling has been shown to induce Th17 responses for clearance of bacteria through the selective induction of IL-23 and IL-1, which is abrogated in DCs from Crohn’s disease patients carrying a mutation in NOD2. ATG16L1, an autophagy-related gene, was also identified in Crohn’s disease. In mice, deletion of the ATG16L1 gene leads to hyperactivation of inflammasomes and subsequent secretion of bioactive IL-1β and IL-18, contributing to intestinal inflammation. Therefore inappropriate responses to bacterial pathogen by DCs may contribute to the pathogenesis of Crohn’s disease.

**Apoptotic Cell Recognition Receptors**

In addition to sampling for immunogenic antigens, DCs help in clearance of apoptotic cells in the periphery. Upon uptake of apoptotic cells or apoptotic microparticles, DCs maintain an immature phenotype and induce immunomodulatory effects. These DCs downregulate co-stimulatory molecules (CD80/86) and produce TGF-β, a cytokine necessary for differentiation of naïve T cells to Tregs. In contrast, uptake of necrotic cells leads to maturation and secretion of proinflammatory cytokines. Inability to clear apoptotic cells and consequent secondary necrosis have been implicated in the development of SLE. Recognition and engulfment of apoptotic cells involve multiple ligands: “eat me” signals, bridging molecules, and phagocytic receptors. One of the principal “eat me” signals is translocation of phosphatidylserine (PS) to the outer leaflet of the plasma membrane of apoptotic cells. PS can be recognized by multiple receptors, such as T cell immunoglobulin mucin (TIM-1), TIM-4, Stabilin-2, brain-specific angiogenesis inhibitor 1 (BAI1), and scavenger receptors (e.g., CD36, CD68, LOX-1).

In addition, the β2-glycoprotein I (β2-GPI) receptor, αv integrins, and MER tyrosine kinase recognize PS through bridging molecules β2-glycoprotein 1 (β2-GPI), milk fat globular-EGF factor 8 protein (MEGF-8), and Gas6, respectively. CD36 and αβ3 on DCs bind apoptotic cells via bridging molecule thrombospondin-1 (TSP-1); complement receptor 3 (CR3 [CD11b/CD18]) and CR4 (CD11c/CD18) via complement protein C3b; and scavenger complex calreticulin-CD91 via surfactant proteins A (SP-A), SP-D, and C1q. CD14 and ATP-binding cassette transporter (ABC) 7 have also been implicated in recognition of apoptotic cells. In addition to providing “eat me” signals, apoptotic cells modify CD31, which prevents ingestion of viable cells.

DCs can also phagocytose necrotic cells; LOX-1 binds to heat shock proteins (HSPs) and allows cross-presentation to T cells of antigens bound to HSP. Recently, variants of ITGAM (CD11b) were associated with SLE. Polymorphisms in C1q, a complement factor, have also been linked with SLE. Polymorphisms in C1q and CD11b could lead to decreased uptake of apoptotic cells and subsequent necrosis and proinflammatory responses.

A common feature of lupus patients is the presence of dying cells in both lymph nodes and inflamed tissues. It is not clear whether this is due to increased cell death or innate or acquired inability to clear apoptotic cells, which normally occurs very rapidly. Increased cell death might be the consequence of oxidative stress or infection, both of which are thought to trigger autoimmune disease. Regardless of the mechanism, increased cell death increases the presence of autoantigens available for presentation by DCs. Necrotic cells attract phagocytes by secreting hyaluronic and uric acid and high mobility group box 1 protein (HMGB1), which are known inducers of inflammasome and inflammation.
**Fc Receptors (FcRs)**

Fc receptors are expressed by many immune cells and are important in regulation of immune responses to immune complexes (ICs), a complex of antigen bound to antibody and sometimes to components of the complement system. DCs express activating receptors FcγRI, FcγRIIA, FcγRIIIA, and inhibitory receptor FcγRIIB.

Infected cells or pathogens coated with IgG activate FcγR-mediated clearance by antibody-dependent cell-mediated cytotoxicity (ADCC) or phagocytosis and/or indirectly through the release of cytokines. This is mediated by ITAM phosphorylation and subsequent activation of SYK and the initiation of the downstream signaling cascade resulting in cell activation.58 In contrast, inhibitory FcγRIIB has an ITIM domain in its cytoplasmic domain, resulting in inhibition of phagocytosis and cytokine secretion.

FcγRIIB also controls IC-mediated DC maturation.59 Low levels of IC can be identified in the serum of healthy individuals, thus FcγRIIB can prevent spontaneous DC maturation. In a small subset of well-controlled RA patients who were able to stop disease-modifying drugs, peripheral blood monocyte-derived DCs expressed high levels of FcγRIIB, signaling through which inhibited TLR4-mediated DC activation.60 In addition, ICs taken up by FcγRIIB are degraded inefficiently and are presented on the cell surface in native conformation, where they can interact with B cells.61 An FcγRIIB gene polymorphism has been identified in SLE and RA patients. DCs from RA patients with FcγRIIB variant showed an increase in IC-mediated inflammation in vitro.62

FcγRs also alter CR3-mediated signaling of phagocytosis upon co-stimulation in vitro. Co-stimulation of CR3 with FcγRI inhibits and FcγRIIA enhances CR3-mediated phagocytosis—a potential mechanism for selective modulation of immunity.

**ANTIGEN PRESENTATION**

DC activation is tightly associated with processing and presentation of antigens by MHC class I and MHC class II, and lipid presentation by CD1 molecules (Figure 9-2).
MHC Class I Antigen Presentation

MHC I molecules are expressed in almost all cell types and present cytosolic proteins, including cytosolic viral or bacterial proteins, to CD8+ T cells. They are composed of heavy chains and β2 microglobulin, an invariant light chain.

Antigenic peptides are predominantly generated from the ubiquitin-proteasome system of ubiquitinated nascent proteins, misfolded proteins (retrotranslocated from endoplasmic reticulum), or the cytosol through the ER-associated degradation pathway (ERAD), neosynthesized defective proteins, defective ribosomal products (DRiPs), and proteins expressed in the cytoplasm. IFN-γ strongly influences processing efficiency by inducing immunoproteasome formation and proteasome activator PA28 synthesis. DCs express immunoproteasomes containing the active site subunits LMP2, LMP7, and MECL-1, which enhance antigen processing and the generation of a different spectrum of peptides from the standard proteasome. Recently, intermediate immunoproteasomes were identified; they contain only one or two inducible catalytic subunits of immunoproteasome and confer different cleavage specificities. This broadens the repertoire of antigens presented to CD8+ T cells.

The peptides are transported to the ER through the transporter associated with antigen processing (TAP), where long peptides are further trimmed by ER aminopeptidase-I (ERAP-1) to 8-mer or 9-mer peptides for loading onto MHC class I molecules. The trimeric complex of MHC class I heavy chains, β2 microglobulin, and peptide allows for optimal folding, glycosylation, and delivery to the cell surface (see Figure 9-2).

Cross-Presentation

DCs can acquire antigens exogenously and process them for presentation by MHC class I molecules, referred to as cross-presentation. This property is atypical because MHC class I molecules exclusively present endogenous proteins in most cells. Cross-presentation requires peptides to gain access to cytosol, to be processed by proteasome, and to be transferred to the ER for loading onto MHC I class molecules (see Figure 9-2C). Although the precise mechanism of cross-presentation remains controversial, it is well established that DCs utilize this process to activate CD8+ T cells. DCs may acquire antigens in the form of apoptotic cells, necrotic cells, antibody-opsonized cells, immune complexes, heat shock proteins, and exosomes, and even by nibbling of live cells. Autophagosomes may be another source of peptides for MHC class I molecules. Autophagy clears ubiquitinated cytoplasmic protein aggregates. Activation of DCs through TLR or exposure to stress induces DC aggresome-like structures (DALIS) containing polyubiquitinated proteins; peptides derived from DALIS can also be presented by MHC class I molecules.

MHC Class II Antigen Presentation

The MHC class II pathway is constitutively expressed only by APCs. MHC class II αβ heterodimers rely on a specialized type II transmembrane chaperone protein, the invariant chain (Ii), for stable assembly in the endoplasmic reticulum. Ii stabilizes the MHC II complex and contains an endosomal sorting and retention signal. MHC class II molecules are transported and concentrated in multivesicular and multilamellar late endosomal compartments called the MHC class II-containing compartment (MIC).

Antigens are endocytosed and retained within phagosomes before fusing with lysosomes to form phagolysosomes. Concomitant TLR signals induce activation of the vacuolar proton pump that enhances lysosomal acidification and antigen proteolysis within phagolysosomes. Acidification of this compartment allows optimal activity of cathepsins. Cathepsin S degrades the cytoplasmic tail of Ii, leaving a short peptide, the MHC class II–associated invariant-chain peptide (CLIP), bound to the peptide-binding groove and thus protected from proteases. CLIP is replaced by an antigenic peptide in the endosome derived through the action of endosomal proteases on essentially any proteins accessing the endosome. The rate of dissociation of CLIP is too slow for quantitative peptide loading; the catalyst-chaperone protein HLA-DM accelerates the rate of CLIP release and peptide exchange in MIC. The transport of loaded MHC class II molecules is thought to involve the transformation of MIC into tubular structures directed toward the site of T cell interaction at the plasma membrane.

TLRs regulate phagosome maturation, enhance lysosomal acidification, and increase antigen uptake transiently. MHC class II surface expression and turnover rates are regulated by cytoplasmic domain ubiquitination in DCs. This explains the expression of low levels of MHC class II molecules by immature DCs and their increase in half-life after maturation, sustaining antigen presentation after migration into secondary lymphoid organs.

RA is characterized by autoantibodies against citrullinated proteins. A possible mechanism of generation of the immune response against citrullinated proteins has been recently described in mice. It has been shown that two distinct conformers of MHC-peptide complexes are known for the same peptide; these are referred to as type A for stable conformation and type B for less stable conformation. Peptides generated from intact protein are loaded onto MHC class II complex in the late endosome/lysosome. In this compartment, HLA-DM catalyzes peptide loading and acts as a conformational catalyst, hence only more stable type A conformers are loaded. In comparison, the same peptide may be loaded exogenously at the cell surface and in the early endosome in exchange for poorly binding peptides occupying MHC class II molecules in the absence of HLA-DM. These mechanisms of loading give rise to both stable type A and flexible type B conformers. Hence only type A T cells are primed with native protein, and exogenous peptide gives rise to priming of both type A and B T cells. In the thymus, T cells recognizing the type A conformation for a peptide are deleted, whereas type B T cells escape thymic negative selection. Thus self-reactive type B T cells are present in the naïve peripheral repertoire and have potential for autoimmune activation. This was recently shown in nonautoimmune prone mice after immunization with unmodified hen egg lysozyme (HEL); T cells specifically reactive to citrullinated epitopes of HEL were present among the responding repertoire to immunization. The citrullinated HEL epitopes were processed and loaded onto MHC class II complexes and were presented to T cells by
DCs.\(^9\) Hence naturally occurring self-reactive T cells, which have evaded thymic negative selection, may recognize citrullinated peptides loaded as flexible type B conformers onto MHC class II molecules and subsequently provide help for B cell–mediated antibody responses against citrullinated proteins.

### Lipid Presentation

CD1 molecules present lipid antigen to T cells. The CD1 family of MHC I-like glycoproteins includes CD1a-d on the surfaces of DCs and CD1e, which remains in the endoplasmic reticulum. CD1a-c–restricted T cells express CD4 or CD8 or lack both CD4 and CD8 (double negative). By contrast, most CD1d–restricted T cells express a semi-invariant T cell receptor (TCR) and markers of NK cells and are identified as invariant NK T (iNKT) cells.\(^80\)

Similar to MHC I and MHC II molecules, CD1 molecules assemble in the ER, where they noncovalently bind the chaperones calnexin, calreticulin, ERP57, and β2 microglobulin.\(^81\) Lipid antigens are generated by lysosomal hydrolases of internalized phagosomes, membrane-bound lipids in clathrin-coated vesicles, or internalized apoptotic bodies or exosomes and are loaded onto CD1 molecules.

### T CELL ACTIVATION

cDCs interact with naïve T cells upon arrival in the lymph node and with effector memory T cells at inflammatory sites. DC–derived signals that determine the outcomes of these interactions include the following: antigen presentation in the context of MHC molecules (Signal 1), co-stimulatory molecules, including CD80 and CD86 (Signal 2), and cytokines (Signal 3). Expression of these signals by DCs is induced by exposure to PAMPs or DAMPs. Some cell surface molecules are inhibitory, such as PD1 through negative interactions with T cell PD1, or CD80/CD86 with T cell CTLA-4. Signal 3—provided in cis by APCs to antigen-specific T cells—includes IL-12, IL-23, TNF, IL-6, IL-1β, and type I IFN produced by DCs. Because cytokine secretion varies according to PRR activation, PAMPs and DAMPs are important factors, which skew the adaptive immune response.

T cell activation is dependent on the intensity and length of DC–T cell interactions, mediated by the immunologic synapse (IS), which is a region of spatially and temporally organized motifs of membrane proteins and cytosolic molecules formed at the interface between a T cell and a DC.\(^82\)

Stable long-lasting contacts of APCs with T cells are dependent on co-stimulatory molecules and integrin molecules ICAM-1 and ICAM-3. These integrins form a ring in the outer part of the IS, in a region adjacent to the corresponding area containing lymphocyte function associated antigen-1 (LFA-1).\(^83\) It is hypothesized that the stability of the IS confers temporal inhibition of DC apoptosis, which would enhance activation of clonal T cells in the lymph node.\(^82\)

CD4+ T helper (Th) cells play a vital role in the induction of humoral and cellular immunity, providing signals for the activation of B cells and CD8+ T cells. Controlled by the activated DCs, CD4+ T cells can differentiate into Th1 cells, producing IFN-γ and IL-2; Th2 cells, producing IL-4, IL-5, and IL-13; follicular helper T cells (Tfh), expressing inducible T cell co-stimulator and producing IL-21; Th17 cells, producing IL-17; or regulatory T cells (Tregs), producing TGF-β or IL-10 (see Chapter 13).

DCs promote a Th1 phenotype by acting on both NK cells and naïve CD4+ T cells. Collaboration of IFN-γ and IL-12 induces and sustains Th1 differentiation. The Th2 response is promoted by induction of OX40 ligand (OX40L) on DCs, which is mediated by release of thymic stromal lymphopoietin (TSLP) by epithelial cells, mast cells, and basophils.\(^94\) Absence or depletion of IL-12 production by DCs also skews the T cell response to Th2. In turn, IL-10 secretion by Th2 cells can negatively regulate DC function.\(^85\) DCs make a deficient response to TSLP stimulation in Crohn’s disease, an inflammatory bowel disease associated with Th1 and Th17 immunity.\(^96\)

Proinflammatory T cells expressing IL-17 were described in RA synovium as long ago as in 1998.\(^87,88\) Recently, Th17 cells expressing the RORγt transcription factor were described as a new subset of Th cells. Upon TLR stimulation, innate immune cells including DCs secrete IL-6 and IL-1β. The presence of IL-6 and TGF-β induces Th17 differentiation, subsequent production of IL-21, and expression of IL-23.\(^89,90\) IL-1β is also a critical amplifier of Th17 differentiation, especially in humans. IL-1β and IL-6 also promote the reprogramming of FoxP3+ regulatory T cells to Th17.\(^90\) This scenario is relevant to the development of autoimmune disease, as Tregs are selected based on intermediate affinity for self-peptides in the thymus.\(^91\) IL-21 amplifies Th17 differentiation; however, IL-23—produced by DCs—is needed for maintenance of Th17 cells.\(^89\)

Although Th1 cells were the major subset of T cells identified in RA synovium, identification of Th17 cells and the cytokines IL-17 and IL-23 within affected tissues and/or fluids has implicated both Th1 and Th17 cells in the pathogenesis of both RA and spondyloarthropathy.\(^88,92\) Synovial tissue DCs express IL-12p70 and IL-23p19, providing a mechanism by which Th1 and Th17 cell production is locally perpetuated.\(^93\)

pDCs, potent producers of type I IFN, also present viral and tumor antigens to initiate both CD4+ and CD8+ T cell responses, thereby providing an important link between innate and adaptive immunity.\(^94,95\) Increased production of IFN-α may activate CD8+ T cells, which are increased in kidneys in lupus nephritis patients.\(^96,97\) In addition, pDCs can induce Treg generation by induction of indoleamine-2,3-dioxygenase (IDO) (discussed later).

### B CELL ACTIVATION

B cells express both antigen-specific B cell receptors (BCRs) and various TLRs, thus allowing them to play a role in innate and adaptive immunity.

Although DCs present processed antigens to naïve T cells, B cells recognize antigen in its unprocessed native state. Two mechanisms of resistance to intracellular antigen degradation in DCs are described: (1) the internalization of antigen-containing immune complexes into nondegradable intracellular compartments through FcγRIIb, followed by receptor recycling to the cell surface and presentation to the BCR; and (2) accumulation of intact antigen in neutral endosomes by DC-SIGN (carbohydrate-containing antigen).
When DCs and B cells take up the same antigen, DCs can activate B cells in an antigen-specific CD4+ T cell–dependent manner. This results in B cell activation and isotype switching to IgG, IgA, and IgE, as well as memory B cell generation in response to T-dependent antigens.

DCs can also stimulate B cell proliferation by expression of BAFF and its closely related tumor necrosis family member APRIL, a proliferation-inducing ligand. Furthermore, inflammatory cytokines secreted by DCs can also affect B cell activation. IFN-α and IL-6, or ICAM-1, expressed by activated pDCs, regulate B cells to differentiate into plasma cells for T cell–independent antibody production. IFN-α can enhance antibody secretion in vivo and can induce cDCs to produce BAFF and APRIL; this triggers isotype switching independently of CD40 ligation.

**NK CELL ACTIVATION**

DCs can induce activation and/or proliferation of NK cells, resulting in increased NK cell cytolytic activity and/or IFN-γ production. Both cell-cell contact–dependent interactions and soluble cytokine signals are involved. NKG2D ligands are cell contact–dependent signals expressed by DCs in response to stressful stimuli. LFA-1/ICAM-1 interactions have also been implicated in NK cell activation by DCs. Type I IFN augments NK cell cytotoxicity; IL-12 induces IFN-γ secretion and cell proliferation; IL-15 determines cell differentiation and survival; and others such as IL-18 and IL-2 are mediators in DC-mediated NK cell activation.

Reciprocally, NK cell–mediated activation of DCs depends mainly on cytokines, principally TNF and IFN-γ, whereas cell contact–dependent interactions are required for DC elimination. NK cells can lyse imDCs through NKP30—a process referred to as DC editing. A TRAIL–mediated mechanism and CD94-NKG2A–mediated signals also induce DC lysis.

**DENDRITIC CELLS AND TOLERANCE**

Self-tolerance is the ability of the immune system to deal with invading microorganisms (pathogenic antigens) while remaining inert to the body’s constituents (self-antigens). A recent autoimmune model, in which cDCs, pDCs, and Langerhans cells were constitutively depleted from thymus and periphery, demonstrated the overall importance of DCs for self-tolerance.

**Central Tolerance**

Central (or thymic) tolerance defects are important and probably essential contributors to spontaneous autoimmune disease. T cells are selected in the thymus according to their affinity for self-MHC molecules loaded with endogenous self-antigens displayed by thymic cortical epithelial cells. Those T cells reactive to self-antigen above a threshold of affinity for self-antigen–MHC complexes expressed and presented by medullary APCs are then deleted by negative selection. These APCs include medullary epithelial cells (mTECs) and medullary DCs expressing the marker SIRPa, and they delete self-reactive T cells in the thymus in experimental settings.

In the medulla, mTECs express the transcription factor known as autoimmune regulator (AIRE), which controls the expression of peripheral tissue antigens (PTAs) such as insulin and salivary protein-1. In the absence of AIRE, glandular (salivary and lacrimal glands, liver, pancreas, thyroid) organ-specific autoimmunity develops. In humans, this monogenic syndrome is known as autoimmune polyglandular syndrome and candidiasis (APECED).

In mice, thymic medullary DCs include CD11b+CD8+ SIRPa+ DCs arising from proliferating intrathymic precursors, and extrathyrmically derived CD11b+CD8+ SIRPa+ DCs. Both are located in the cortico-medullary junction, the site of negative selection. Medullary DCs can present self-antigens that they have captured from peripheral tissues before their traffic into the thymic medulla, or after uptake of apoptotic mTECs, which themselves express PTAs. mTECs progress to apoptosis after terminal differentiation and expression of AIRE, delivering PTA to DCs for the induction of central tolerance to self-peptides in the thymus.

Although an affinity threshold applies for central deletion of self-reactive T cells, this threshold varies according to the susceptibility of thymocytes to death and the capacity of the T cell receptor and its signaling pathways to transmit an activation signal. Efficiency of self-antigen presentation also depends on the ability of thymic APCs to process and present self-antigen, and the density of MHC and co-stimulatory molecules expressed by the APCs.

Although selection of CD4+CD25 SP thymocytes occurs in the cortex, FoxP3+ thymocytes are located in the medullary regions, and DCs from the periphery can migrate to the medullary and play an important part in their development.

**Peripheral Tolerance**

Peripheral tolerance mechanisms provide a second level of control to restrain potentially autoreactive T cells, which migrate into the periphery from the thymus. This is necessary because of the stochastic nature of deletional central tolerance, which is based on affinity thresholds. Deletion is also used as a peripheral tolerance mechanism, where resting DCs derived from hematopoietic progenitor cells have been shown to delete autoreactive CD8+ T lymphocytes. This deletion of T cells autoreactive toward specific autoantigenic determinants is also known as recessive tolerance. In addition, “dominant” peripheral tolerance mechanisms involve suppression or regulation of the proliferation and effector functions of autoreactive T cells by Tregs. Although antigen-specific, regulatory mechanisms are able to suppress toward a wider range of determinants as the result of bystander or infectious tolerance mechanisms.

"Natural (n)” CD25+FoxP3+ Tregs are produced centrally, selected by recognition of self-antigen at intermediate affinity by cTECs, and potentially also by mTECs and thymic DCs. However, DCs also induce or promote the generation of peripheral Tregs, known as induced (i)Tregs. This occurs in several contexts. CD103+ DCs located in gut epithelium and draining mesenteric lymph nodes stimulate the peripheral conversion of FoxP3+ Tregs from naïve T cell precursors. These specialized lamina propria DCs express retinal dehydrogenase, which catalyzes the conversion of...
retinal to retinoic acid.\textsuperscript{16,120,123} In vitro, this conversion is recapitulated by stimulation of naïve T cells with DCs in the presence of mitogenic anti-CD3 and TGF-β.\textsuperscript{122} Another FoxP3+CD25+ iTreg population known as regulatory type I (Tr1) cells expresses high levels of IL-10 and IFN-γ and typically is induced by DCs generated in the presence of, and expressing high levels of, IL-10.\textsuperscript{123} In humans this small DC subset is marked by expression of CD16, CD14, immunoglobulin-like transcript (ILT4), and HLA-G.\textsuperscript{125} Finally, FoxP3+ Tregs may be induced in the periphery and may develop greater suppressive function as a result of T cell exposure to the enzyme indoleamine-2,3-dioxygenase (IDO). IDO is expressed upon activation of various APCs, including pDCs and cDCs, and depletes tryptophan from the environment by metabolizing it to kynurenine, thus starving T cells of an essential amino acid required for cell growth and viability.\textsuperscript{22,124} Although IDO probably plays little role in the maintenance of tolerance in the resting state, it is strongly induced by NFκB signaling in response to certain inflammatory signals, including TLR7 and TLR9 activation by single-stranded RNA and CpG, respectively, immunostimulatory sequences, type I and type II IFN, and “back-signaling” by the T cell activation--induced cell surface molecules CTLA4 and glucocorticoid-induced TNFR-related protein (GITR), which ligate CD80/CD86 and GITR-ligand respectively, expressed by DCs.\textsuperscript{125,126} GITR-ligand is also induced by glucocorticoids. Thus, IDO+ DCs induced by inflammatory signals and activated T cells promote the generation of FoxP3+ Tregs with timing that ensures regulated expansion of activated T cells and other innate inflammatory pathways. pDCs appear to induce IDO activity and FoxP3+ Tregs particularly effectively; this role may be even more important than that of immunogenic antigen presentation for this subset.\textsuperscript{122,127} They may use additional mechanisms, such as granzyme B, to regulate effector T cell function.\textsuperscript{128}

In chronic inflammatory settings, such as RA and juvenile rheumatoid arthritis (JRA), IDO+ APCs and FoxP3+ T cells are thus enriched, relative to healthy tissues.\textsuperscript{129,132} However, mechanisms such as inhibition of regulatory function by proinflammatory actions of TNF and IL-6 or expression of tryptophanyl-t-RNA-synthetase inhibit regulatory function or the response of effector cells to regulation.\textsuperscript{90,129,133}

The discovery that DCs generated ex vivo could be exposed to antigen and transferred to naïve recipients to prime immunity spawned many experiments in mice with subsequent translation to human clinical trials in which DCs were employed as a natural adjuvant for vaccination.\textsuperscript{134,135} The subsequent discovery that Tregs could be induced by adoptive transfer of antigen-exposed modified DCs opened possibilities of restoration of tolerance and suppression in mouse models of autoimmune disease and allografting, with recent translation to clinical trials in RA and type 1 diabetes.\textsuperscript{136} DCs with the capacity to regulate inflammation through iTregs typically present antigen with intermediate affinity to T cells (e.g., suboptimally activated DCs, which present small amounts of antigen, DCs lacking co-stimulatory molecules or expressing suppressive molecules such as IL-10, IL-4, or Fas-ligand).\textsuperscript{137,138} Antigen-exposed DC deficiency in the NFκB subunit RelB induced antigen-specific Tregs and suppressed disease after adoptive transfer into wild-type hosts with inflammatory arthritis.\textsuperscript{139,140} Similar effects occurred upon transfer of antigen-loaded DCs that had been transfected with RelB siRNA.\textsuperscript{141} These experiments identified NFκB RelB as a critical regulator of DC antigen-presenting function.\textsuperscript{142} The PI3K/mammalian target of rapamycin (mTOR) is a second major regulatory pathway regulating the interactions between DCs and regulatory and effector T cells.\textsuperscript{118} Thus DCs generated in the presence of rapamycin, an inhibitor of mTOR, induced antigen-specific Tregs and suppressed rejection after adoptive transfer into allograft recipients.\textsuperscript{143} Wnt/β-catenin is a third major regulatory pathway involved in DC function. β-Catenin was recently shown to be constitutively active in intestinal DCs and to regulate the balance between FoxP3+ Tregs and Th17 effector cells in the lamina propria through induction of retinal dehydrogenase.\textsuperscript{144} In vitro, disruption of E-cadherin signaling in DC cultures activated this pathway and their capacity to induce Tregs.\textsuperscript{145} Enhanced and prolonged ERK signaling has also been shown to be important for suppression of IL-12 and induction of IL-10 secretion by DCs, which promote tolerance.\textsuperscript{146} This includes TLR2-dependent (e.g., by bacterial teichoic acids) and TLR2-independent signals (e.g., by C5a and cigarette smoke--induced oxidative stress).\textsuperscript{147}

Many other approaches have been developed using immunomodulatory drugs for generation of DCs with similar capacity to induce Tregs and suppress disease after adoptive transfer. These include combinations of immunomodulators, which turn out to block NFκB and to activate IDO and ERK, such as vitamin D, glucocorticoids, and lipoteichoic acid,\textsuperscript{148} or molecules with the capacity to induce partial activation by DCs, thus preventing further induction of immunogenic molecules, such as IL-12, resulting in a “semi-mature” or “tolerant” state. The semi-mature state can be induced by chronic exposure to TNF, IL-6, or low-dose endotoxin.\textsuperscript{149,150} It has also been proposed that exposure to chronically increased levels of IL-6—as occurs in RA or chronic infection, or in the context of various tumors—can promote a tolerant DC phenotype in vivo.\textsuperscript{151}

Appreciation of the critical pathways regulating DC function has led to a number of successful approaches in mouse models to target DC to promote tolerance. Inflammatory arthritis was suppressed in mice delivered arthritogenic antigen packaged into liposome nanoparticles with an NFκB inhibitor targeting phagocytic APCs in situ.\textsuperscript{152} CD40-antisense oligonucleotides packed into liposomes similarly downregulated CD40 expression after targeting phagocytic APCs and blocked collagen-induced arthritis (CIA).\textsuperscript{153} Delivery of CTLA4-Ig (abatacept) to mice with CIA modified DCs and induced FoxP3+ Tregs.\textsuperscript{154} More specific targeting of CD205+ DCs with proteolipid protein autoantigen conjugated to anti–DEC-205 suppressed experimental allergic encephalomyelitis in mice through induction of Tregs.\textsuperscript{155} More indirectly, regulatory DCs expressing high levels of IL-10 and TGF-β were induced by administration of probiotics.\textsuperscript{156}

**AUTOIMMUNITY**

In autoimmune diseases, when tolerance against self-determinants is impaired, activated autoreactive T and B
cells and inflammatory innate immune and parenchymal cells participate in the process of tissue damage. DCs play an essential role in T cell priming in autoimmune disease, following translocation of antigens (including those modified [e.g., citrullinated] by the inflammatory process) from the periphery to secondary lymphoid organs and their presentation to naïve T cells, but also locally at the inflammatory site. Here they act as both local APCs and effector cells, producing proinflammatory molecules and regulating other effector cells. In different autoimmune conditions, cDCs and pDCs are enriched in inflamed tissues, particularly in a perivascular distribution. Moreover, DCs infiltrate tissues at very early stages of disease, where they contribute to the recruitment and organization of other immune cells. For example, in mice, DC infiltration is an early feature of islet cell autoimmunity in diabetes mellitus and contributes to local lymphoid tissue formation in the pancreas. Organization of tissue into lymph node–like structures, including lymphoid follicles, is a common feature of tissues affected by autoimmune inflammation, which contain follicular DCs (FDCs) and nurse-like fibroblast-like synoviocytes, supporting chronic production of autoantibodies.

Here we discuss prototypic autoimmune rheumatic diseases and the role that DCs play in their initiation and progression.

**Systemic Lupus Erythematosus**

In SLE, an increased load of apoptotic material has been related to induction of the disease and its severity. Complement receptor 3 (CR3), a heterodimer of CD11b and CD18, is an apoptotic receptor that takes up apoptotic cells coated with C3bi. Recently, a variant of ITGAM (CD11b) has been discovered as a risk factor for SLE. The IFN-α pathway is strongly linked to SLE pathogenesis, both biologically and genetically. Several genes for factors upstream and downstream of IFN and TNF production, such as IRF5, STAT4, TNAIP3, and TYK2, have been associated with SLE susceptibility. Additionally, TREX1, a gene encoding protein involved in degradation of nicked DNA, has been associated with SLE susceptibility. It is proposed that reduced apoptotic clearance and thus availability of autoantigens in an immunogenic format for antigen presentation and uncontrolled IFN production by DCs can mediate SLE pathogenesis and maintenance.

Under normal conditions, uptake of apoptotic material in the absence of inflammatory cytokines promotes DC tolerance through stimulation of TGF-β and induction of Tregs. Dysregulated clearance of apoptotic cells leads to an accumulation of blebs with modified autoantigens. Apoptotic blebs containing endogenous danger signals or alarmins such as HMGB1 or antimicrobial peptides complexed with nucleosomes can activate DCs. Also, in already primed individuals, DNA-containing chromatin bound to autoantibodies can activate pDCs through TLR9. However, cDC activation by TLR9–independent activation and TLR7–dependent mechanisms has also been described because cDCs do not express TLR9. Moreover, uptake of autoantigens into lysosomal-associated membrane protein 1 (LAMP1) endocytic compartments of cDCs may be critical for effective antigen presentation. Furthermore, an increase in the number of circulating Th17 cells and a decrease in the number of Treg cells have also been recorded; these were associated with disease flares of SLE. In addition, hyperactivation of IL-17 and IL-23 has been observed in SLE patients.

Although T cells initially primed by DCs may be specific for modified autoantigens, epitope spreading accounts for activation of B cell responses against modified and unmodified autoantigens. A model, which proposes a pivotal role for DCs in the initiation and progression of SLE, indicates that antibodies opsonize autoantigens and activate pDCs, which will amplify autoimmune responses primed by cDCs.

**Rheumatoid Arthritis**

DCs are thought to contribute to ongoing inflammation through presentation of autoantigens, as suggested by animal models of RA in which DCs have been depleted, or through proinflammatory (including capacity for cartilage destruction) and regulatory effector functions. For instance, joint DCs and monocyte-derived DCs can present human cartilage glycoprotein 39 (HCgp39) and epitopes from synovial fluid (SF) to antigen-specific T cells, respectively. Further evidence for the contributing role of DCs to disease initiation and perpetuation in RA and the spondyloarthropathies derives from the substantial accumulation of immature and mature DC subsets, including cDCs and pDCs perivasculary in inflamed synovium, in close association with T and B cell follicles. This accumulation of DCs contains the most significant numbers of mDCs and pDCs identified in any human inflammatory site. DCs expressing markers of both differentiated and undifferentiated cells have been identified in RA synovial tissue. Evidence of prior activation in vivo includes upregulation of MHC, co-stimulatory molecules such as CD86, and RElB; expression of receptor activator of nuclear factor κB (RANK) and its ligand (RANKL); and heightened production of proinflammatory cytokines ex vivo (e.g., IL-1, IL-6, TNF) in response to stimulation with immune complexes or TLR agonists. Migratory DCs may undergo maturation in response to locally produced cytokines, or to endogenous factors released during inflammation (e.g., HMGB1, HSPs). These DCs may have migrated into the joints in response to locally produced cytokines and chemokines (CCL19 to CCL21) or may differentiate locally from myeloid progenitors in response to growth factors contained within SF. Expression of nuclear RElB and CCR7 is associated with cells expressing CCL19 and CCL21. Ectopic expression of CCL19 is sufficient for formation of lymphoid tissue similar to that seen in rheumatoid synovial tissue. In contrast, immature DCs in the synovial lining and in sublining layers are characterized by CCR6 expression and are associated with CCL20–expressing cells. The number of CD304+ pDCs expressing type I IFN and IL-18 was found to exceed that of the CD1a+ subset of cDCs, expressing IL-12 and IL-23 in RA synovium, particularly in patients positive for RF and anticitrullinated protein antibodies (ACPAs). RA synovial pDCs also displayed immunoreactivity for IDO. In RA and JRA peripheral blood, cDCs were found to be reduced relative to healthy
peripheral blood, and this reduction correlated with disease severity, supporting the idea that migration to synovium is increased by inflammatory chemotactic activity. On the other hand, the number of peripheral blood pDCs in RA correlated inversely with the development of antinuclear antibodies and increased serum IFN-α following TNF inhibition with infliximab. In the peripheral blood of children with systemic JRA, cDCs were decreased and peripheral blood (PB) monocytes were increased during flares.

Increased numbers of cDCs and pDCs have been shown in SF. cDCs are threefold to fourfold enriched in fresh synovial fluid compared with the blood or with osteoarthritis SF and are more differentiated than blood DC precursors. SF DCs and blood DCs stimulate resting allogeneic T cells with equivalent efficacy, and more efficiently than SF monocytes. However, SF DCs are markedly more efficient stimulators of autologous T cells in the absence of exogenous antigen, suggesting that their MHC molecules are loaded with relevant RA autoantigens.

RA was initially believed to be exclusively Th1 driven. However, identification of Th17 cells and its cytokine IL-17 within affected tissues and/or fluids now implicates both Th1 and Th17 cells in its pathogenesis. DCs are required for the differentiation of both subsets from naive precursors. Th1 cells, which produce IFN-γ, activate macrophages to produce proinflammatory cytokines (IL-1, IL-6, and TNF), induce immunoglobulin class switching to complement-fixing antibodies. IL-17 recruits leukocytes and stimulates IL-1 and TNF production by human macrophages, which altogether stimulate activation of synovial fibroblasts.

Th17 cells mediate bone destruction in RA by stimulating osteoclastogenesis through RANK-RANKL interactions. Subsets of synovial DCs express IL-12 or IL-23—cytokines essential for full differentiation of Th1 and Th17 cells, respectively—providing a mechanism by which their production is locally facilitated or perpetuated.

In RA synovium, ectopic lymphoid organs resembling germinal centers (GCs) support autoantibody production because they express activation-induced cytidine deaminase (AID) and are surrounded by ACPA-producing plasma cells. In RA, ACPA and rheumatoid factor antibodies are presumed to bind to Fc receptors on macrophages and DCs, thereby inducing their activation and production of proinflammatory cytokines. Indeed, synovial DCs from active RA patients have markedly upregulated FcγRII. pDCs have also been found in rheumatoid synovium and through production of IFN-α may enhance autoantibody production. Indeed a substantial number of RA patients express an IFN-α signature associated with ACPA production.

Altogether, human data strongly suggest multiple roles for DCs in rheumatoid synovium, in local inflammation, joint destruction, and immune regulation. These data complement murine studies of inflammatory arthritis, which establish the important role of DCs in priming, inflammatory responses, and regulation of arthritic autoimmune responses.

Sjögren’s Syndrome

Sjögren’s syndrome (SS) is characterized by infiltration of mononuclear cells into exocrine, predominantly salivary and lacrimal, glands. Infiltrating cells are mainly CD4+ T cells but also may be CD8+ T cells, B cells, and some plasma cells. Sjögren’s syndrome–related autoantibodies include SSA/Ro, SSB/La, antinuclear antibodies, and rheumatoid factor.

Germinal center formation with FDC networks has been shown in the salivary glands of SS patients, possibly attracting lymphocytes to the site. FDCs have also been implicated to function as accessory cells for B cell function.

Mouse models of SS showed that DC infiltration precedes lymphocyte infiltration into the submandibular gland, suggesting the initiating role of DCs in disease development. Increased numbers of mature DCs were detected in a mouse model and in human SS patients. pDCs were detected in the salivary glands of SS patients but not of healthy controls, whereas decreased levels of pDCs in the blood of SS patients indicate their recruitment into salivary glands. Given the role of pDCs and the increased levels of antinuclear antibodies and of type I interferon in the salivary glands, several studies have investigated the role of Epstein-Barr, hepatitis C, human T-lymphotropic virus type I (HTLV-1), and coxsackievirus in patients with SS, but no evidence has been found to support their direct involvement.

Seronegative Spondyloarthropathies (SpA) and Psoriasis

Bacteria are thought to be involved in the pathogenesis of psoriasis and psoriatic arthritis. Because the innate immune system is the first line of defense against bacteria, innate responses have been studied in these patients. PB DCs in psoriatic arthritis (PsA) patients and keratinocytes and dermal DCs of psoriatic skin lesions express high levels of TLR2. pDC numbers are increased and the type I IFN signaling pathway is activated in psoriatic skin. Furthermore, blocking of pDC or IFN signaling in skin xenograft models prevented the development of psoriasis.

Keratinocytes play important roles in secreting antimicrobial peptides and DC chemoattractants. Antimicrobial peptide LL37 has been shown to bind self-DNA and to trigger type I IFN in psoriasis.

Dermal DCs are increased in psoriatic skin and induce proliferation of autoreactive T cells, inducing a Th1 response. A subset of dermal DC, known as Tip-DC, produces inducible nitric oxide synthase (iNOS) and TNF, promoting a proinflammatory environment.

In psoriatic lesions, a vicious cycle develops, in which activated T cells produce TNF and IFN-γ, which activate keratinocytes to produce a wide variety of cytokines, resulting in DC activation. Chemokines attract lymphocytes into the site. However, the trigger is unknown.

The importance of HLA-B27 in spondyloarthropathies is well defined, but recent genetic studies demonstrate the involvement of innate immune system defects in the development of not only ankylosing spondylitis (AS) but also psoriasis and Crohn’s disease. These include IL23R, IL12B, STAT3, and CARD9. These findings support research indicating the importance of IL-17 in disease pathogenesis in SpA, and suggesting possible involvement
of fungal recognition in DCs,\textsuperscript{237} inducing a Th17 response in T cells.

In HLA-B27 transgenic rats, a defective stimulatory capacity of DCs has been detected.\textsuperscript{238,239} HLA-B27 in this model also induces the unfolded protein response,\textsuperscript{240} leading to reduction of MHC class I\textsuperscript{241} and IL-23 production,\textsuperscript{242} which seems to play an important role in SpA.

Similar to other seronegative arthritides, pDCs are increased in synovial fluid but not in PBs in SpA patients.\textsuperscript{183} These pDCs were more immature than cDCs when co-
stimulatory markers, including CD40, CD83, and DC-LAMP , were analyzed. Thus defective DC APC function may be an important factor in the initiation or perpetuation of SpA.

Selected References


154. Kaji R, Kyoshima-Shibata J, Nagaoka M, et al: Bacterial teichoic acids reverse predominant IL-12 production induced by certain lactobacillus strains into predominant IL-10 production via TLR2-


Mononuclear phagocytes are widely distributed, biosynthetically active cells derived from hematopoietic precursors that circulate as monocytes and enter tissues constitutively and in response to inflammatory stimuli. In connective tissue and bone, they play a major role in homeostasis, growth, and remodeling as mature macrophages and osteoclasts (Figure 10-1). Myeloid dendritic cells (DCs) represent a distinct form of differentiation, specialized to maintain immune tolerance or to induce humoral and cellular immunity through B lymphocytes and T lymphocytes. Through their recognition, antigen-presenting, regulatory, and effector mechanisms, mononuclear phagocytes contribute to a range of inflammatory, infectious, autoimmune, metabolic, and degenerative rheumatic diseases, providing targets for therapeutic intervention.

Growth of knowledge in mononuclear phagocyte biology has developed in close association with understanding of pathogenesis and treatment of chronic arthritis. Relevant examples include the development of steroidal and nonsteroidal anti-inflammatory agents and of anti–tumor necrosis factor (TNF) monoclonal antibodies. Genetic lesions in macrophage colony-stimulating factor (M-CSF) cause osteoclast deficiency and osteopetrosis in mouse models, whereas human mutations in cytosolic nucleotide oligomerization domain (NOD)-like receptors (NLRs) result in interleukin (IL)-1β overproduction and hyperinflammatory syndromes, often associated with persistent joint disease. Immune complex deposition and complement activation combine with Toll-like receptor (TLR) recognition to induce effector pathways of tissue destruction and repair.

This chapter reviews general aspects of mononuclear phagocyte differentiation, recruitment, and activation, integrating the properties of different cellular subtypes. Studies on the biology of macrophages, DCs, and osteoclasts have diverged to the extent that common features have been overlooked in the understandable search for specificity. This review is an attempt, in part, to reintegrate these differentiated sublineages. General features especially relevant to physiologic and pathologic consequences of their presence in bone, joints, and connective tissues are discussed, and gaps in our knowledge are pointed out. The emphasis is on studies in humans, where available, with reference to murine models where applicable. The subject area encompasses innate and acquired immunity, autoimmunity, and osteoimmunology, a reflection of local specialization of mononuclear phagocyte and lymphocyte biology. Relevant topics discussed elsewhere in this textbook include innate immunity (see Chapter 18), cytokines and chemokines (see Chapter 26), osteoclast functions (see Chapter 4), and anti–TNF therapy (see Chapter 63). For further details in connection with macrophage and DC biology, see reviews cited in this chapter.

OVERVIEW

Circulating monocytes give rise to tissue macrophages, which display considerable phenotypic microheterogeneity in different organs. Similarly, and with some overlap in properties, myeloid DCs are present as heterogeneous sentinel cells at mucosal and cutaneous surfaces undergoing a complex maturation process as they migrate to lymphoid organs after capture of antigens. Plasma-lymphoid DCs may represent a distinct sublineage, specialized to produce high levels of type I interferon in response to viral stimuli. Mononuclear precursors in the blood give rise to multinucleated osteoclasts, specialized to resorb bone. Circulating monocytes are themselves heterogeneous, using distinct chemokine and adhesion receptors to give rise to different tissue mononuclear phagocytes. These cells differ in life span depending on the recruitment stimuli and on local factors in their environment, and they express diverse plasma membrane receptors, making it possible for them to interact with many different cell types and with microbial and modified host components. Phagocytosis is a hallmark of the ability of macrophages and DCs to engulf particulates, including foreign materials, bacteria, and dying host cells generated by apoptosis or necrosis.

KEY POINTS

- Macrophages play an important role in tissue homeostasis and in cellular and humoral innate and adaptive immunity.
- Macrophages, dendritic cells, and osteoclasts are derived from blood monocytes.
- Tissue macrophages display marked phenotypic heterogeneity, depending on their local microenvironment.
- Distinct subsets of monocytes are recruited to inflammatory sites in tissues.
- Microbes and cytokines (e.g., interferon-γ, interleukin [IL]-4, IL-10) selectively regulate macrophage activation.
- Macrophages express a range of opsonic and nonopsonic receptors to recognize and take up foreign and altered host components.
- Toll-like membrane and nucleotide oligomerization domain (NOD)-like cytosolic receptors induce alterations in macrophage gene expression and secretion of inflammatory mediators.
- Macrophages and their products contribute to tissue damage and repair and to chronic inflammation and autoimmunity.
These particulates are recognized by nonopsonic receptors, including scavenger and lectin-like receptors, or after opsonization with antibodies or complement or both that enhance uptake via Fc and complement receptors. In addition, other humoral proteins, such as Pentraxins, interact with their target ligands, bridging them to less well-defined macrophage receptors, which regulate early cellular responses during innate immunity. TLRs play an important role in sensing the nature of the captured cargo, often interacting with the extensive repertoire of non-TLR receptors (NTRs). An extended family of NOD-like receptors (NLRs) senses various cytosolic ligands, resulting in a complex assembly of proteins (inflammasomes), caspase activation, and release of IL-1β. Uptake of particulate and soluble antigen, directly or by “cross-priming,” induces DC maturation, processing, and association with major histocompatibility complex (MHC) II molecules, as well as presentation of peptides to naive CD4 T lymphocytes. Endogenously generated or foreign peptides generated during biosynthesis of virus glycoproteins associate with MHC I molecules for surface recognition by cytotoxic T cells (mainly CD8). DCs and possibly macrophages regulate T cell activation or tolerance through additional co-stimulatory surface antigens and cytokines, depending in part on concomitant TLR stimulation. Activated T lymphocytes and their products, such as interferon-γ, IL-4/IL-13, and IL-10, regulate the effector functions of mononuclear phagocytes.

Digestion of macromolecules occurs in the vacuolar compartment involving dynamic membrane traffic and interactions with the cytoskeleton. Complex intracellular pathways transduce surface and vacuolar-derived stimuli to initiate formation of intracellular protein signaling complexes. Transcription factors translocate to the nucleus and form activation or inhibitory chromatin-binding complexes that regulate gene expression and biosynthesis of secretory products, such as TNF. Other low-molecular-weight metabolites are generated by nontranscriptional mechanisms, yielding inflammatory mediators and antimicrobial activities. Osteoclasts are able to secrete hydrochloric acid (HCl) and potent proteolytic enzymes into sealed-off localized areas of bone, to which they adhere tightly through specialized podosomes and actin rings.

The secretory activities of mononuclear phagocytes influence a range of cellular and extracellular targets to maintain tissue homeostasis but also are responsible for tissue destruction. Their trophic actions, through cellular contact with other stromal cells and extracellular matrix and secretion, regulate tissue catabolism, cell growth, angiogenesis, and repair. In addition to local effects, macrophages contribute to systemic integration of proinflammatory and anti-inflammatory effects, acting on the central nervous system, endocrine organs, liver, and energy stores.

Overall, the activation phenotype of mononuclear phagocytes is modulated by extrinsic factors (e.g., cytokines, hormones), by balance of surface receptors with activating/inhibitory cytoplasmic motifs, by cytosolic regulators such as suppressors of cytokine synthesis (SOCS) proteins, by phosphorylation/dephosphorylation of signaling molecules, and by assembly of transcription factor complexes on chromatin. Microarray analysis of macrophages has made it possible to distinguish characteristic signatures of gene expression after innate activation via TLRs, deactivation via glucocorticoids, or modulation via cytokines. As a result, the phenotype of tissue macrophages is markedly heterogeneous, making for complexity of function in different sites in health and disease, but also providing opportunities for novel state-specific or tissue-specific targeting by drugs.

**LIFE HISTORY AND HETEROCENEGITY (MACROPHAGES, DENDRITIC CELLS, AND OSTEOCLASTS)**

In an adult, all three sublineages of mononuclear phagocytes originate from CD34+ committed progenitor cells in the bone marrow (Figure 10.2). These diverge from lymphoid cells and subsequently from polymorphonuclear leukocytes, although many genes are still expressed but not translated in both types of phagocytic cell. In vitro, bone marrow and blood monocytes can be stimulated by growth factors to generate macrophages (M-CSF or granulocyte-macrophage colony-stimulating factor [GM-CSF]), DCs (GM-CSF, with or without IL-4), or osteoclasts (M-CSF and RANK ligand). Some of these growth factors also are essential in vivo (e.g., osteopetrotic, M-CSF–deficient mice lack many but not all populations of tissue macrophages and osteoclasts). Monocytopoiesis is less well understood in humans and may depend on M-CSF and GM-CSF. Trophic interactions between hematopoietic precursors and stromal cells in the bone marrow (mesenchymal and hematopoietic in origin) are mediated by cell contact, via surface receptors, and by soluble factors (e.g., c-kit ligand and IL-1). Transcription factors that are essential for monocyte/macrophage production and related cells include Pu-1,
other Ets family members, Maf, and Mi, implicated in microphthalmia.

Although bone marrow precursors proliferate vigorously as they differentiate in the presence of M-CSF or GM-CSF, monocytes become refractory to these growth stimuli. Local macrophage proliferation can be induced by IL-4 during Th2 inflammation. Restriction of DNA synthesis is associated with chromatin condensation, whereas RNA and protein synthesis persists and can be modulated by a variety of stimuli, as is discussed further subsequently.

Tissue macrophages, DCs, and osteoclasts all derive from circulating monocytes, although these already may display heterogeneity (see Figure 10-2). Recent studies in transgenic mouse models and in humans have described progenitor cells and precursor-product relationships in monocyte-macrophages, as well as DC differentiation. Cells are recruited constitutively to peripheral sites in the steady state. A subset of patrolling monocytes may not leave the vasculature in the steady state. Additional monocytes can be recruited to local sites in response to infectious, inflammatory, and metabolic stimuli. Such “elicited” cells display distinct properties from the “resident” cells, which, in the case of macrophages and DCs, also display marked diversity, depending on their location. A considerable body of evidence indicates that different subsets of macrophages and DCs originate from distinct monocyte populations in peripheral blood. Apart from marker antigens (e.g., CD14, the receptor for lipopolysaccharide-binding proteins; Gr-1, a mouse Ly-6 antigen expressed by polymorphonuclear neutrophils and by some monocytes), levels of chemokine receptors for fractalkine (CX3CR1) and CCR2 (MCP1) seem to distinguish monocyte subsets that give rise to inflammatory and resident macrophages (see Figure 10-2). Resident macrophages and immature DCs are present in many lymphohematopoietic and nonlymphoid organs; selected markers and properties are listed in Table 10-1. The CD68 antigen, a late endosomal mucin-like glycoprotein related to the lysosome-associated membrane protein (LAMP) family, is the most broadly expressed marker for all mononuclear phagocytes, although its function is still obscure.

Osteoclasts and resident macrophages are found on the surface of bone and display distinct phenotypes. Although antigenic markers for human and mouse mononuclear phagocytes are useful for phenotypic analysis, no single marker is definitive in their classification, and it is impossible, in most tissues, to distinguish locally activated resident cells from “elicited,” newly recruited, and locally activated monocytes. This problem is due in part to rapid modulation of antigen expression in blood monocytes when they enter a particular tissue microenvironment. Species differences in anatomy and in marker expression are a further confounding factor.

Apart from the expression of chemokine receptors, adhesion molecules play a role in selective monocyte “homing,” but their differential expression by macrophages, DCs, and osteoclasts is less well defined than for lymphocyte subpopulations. These include various heterodimeric integrins implicated in adhesion to endothelium, to extracellular matrix, and to bone. There is considerable scope for

---

**Figure 10-2** Differenniation and distribution of mononuclear phagocytes. Distinct subpopulations of circulating monocytes are thought to give rise to resident tissue macrophages (MØ), dendritic cells (DCs), and osteoclasts compared with cells recruited by an inflammatory stimulus. Further phenotypic heterogeneity arises from microenvironmental stimuli, such as cytokines and microbial products. For additional details of morphologic and other properties of cells in different tissues, see Gordon and Gordon and Hughes.
characterization of additional receptors and markers; one example is the EGF-TM7 family of leukocyte receptors illustrated in Figure 10-3.\textsuperscript{10} Although the F4/80 antigen has been extremely useful as a differentiation marker in the mouse, the human counterpart (EMR1) has not been of comparable value, showing a predilection for eosinophils. EMR2, a related human myeloid cell receptor, is a useful marker, however, expressed by many tissue mononuclear phagocytes; it binds chondroitin sulfate proteoglycans broadly present in connective tissue and has been implicated in leukocyte adhesion, migration, and activation.\textsuperscript{11}

In joints, a resident synovial macrophage population is present, and recruited monocytes and macrophages are prominent in inflammatory, autoimmune, and infectious diseases, together with DCs and other myeloid and lymphoid cells. These resident and recruited mononuclear phagocytes have not yet been attributed to particular monocyte subsets. These may not be the same monocyte subpopulations that give rise to resident and recruited mononuclear phagocytes in other tissues and in response to different pathologies.

**MOBILIZATION OF MONONUCLEAR PHAGOCYTES**

Similar to other white blood cells, mononuclear phagocytes are distributed in intravascular and extravascular compartments, sharing common mechanisms of mobilization but also displaying distinct features among themselves and compared with other cell types. Exit from the bone marrow is controlled constitutively and on demand, but this process is not well understood, except for the role of chemokine receptors, such as CCR2 (ligands MCP1 through MCP4).\textsuperscript{32} What determines differentiation into the mature macrophages that form part of the stromal microenvironment in bone marrow is unknown; one possibility is re-entry of circulating monocytes from blood to bone, as for osteoclasts. Apart from the circulating pool of monocytes, many mature macrophages adopt a sinusoidal distribution in liver (Kupffer cells) and in selected lymphoid and endocrine organs. These cells are distinct from, but share endocytic properties with, sinusoidal endothelial cells. The constitutive exit through vascular endothelium to become tissue macrophages and DCs is not understood, whereas induced mobilization is well characterized, sharing many features with polymorphonuclear neutrophils.

Figure 10-4 summarizes stages and molecules implicated in monocyte egress. Although most monocytes exit the microvasculature by diapedesis between endothelial cells, evidence of an alternative transcytotic mechanism is

---

**Table 10-1  Selected Properties of Mononuclear Phagocytes and Related Cells**

<table>
<thead>
<tr>
<th>Monocytes/Macrophages: Antigen Marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>F4/80 (mouse)</td>
</tr>
<tr>
<td>EMR2-EGF module containing receptor (human)</td>
</tr>
<tr>
<td>CD68</td>
</tr>
<tr>
<td>Complement receptor 3 (CR3) (CD11b)</td>
</tr>
<tr>
<td>Sialoadhesin (SIGLEC-1)</td>
</tr>
<tr>
<td>Scavenger receptors:</td>
</tr>
<tr>
<td>Scavenger receptor A (SR-A)</td>
</tr>
<tr>
<td>Macrophage receptor collagenous domain (MARCO)</td>
</tr>
<tr>
<td>Mannose receptor</td>
</tr>
<tr>
<td>Macrophage colony-stimulating factor receptor (M-CSFR)</td>
</tr>
<tr>
<td>Other properties:</td>
</tr>
<tr>
<td>Oposonic phagocytosis</td>
</tr>
<tr>
<td>Lysozyme secretion</td>
</tr>
<tr>
<td>Abundant acid hydrolases</td>
</tr>
</tbody>
</table>

**Myeloid Dendritic Cells: Antigens**

| Major histocompatibility complex class II |
| Co-stimulatory molecules                 |
| CD11c                                  |
| CD8α                                  |
| DEC205                                 |
| Dendritic cell–specific ICAM-grabbing non-integrin (DC-SIGN) |
| Dendritic cell–lysosome-associated membrane protein (DC-LAMP) |
| Other properties                        |
| Activation of naive CD4 T lymphocytes   |

**Plasmacytoid Dendritic Cells: Antigens**

| CD123                                 |
| B220                                  |
| Lectin-like receptors (SIGLEC H)      |
| Other properties                      |
| In vitro growth by fli-3 ligand       |

**Osteoclasts**

| CD68                                  |
| Tartrate-resistant acid phosphatase (TRAP) |
| Calcitonin receptor                   |
| αvβ3                                  |
| Vacuolar H⁺-ATPase                     |
| Proteinase K                          |
| Resorption of living bone             |

Note: Marker expression varies, depending on cell localization, maturation, and activation. Some markers also are present on other myeloid cells (e.g., polymorphonuclear neutrophils) and on selected endothelial cells. Structures and functions of receptor antigens are described elsewhere in this chapter.

---

---

---

---
Mononuclear phagocytes use integrins and receptors for extracellular matrix such as CD44, regulating the dynamics of their adhesion and migration. Osteoclast adhesion to bone depends on $\alpha_\nu \beta_3$ and possibly other adhesion molecules implicated in podosome attachment. Plasma membrane receptors such as CD97 and EMR2, members of the EGF-TM7 family, and TREM-1, an immunoglobulin superfamily molecule with an immunotyrosine activation motif (ITAM)-based activation motif (see later), regulate myeloid cell effector functions and adhesion and migration within extravascular tissue compartments.

Monocytes recruited to tissues by poorly degradable foreign materials or by selected microbial pathogens or parasites form granulomas, organized structures rich in macrophages, incorporating other myeloid and lymphoid cells and fibroblasts and extracellular matrix. Pathogen-induced granuloma formation depends on adhesion molecules such as CR3, a $\beta_2$ integrin, TNF, and chemokine receptors such as CCR4; granuloma macrophages express abundant secretory products such as lysozyme and proinflammatory cytokines. Characteristic morphologic evidence of macrophage differentiation in granulomas includes epithelioid cell and multinucleated giant cell formation, such as with Mycobacterium tuberculosis infection. Giant cells arise by monocyte recruitment and macrophage fusion, rather than by impaired cytokinesis. Cytokines such as IL-4 and IL-13, acting through a common receptor chain, and GM-CSF promote macrophage fusion. Foreign surfaces including biomaterials also can play a role in granuloma formation, and chemokine receptors have been implicated in the induction of foreign body–induced giant cells.

The surface molecules involved in these examples of macrophage homokaryon formation and the functional significance of induced fusion are still poorly understood. By contrast, multinucleation in osteoclasts is a physiologic process, dependent on M-CSF and RANK ligand, and several plasma membrane (CD44, CD9, TREM-2, DC-STAMP) and intracellular (c-src, c-fos) molecules have been implicated in this differentiation process. Multinucleation is thought to favor efficient localized resorption of bone. Osteoclasts become polarized for secretion and display highly active plasma membrane ruffling. Interactions with...
Figure 10-5  Positioning of dendritic cells (DCs) within lymphoid tissues. Blood-derived monocytic precursors enter skin and mucosal peripheral tissues and differentiate into Langerhans cells within epithelia and into other immature DCs. When established, Langerhans cells turn over independently from bone marrow–derived cells (e.g., after ultraviolet irradiation). On exposure to stimuli (e.g., foreign antigens, local infections) and constitutively (after uptake of apoptotic cells, such as in the gastrointestinal tract), DCs undergo maturation, upregulate CCR7, and enter afferent lymphatics and secondary lymphoid tissues, where they interact with CD4 T lymphocytes. CD4 T lymphocytes become activated, interacting with B cells or CD8 T lymphocytes, and re-enter blood, homing to peripheral sites. DCs also are able to interact with innate lymphocytes and natural killer cells. Follicular DCs in B cell areas have a distinct, poorly defined bone marrow origin, express novel antigenic markers, and are able to capture immune complexes through complement activation. Plasmacytoid DCs have a distinct interfollicular location and express markers of a myeloid and a lymphoid nature. (Courtesy R. Steinman.)

Figure 10-6  Dendritic cell (DC) maturation. In the steady state, immature DCs are actively endocytic, capturing antigens efficiently through pattern recognition receptors. In response to a range of exogenous and endogenous stimuli, they translocate major histocompatibility complex class II to the surface, express a range of co-stimulatory molecules and secretory products, and become efficient antigen-presenting cells. As a result, DCs control adaptive immunity or induce tolerance. TNF, tumor necrosis factor. (Courtesy R. Steinman.)

| Table 10-2  Maturation Markers for Human Monocyte–Derived Dendritic Cells |
|-------------|-----------|-----------|
| **Enhancer/Co-stimulatory Molecules** | **Immature** | **Mature** |
| CD80 (B7-1) | Low | High |
| CD86 (B7-2) | Low | High |
| CD83 | De novo | Moderate/low |
| EMR2 (CD312) | High | Moderate/low |
| **Antigen Uptake** | | |
| FcγRII | High | Low |
| **Signaling** | | |
| CD40 | Low | High |
| CXC4 | Low | High |
| CCR5 | High | Low |
| CCR6 | High | Low |
| CCR7 | Low | High |
| **Antigen Presentation** | | |
| HLA-DR (major histocompatibility complex class II) | Moderate | High |
| HLA-DQ (major histocompatibility complex class II) | Moderate | High |
| CD1a | High | Low |
| **Others** | | |
| DC-SIGN (CD209) | High | Low |
| CD14 | High/moderate | Negative |
| CD123 (IL-3R) | Low | High |

Note: In addition to the listed antigens and antigen-presenting cell functions, dendritic cell production of, and response to, other growth factors, chemokines and cytokines, microbial products, and immune complexes varies during maturation and activation in vitro. Validation in vivo is incomplete.
osteoblasts; local cytokines, such as osteoprotegerin; and circulating hormones, including calcitonin, parathormone, and vitamin D metabolites regulate their gene expression and function, as is discussed in Chapter 4. Vitamin D receptors also modulate the function of macrophages and DCs.

The turnover of different mononuclear phagocytes varies, depending on their activation status and tissue localization.42 Resident macrophages can live for weeks or months, whereas inflammation reduces survival to hours or days. DCs are relatively short-lived cells; osteoclast turnover in vivo has not been studied in detail. The role of apoptosis in mononuclear phagocyte turnover is poorly characterized, in contrast to that in polymorphonuclear neutrophils, but cell survival is regulated by growth factors such as M-CSF and by interactions with neighboring cells and pathogens.

**RECOGNITION**

In the past decade, emphasis on the problem of immune recognition has shifted to a considerable extent from somatically rearranged receptors for peptide, on lymphocytes, to germline-encoded receptors on myeloid antigen-presenting cells (APCs).43-45 Previous studies on myeloid cells focused on well-known opsonic receptors for antibody (Fc receptors) and complement (complement receptor). The concept of direct pattern recognition receptors for conserved structures on microbes provided an impetus to studies on innate immunity, reinforced by the discovery of TLRs. More recently, investigation of numerous NTRs, including a range of lectin-like recognition molecules and of scavenger receptors, markedly enhanced knowledge of the recognition repertoire.41 Finally came the discovery of cytosolic recognition proteins, the NLR, RIG-I families, and AIM2,46 implicated in genetic and microbially induced hyperinflammatory syndromes, Crohn’s disease, and nucieic acid sensing.

These molecules are mainly, although not exclusively, expressed by macrophages and DCs and play an important role in immune responses to foreign or modified host ligands, resulting in inflammation and autoimmunity. Given the complexity of microbial and cellular particulates compared with discrete soluble ligands, different receptors collaborate and synergize with one another to activate or inhibit subsequent APC responses. The original distinction drawn between pattern recognition receptors for exogenous, so-called pathogen-associated molecular patterns and for endogenous, host-derived ligands has been eroded to some extent because individual receptors are able to bind both types of ligand,47 although their distinct cellular expression and signaling responses may account for discrimination by host APCs. In particular, recognition and uptake of apoptotic cells result in downregulation of macrophage effector molecules, as opposed to the proinflammatory effects induced by microbial ligands. The signaling pathways induced by other, modified host-derived ligands, such as oxidized lipoproteins and hyaluronates, are not well characterized.

This section summarizes selected receptor structures, indicates some of their ligands, and addresses briefly signaling and antigen processing pathways. Further details are provided in Chapter 18, which also illustrates interactions between humoral and cellular arms of the innate response.

**Toll-Like Receptors**

Key features of TLRs are illustrated elsewhere (see Chapter 18.48-49). TLRs consist of homodimers or heterodimers of transmembrane glycoproteins containing extracellular leucine-rich repeats (LRRs) and cytoplasmic Toll/IL-1 receptor (TIR) domains, similar to the intracellular domain of the IL-1 receptor, which contains extracellular immunoglobulin superfamily domains. TLRs are expressed on the surface of APCs or, in the case of TLR3, TLR7/8, and TLR9, on vacuolar membranes, contributing to proinflammatory and immunogenic signaling after “sensing” of cargo. Isolated TLRs do not bind ligands directly but collaborate with humoral factors and with other plasma membrane glycoproteins, as in the well-studied case of TLR4, the glycosyl phosphoinositol (GPI)-anchored CD14 receptor for the plasma lipopolysaccharide-binding protein and the membrane-associated molecule MD2.

**Non–Toll-Like Receptors**

NTRs constitute a variety of lectin-like,45,50 scavenger, and other plasma membrane glycoprotein nonopsonic receptors; selected examples are illustrated in Figures 10-7 and 10-8. They include type 1 and 2 transmembrane molecules, with C-type lectin or lectin-like, collagenous, or immunoglobulin superfamily domains. In addition, macrophages express sialic acid–binding receptors, members of the SIGLEC family.51 Table 10-3 summarizes known ligands and potential functions relevant to their role in homeostasis and immunity. Apart from direct recognition of selected self, modified self, and foreign structures, they play an important role in cell-cell and cell-matrix interactions; in phagocytosis and endocytosis; and in less well-understood processes, such as targeted delivery of antigens to peripheral lymphoid organs.52

Receptors that promote rapid clearance of apoptotic cells, directly or after opsonization by complement and other extracellular proteins, are shown in Figure 10-9. More recent research has shown collaboration of different receptors at the cell surface,53 regulating cellular responses. Other regulatory surface receptors include transmembrane molecules with tyrosine-based activating (ITAM) or inhibitory (ITIM) cytoplasmic motifs, for example, C-type lectin-like molecules, immunoglobulin superfamily members such as Fc receptors, discussed further subsequently, and receptor-ligand pairs that regulate macrophage activation (CD200/CD200R)44 and phagocytosis (SIRPα and CD47) (Figure 10-10).

**Complement Receptors**

Complement receptors are heterogeneous receptors expressed by APCs (Figure 10-11) that mediate direct and lectin- and antibody-dependent binding of activated complement components and play a role in cell migration and phagocytosis and immune regulation.54 The discovery of a novel complement receptor selectively expressed by Kupffer cells suggests that mononuclear phagocyte heterogeneity in complement receptor function provides evidence of tissue specificity for some macrophage subsets.55 Genetic deficiencies in complement receptor expression and their ligands...
Figure 10-7  Selected lectin-like receptors expressed by macrophages and dendritic cells (DCs). (See Table 10-3 for ligands.) These receptors play an important role in nonopsonic adhesion, cell-cell interactions, endocytosis, and phagocytosis. The cysteine-rich domain of the mannose receptor has been postulated to play a role in transport of glycoconjugates to peripheral lymphoid organs for clearance or antigen-dependent activation of B lymphocytes. DEC205 has been used to show efficient targeting of antigens to DCs. The ITAM-like motif of Dectin-1, the β-glucan receptor, is essential for myeloid cell responses to fungal particles, in collaboration with TLRs and nLR. CTLD, C-type lectin domain.

Figure 10-8  Selected scavenger receptors. These non-TLRs consist of diverse structures and bind a range of microbial and endogenous ligands (see Table 10-3). They differ in regulation and expression, providing useful markers of macrophage activation (CD14, MARCO), heterogeneity (CD163, the scavenger receptor for clearance of haptoglobin-hemoglobin complexes), and lipid homeostasis (SR-A, CD36, LOX-1). Scavenger receptors such as MARCO can collaborate with TLRs in APC responses to microbes. APC, antigen-presenting cell; CTLD, C-type lectin domain; TLRs, Toll-like receptors.
Identification of microbial peptidoglycan breakdown and other products as ligands for an intracellular family of LRR-containing cytosolic sensors that regulate caspase activation and IL-1β secretion, especially in macrophages, provided a major impetus to this burgeoning field. Other chapters describe examples of NLRs and their role in hyperinflammatory syndromes, as well as RIG-I–like and other nucleic acid sensing receptors. NLRs and other cytosolic receptors participate in signaling pathways implicated in TLR-induced functions and play a major role in inflammasome activation, caspase-1 activation, and IL-1 and IL-18 release. RIG-I and MDA-5 are members of the RIG-I–like receptor (RLR) RNA helicase family, found in macrophages and other myeloid cells, as well as in nonhematopoietic cells. They play a role in recognition of viral RNA, such as single-stranded RNA with 5′-triphosphate and/or double-stranded RNA. Although distinct, they share signaling features (e.g., recruitment by adapter IPS [also known as MAVS, CARDIF, or VISA] to the outer membrane of mitochondria) and activate transcription factors involved in type I interferon induction and antiviral immunity. RLR recognition of RNA is predominant in macrophages, cDCs, and fibroblasts, but pDCs utilize TLR preferentially. Gene deletion or silencing has shown that RIG-I modulates the generation and have provided insights into the pathogenesis of autoimmune disease and of rheumatic joint injury. Complement deposition on follicular DCs has been implicated in the activation of B cell responses. Complement regulatory proteins are expressed by many cell types and APCs and play an important role in limiting cell activation. In addition, complement receptors collaborate with other receptors to modulate myeloid cell responses. Their signaling pathways are less well defined than those of Fc receptors.

### Fc Receptors

APCs express diverse receptors for monomeric and complexed immunoglobulins regulating effector responses to antigens. Figure 10-12 illustrates the properties of human Fc receptors with activating or inhibitory motifs in their cytoplasmic tails. Fc receptor polymorphisms have been implicated in autoimmune human diseases such as systemic lupus erythematosus. Studies on genetically manipulated mouse models have led to considerable insights into the role of individual Fc receptors in immunopathogenesis. Fc receptors are able to cooperate with chemokine and other receptors. The nature and role of Fc-independent interactions of antibody with APCs are under investigation and include lectin-like recognition of sialylated and mannosyl residues.

### Cytosolic Receptors

<table>
<thead>
<tr>
<th>Class</th>
<th>Receptor</th>
<th>Microbial Ligands</th>
<th>Endogenous Ligands</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scavenger receptors</td>
<td>SR-A I/II</td>
<td>Gram-positive/gram-negative bacteria Lipoteichoic acid</td>
<td>Apoptotic cells Modified low-density lipoproteins AGE-modified proteins β-Amyloid Marginal zone B lymphocytes</td>
<td>Phagocytosis Endocytosis Adhesion</td>
</tr>
<tr>
<td></td>
<td>MARCO</td>
<td>Lipid A Neisserial surface proteins Gram-positive/gram-negative bacteria Neisserial surface proteins</td>
<td>Uteroglobin-related protein</td>
<td>Foam cell formation Adhesion</td>
</tr>
<tr>
<td></td>
<td>CD36</td>
<td>Diacylated lipopeptide from gram-positive bacteria Plasmodium falciparum parasitized erythrocytes</td>
<td>Apoptotic cells (with thrombospondin and vitronectin receptor) HDL Outer rod segments</td>
<td>Innate activation Uptake, exchange of lipids, macrophage fusion</td>
</tr>
<tr>
<td>Lectins</td>
<td>Dectin-1</td>
<td>β-Glucan</td>
<td>T lymphocytes (noncarbohydrate) ICAM 2/3 T lymphocytes Lysosomal hydrolases</td>
<td>Fungal uptake and immunomodulation Endocytosis</td>
</tr>
<tr>
<td></td>
<td>DC-SIGN</td>
<td>Mannosyl/fucosyl glycoconjugates including viruses (e.g., human immunodeficiency virus-1, dengue)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mannose receptor CRD</td>
<td>Mannosyl/fucosyl glycoconjugates on bacteria, viruses, fungi, parasites</td>
<td>Thyroglobulin Ribonuclease β-Amylase</td>
<td>Adhesion</td>
</tr>
<tr>
<td></td>
<td>Cysteine-rich domain</td>
<td></td>
<td></td>
<td>Antigen targeting</td>
</tr>
<tr>
<td></td>
<td>Fibronectin type II domain</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This table illustrates the dual recognition properties and diverse cellular functions of selected non–Toll-like receptors. For additional details, see references 45, 101, and 107-110. AGE, advanced glycation end product; CRD, carbohydrate recognition domains (Ca+++-dependent); HDL, high-density lipoprotein; ICAM, intercellular adhesion molecule.
**Figure 10-9** Phagocytic receptors for apoptotic cell phagocytosis. Macrophages (MØ) and immature myeloid dendritic cells (DCs) are the main immune cells involved in the clearance of apoptotic cells. They express broadly similar multiple receptors, which can bind directly or via opsonic soluble proteins (e.g., mannose-binding lectins [MBLs] to ligands). Phosphatidylserine (Ptdser) becomes exposed on the outer surface of the apoptotic cell, and a receptor for this ligand has been long sought. A new receptor (TIM4 and related TIM1) has been discovered on resident MØ, with specificity for PtdSer. Other MØ populations use MFGE8 (a milk fat globulin protein secreted by MØ) as an opsonin. Discrimination of nonself and altered self may involve combinations of different phagocyte receptors. Apoptotic cell uptake results in an anti-inflammatory response by MØ (e.g., release of TGF-β and PG E2) and has been implicated in cross-presentation by DC. MFGE8, milk fat globule-EGF factor 8 protein; PGE2, prostaglandin E2; TGF, transforming growth factor; TIM, T cell immunoglobulin mucin. (From Savill J, Dransfield I, Gregory C, et al: A blast from the past: clearance of apoptotic cells regulates immune responses, Nat Rev Immunol 2:965–975, 2002.)

**Figure 10-10** Surface receptors that, singly or as paired receptors, modulate macrophage responses. Illustrated are immunoglobulin superfamily and associated transmembrane molecules using ITAM or ITIM motifs to generate activating or inhibitory signals. DAP12 can associate with a range of other membrane molecules as a signaling partner in macrophages, dendritic cells, and natural killer cells. Note the charged amino acid residues in transmembrane domains. CD200 and CD47, which are broadly expressed in tissues, generate inhibitory signals in macrophages via CD200R54 and SIRPα. In CD200 knockout mice, macrophages show spontaneous activation. Macrophages themselves are induced by Toll-like receptor stimuli to express CD200.54
Cells Involved in Autoimmune Diseases and Inflammation

PART 2

Figure 10-11 Complement receptors and membrane regulators expressed by macrophages. CR1 is broadly expressed by nucleated cells, acting as a "sink" for activated complement; CR3 (CD11b/CD18), a phagocytic receptor for C3b-coated particles, and CR4 (CD11c/CD18) are β2 integrins, which, together with LFA-1 (CD11a/CD18), mediate adhesion of myeloid cells to endothelium and extracellular matrix and migration. huCRig L and S are long and short forms of a newly described complement-binding receptor on Kupffer cells, which mediate uptake of opsonized bacteria. CD55 and CD59 are glycosyl phosphoinositide (GPI)-anchored regulators of complement activation.

Figure 10-12 Human Fc receptors. Myeloid cells express a range of classic Fc receptors that initiate a variety of cellular responses, including phagocytosis, antibody-dependent cell-mediated toxicity, antigen presentation, respiratory burst, and release of inflammatory mediators. Immunoglobulin subclasses are bound by extracellular domains; signaling via cytoplasmic ITAM or ITIM is mediated by associated membrane-spanning polypeptides. Activation and inhibitory receptors usually are co-expressed on the cell surface and function in concert, determining the magnitude of effector cell responses. A range of Fc receptor–like molecules (immunoglobulin superfamily [IGSF] extracellular domains), with similar ITAM/ITIM cytoplasmic motifs, are mainly expressed by B or T/natural killer cells; they may regulate lymphocytic differentiation and responses. GPI, glycosyl phosphoinositide; ITAM, immunotyrosine activation motif; ITIM, immunotyrosine inhibition motif.
proteins AIM-2 and IFI16 form a new family of innate DNA sensors, termed AIM-2–like receptors (ALRs).64

**Damage/Danger-Associated Molecular Patterns (DAMPs)**

Host-derived molecules that activate TLR pathways and inflammasomes such as NLRP3 include nuclear and cytosolic proteins (e.g., HMGB1, S100), matrix components (hyaluronate), adenosine triphosphate (ATP), and uric acid. They are released by cell stress, especially necrosis, and contribute to the initiation and perpetuation of sterile and infectious inflammation. Other receptors include receptor for advanced glycation end products (RAGE) and possibly scavenger receptors, although ligands are promiscuous and often are poorly defined. The concept gained currency from work by Matzinger and colleagues.65 Various hypotheses have been proposed for inflammasome activation, including K⁺ efflux, membrane pore formation, lysosome disruption and cathepsin release, and reactive oxygen species (ROS), which may be necessary but is not sufficient for caspase activation.66

**RESPONSES AND MODULATION**

This section summarizes some of the major cell biologic effects resulting in antigen processing and presentation by APCs. Other chapters address related topics, such as recognition of lipids by CD1 and intracellular trafficking of related molecules.

**Phagocytosis and Endocytosis: Antigen Processing**

The vacuolar apparatus of APCs is illustrated schematically in Figure 10-13.67,68 Internalization of the plasma membrane results in phagosome/endosome formation, with progressive acidification and digestion, depending on delivery of vesicles and their hydrolytic contents. Membrane and receptors are recruited, modified by maturation, and are retrieved by recycling. Further fusion with Golgi-derived vesicles and primary lysosomes yields phagolysosomes and secondary lysosomes, reaching a pH of approximately 5.5 to 6.0. Depending on the bulk of plasma membrane internalized and the size of the particle, the uptake process involves cytoskeletal components and small guanosine triphosphatases (GTPases) and docking machinery. Guanosine triphosphate (GTP) hydrolysis serves as an important mechanism for controlling intracellular membrane traffic and coupling to the cytoskeleton. Cytokines such as interferon-γ can have a major effect on activation and relocation of GTP-binding proteins, contributing to host cell–pathogen interactions.70

Proteomic analysis of isolated phagolysosomes has revealed more than 600 constituents, classified according to the source of the membrane and various putative functions.71 Although most membrane constituents are derived from the plasma membrane, descriptions of contributions from the endoplasmic reticulum have elicited considerable interest, along with different estimates of their extent.71,72 Another controversial aspect relates to the role of TLR engagement in enhancing the kinetics of the maturation process, a transcription-independent process.73 Variations on this theme occur, depending on cell type and maturity, whether the extracellular cargo is microbial or host-derived, and whether the process arises by autophagy,74 rather than heterophagy. The role of autophagy in cellular resistance to intracellular pathogens, such as M. tuberculosis, is receiving increasing consideration.75,76

Immature DCs are actively endocytic and phagocytic but poor APCs, whereas mature DCs downregulate uptake but acquire highly efficient APC function.18 Changes associated

---

**Figure 10-13** Schematic presentation of phagocytosis and endocytosis. Particulates are taken up by actin-dependent sequential maturation processes involving membrane fusion and fission, which intersect with the endocytic pathway at several stages. Cytosolic small guanosine triphosphatases (GTPases) (Rabs) determine organelle-specific interactions. Membrane is recycled to the plasma membrane with processed antigen (see Figure 10-14). Progressive acidification and delivery of lysosomal hydrolases result in terminal degradation. Compartment membranes express marker proteins such as LAMP1; the pan-macrophage CD68 antigen is associated with late endosomes and lysosomes.
with DC maturation and antigen presentation are illustrated in Figures 10-7 through 10-14. Pathogenic intracellular organisms vary in their subversion of the aforementioned process, interfering with different stages, such as cytosolic signaling mechanisms, fusion, or acidification, and in selected cases, inducing a novel membrane composition. Organisms can replicate in immature or mature compartments, or can translocate their genome to the cytosol by acid-induced envelope fusion or by disruption of the vacuolar membrane.

**Signaling**

The best-characterized signaling responses in macrophages include the TLR, type I interferon, and Fc receptor-induced pathways. The various signaling cascades are complex, interact with one another, and result in phosphorylation/dephosphorylation, formation of cytosolic protein complexes, and activation of signaling proteins such as nuclear factor κB (NFκB), which enter the nucleus to regulate transcription. In the case of TLR-dependent sensing, a restricted number of adapters, such as MyD88, TIRAP-MAL, and TRIF, channel the flow of information into the cell. Emerging evidence suggests that NTR-induced signaling (e.g., by the β-glucan receptor [Dectin-1] cytoplasmic ITAM-like motif) collaborates with TLR and NLR signaling pathways, with differential involvement of syk and CARD9 in effector responses in different mononuclear phagocytes. Distinct but interlinked pathways involve interferon regulatory factors as part of an amplification pathway with broader immunoregulatory functions than antiviral responses alone. A more recently defined cytosolic specialized antiviral pathway involves interaction with mitochondrial components. Cell surface or intracellular proteins serve as negative regulators, as exemplified by members of the SOCS family.

Figure 10-14 illustrates the pathway by which effector response–derived antigens are transported to the cytosol during synthesis, are processed in proteasomes, and become associated with MHC class I molecules for presentation at the cell surface. Effector response chaperones, such as heat shock protein, may contribute to the surface delivery of antigens. The inflammasome is another example of a multiprotein assembly within the cytosol by which NLRs bring about IL-1 activation.

Several of the processes already described involve proteolytic enzymes, including cathepsins and caspases, and lipid-interacting chaperones. The ubiquitin pathway provides an important mechanism for cytosolic degradation of proteins. In addition, vacuolar H^+ ATPases contribute to acidification in different mononuclear phagocytes, especially relevant to osteoclast function. Comparison and further elucidation of the coupling between extracellular and intracellular signaling pathways in macrophages, DCs, and osteoclasts would be of great interest.
Efferent Pathways: Gene Expression and Secretion

The complex responses of different mononuclear phagocytes to intrinsic and extrinsic stimuli have been defined by microarray studies and, in some cases, by protein and functional analyses. Although the major differentiation pathways (macrophages, immature and mature DCs, and osteoclasts) involve selective, stereotypic changes in gene expression, the extensive heterogeneity and plasticity of phenotypes characterizing these cells in situ have begun to be appreciated only more recently.4 Macrophages are known to express numerous nuclear receptors,82 which undergo dynamic changes depending on cellular differentiation and the microenvironment, as well as epigenetic histone modification83 and micro-RNA expression. Exposure to particular cytokines, hormones, and other stimuli, such as microbial antigens and immune complexes, modulates gene expression profoundly and selectively and induces posttranscriptional and posttranslational changes. Through alternative splicing, glycosylation, and poorly characterized protein modifications (e.g., methylation and acetylation of nuclear and cytoplasmic macromolecules), this exposure provides different mononuclear phagocytes with extensive diversity in adaptation and function.

Table 10-4 provides a partial list of macrophage secretory products; these include low-molecular-weight metabolites (oxygen, nitrogen, and lipid-derived) implicated in inflammation and its resolution84 and in antimicrobial responses; other products include proinflammatory and anti-inflammatory cytokines and neutral proteinases. The cell biology of the secretory pathway in macrophages has only begun to be studied. The differential effect of T helper type 1 and T helper type 2 cytokines on nitric oxide production versus its destruction is thought to be relevant to microbicidal mechanisms versus repair.4 The secretory activity of DCs and osteoclasts is less defined except with regard to their specialized functions (e.g., lymphocyte activation, bone remodeling).

Table 10-5 shows a classification of characteristic effects induced by prototypic activating and inhibitory stimuli. It is convenient to distinguish the phenotype of mononuclear phagocytes stimulated by an innate stimulus, including exposure to microbial products such as lipopolysaccharide;
Table 10-5 Immunomodulation of Macrophage Phenotype

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Category</th>
<th>Markers</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microbial (bacterial)</td>
<td>Innate Activation</td>
<td>Induction of MARCO, Co-stimulatory molecules, CD200, Induction of MHC II</td>
<td>Enhanced phagocytosis, Antigen presentation, Inhibition (CD200R)</td>
</tr>
<tr>
<td>Interferon-γ</td>
<td>Classic activation</td>
<td>Potentiation of innate markers, TNF, iNOS induction, NADPH, respiratory burst, LGP47 induction</td>
<td>Proinflammatory, Antimicrobial (NO), signaling, Host defense, inflammation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Downregulation of MR, Modulation of FcR expression, Proteasomal composition</td>
<td>Association with phagosome/intracellular pathogen killing, Unknown</td>
</tr>
<tr>
<td>IL-4/IL-13</td>
<td>Alternative activation Upregulation</td>
<td>Enhanced MR, Induction of arginase, Induction of YM1, FIZZ1 (mouse), Induction of CCL17 (MDC) and CCL22 (TARC)</td>
<td>Antigen presentation, Endocytosis, Humoral immunity, Th2 responses, Allergy, antiparasitic, Immunity</td>
</tr>
<tr>
<td>IL-10 TGF-β</td>
<td>Deactivation</td>
<td>Downregulation of MHC II, NO and ROI</td>
<td>Repair/fibrosis</td>
</tr>
<tr>
<td>Immune complexes</td>
<td>Modified activation</td>
<td>Selective IL-12 downregulation</td>
<td></td>
</tr>
<tr>
<td>Glucocorticoids</td>
<td>Deactivation</td>
<td>IL-10 induction, CD163 induction, Monocyte recruitment downregulated ACE induction, Stabilin induction</td>
<td>Anti-inflammatory, Homeostatic clearance, Hemoglobin-haptoglobin complexes</td>
</tr>
</tbody>
</table>

Note: For additional details, see references 4, 16, and 113.

ACE, angiotensin-converting enzyme; FcR, Fc receptor; IL, interleukin; iNOS, inducible nitric oxide synthase; LGP47, lysosomal glycoprotein; MARCO, macrophage scavenger receptor; MDC, macrophage-derived chemokine; MHC, major histocompatibility complex; MR, mannose receptor; NADPH, nicotinamide adenine dinucleotide phosphate-oxidase; NO, nitric oxide; ROI, reactive oxygen intermediates; TARC, thymus and activation-regulated chemokine; TNF, tumor necrosis factor.

by a classic immune activating cytokine such as interferon-γ, and by IL-4/IL-13 induction of an alternative activation pathway4 by IL-4/IL-13. Other stereotypic responses are induced by immune complexes that selectively modulate IL-12/IL-10 expression.55 Innate and immune cytokine stimuli are able to potentiate one another.

Conversely, deactivating stimuli such as IL-10 and transforming growth factor (TGF-β) and glucocorticosteroids16 induce their own distinctive patterns of gene expression. Upregulation of CD163, a hemoglobin-haptoglobin scavenger receptor, is a striking example of glucocorticosteroid-enhanced endocytic function.87 Stimulation of the vagus nerve attenuates macrophage activation by activating the Jak2-STAT3 signaling pathway.88 Deactivation of macrophages plays a critical role in limiting proinflammatory activation induced by cytokines (e.g., interferon-γ) and opsonic (FcR) and nonopsonic stimulation (TLR, NLR, inflammasomes), acting via a range of signal transduction pathways and naturally occurring antagonists. Deactivation of macrophages is a particular feature of phagocytic uptake of apoptotic cells (see Figure 10-9) thought to be mediated by TGF-β, prostaglandin (PG)E2, and IL-10, which also inhibit the generation of reactive nitrogen and oxygen species. The resolvins and protectins derive from the omega-3 fatty acids eicosapentanoic acid (EPA) and docosahexaenoic acid (DHA) produced by the cyclooxygenase (COX)-2 pathway, especially in the presence of aspirin.

They inhibit production and recruitment of leukocytes to sites of inflammation and counteract pain.89

Plasma membrane signals to deactivate macrophages arise from trans-lation of CD200 (expressed by many cells, including innate-activated macrophages) and its signaling immunoglobulin superfamily (IGSF) counterpart CD200R, expressed selectively by macrophages.54 Another inhibitory receptor pair consists of SIRPα and CD4717; cytotoxic inhibitors include the SOCS proteins.20 These extracellular mediators and poorly defined interactions with surface expressed molecules make the different APCs reciprocally interactive with other immune cells (helper and regulatory CD4 T cells, CD8 T cells, natural killer cells, and natural killer T [NKT] cells) and with nonhematopoietic cells such as fibroblasts, epithelium, and neurons. Induced responses on macrophages include altered MHC class II and co-stimulatory molecule expression, marker antigens involved in cellular interactions, phagocytosis and antigen presentation, and altered secretion. As a result, different mononuclear phagocytes are able to modulate their interactions with CD4 T lymphocytes, with endothelial and other cell types, with extracellular matrix, and with bone to induce trophic and cytostatic/cidal and catabolic effects.

Human inborn errors or defects in macrophage responses to intracellular pathogens such as mycobacteria90 have validated studies on cell activation in the mouse. The immunosuppressive effects of macrophages in immune tolerance
to an allogeneic fetus have been ascribed to induction of indoleamine-2,3-dioxygenase, which catabolizes the essential amino acid, L-tryptophan.\textsuperscript{91} Macrophage products act mainly locally but also can have profound systemic effects, regulating metabolic and regulatory responses within the host. These effector programs and products provide targets for pharmacologic intervention.

**Relevance to Selected Rheumatic Diseases**

Although the different mononuclear phagocytes have been addressed as part of a general host homeostatic system, there are good examples in human rheumatic diseases and in animal models of arthritis in which these cells play a role in the local initiation or effector phase of the pathology. Primary defects in osteoclast differentiation and function contribute to osteopetrosis or, if overactive, osteoporosis. I have referred to autoinflammatory syndromes and the newly discovered cytosolic receptors. Gout might also fall in this category because uric acid crystals provide a metabolic stimulus for inflammasome activation. Macrophages play a role in the pathogenesis and complications of osteoarthritis, including the response to foreign implants. DCs and macrophages play a part in the induction and effector mechanisms of autoimmune arthritis, in concert with T lymphocytes and B lymphocytes and their products, immune complexes, and complement. The catabolic role of TNF, IL-1, neutral proteases, reactive oxygen radicals, and arachidonate metabolites is well known. Less well appreciated are the trophic interactions of macrophages with fibroblasts through production of TGF-β, growth and angiogenic factors, and modulation by alternative activation pathways.\textsuperscript{92} Type I interferons, produced by macrophages and other cells, play an important amplification role in autoimmune inflammation (e.g., systemic lupus erythematosus [SLE]).

Although animal models may not mimic the natural diseases in humans, more recent examples suggest that APC membrane receptors involved in recognition of modified self and foreign microbial ligands could be relevant to arthritis—in streptococcal rheumatic disease and in exacerbation of recurrent arthritis by zymosan in T cell transgenic mice, acting via Dectin-1, the APC receptor for β-glucans.\textsuperscript{93} In addition, inhibitory plasma membrane receptor interactions (CD200/CD200R) limit macrophage activation and collagen-induced murine arthritis.\textsuperscript{34,94} Impaired clearance of apoptotic cells by (uncharacterized) C1q receptors of mononuclear phagocytes may contribute to autoimmune diseases such as lupus.\textsuperscript{38} Finally, macrophages could contribute to wider connective tissue disorders, such as scleroderma, because they are able to regulate fibroblast and matrix synthesis and turnover through metalloproteinases and TGF-β; dysregulation of these pathways has been neglected as a pathogenetic mechanism in these metabolic diseases.

** ISSUES FOR FURTHER INVESTIGATION**

As we take stock of how increasing knowledge of mononuclear phagocyte biology has affected pathogenesis and therapy of rheumatic diseases, additional questions arise regarding basic mechanisms and new selective targets. Treating macrophages, DCs, and osteoclasts as specialized forms of a unitary mononuclear phagocyte family helps to bring out their common and distinctive properties. Selective expression and signaling pathways of pattern recognition receptors by macrophages and DCs are still poorly characterized—regulation of TLR signaling has already received considerable attention as a possible drug target. Discrimination of microbial and host-derived ligands may account for mimicry and autoantigen cross-reactivity, but mechanisms for regulating homeostatic and tolerogenic responses are still unclear.

On the effector side, success with anti-TNF monoclonal antibodies suggests that targeting of individual effector molecules can succeed, especially when these are expressed at the cell surface and are proximal in cascade reactions. The impact of anti-TNF therapy on macrophage functions and mechanisms for resistance to TNF blockade are poorly understood. The role of inhibitory/activating pairs of surface molecules as modulators of APC pathways is unexplored and is likely to move center stage. One neglected aspect of differentiation and function for osteoclasts and granuloma macrophages is the mechanism of fusion and the possible functional significance of multinuclear giant cell formation. Basic knowledge is still lacking and may provide useful insights into the pathogenesis of metabolic bone diseases and enhanced perception of links with the immune system, both innate and adaptive.

Although powerful tools of mRNA and proteomic analysis and gene silencing by RNAi inhibition have become available, the heterogeneity and complexity of mononuclear phagocytes in human tissues, especially at early stages of disease, need improved characterization in situ and at different sites. Much remains to be learned concerning trophic processes and repair and their relationship to resolution of inflammation and fibrosis. These involve cell-cell interactions that are poorly understood but are increasingly amenable to investigation.

**Acknowledgments**

The author thanks Christine Holt and Amelia Molloy-Bland for preparation of the manuscript and present and past members of his laboratory for help with illustrations. Work in the Gordon Laboratory has been supported by the Medical Research Council, United Kingdom; the Wellcome Trust; the Arthritis Research Campaign; and the British Heart Foundation.

**References**


Neutrophils

JOSE U. SCHER • STEVEN B. ABRAMSON • MICHAEL H. PILLINGER

INTRODUCTION: POLYMORPHONUCLEAR LEUKOCYTES

Polymorphonuclear leukocytes constitute a family of hematopoietically derived cells that share the feature of a multi-lobed nucleus. These leukocytes also share a property of possessing highly developed populations of intracytoplasmic granules, divisible into subsets and distinct between cell types. The presence of these granules permits the further designation of polymorphonuclear leukocytes as granulocytes. On the basis of the histocytochemical staining properties of their respective granules, three classes of polymorphonuclear leukocytes have been identified: neutrophils, eosinophils, and basophils. Neutrophil (polymorphonuclear neutrophil) granules stain with neutral dyes; the granules of eosinophils are most effectively stained with acidic dyes such as eosin, and basophil granules stain with basic dyes. In a standard polychromatic Wright stain of a peripheral blood smear, the cytoplasm of neutrophils, eosinophils, and basophils appears blue-pink (neutrophils), pink (eosinophils), and blue (basophils). These classes of polymorphonuclear leukocytes differ with respect to not only appearance but also biochemistry and function. Polymorphonuclear leukocytes constitute an important part of the organism’s system of innate immunity: Their responses to foreign organisms, antigens, or both are preprogrammed and do not depend on prior exposure to the particle.

Neutrophils

Neutrophils are the body's first line of defense against foreign invaders and constitute the major cell type involved in acute and some forms of chronic inflammation. The importance of neutrophils in bacterial defense is illustrated by patients who have hereditary defects in neutrophil function and are prone to repeated and often life-threatening infections. Neutrophils are the most prevalent leukocyte in the bloodstream, typically constituting greater than 50% of all bloodstream leukocytes. During bacterial infection, the percentage of neutrophils may increase to 80% or more. In contrast, tissue concentrations of resting (inactive) neutrophils seem to be quite low. Neutrophils may be considered surveillance cells—sweeping through the bloodstream scanning for infections or other inflammatory events. The capacity of neutrophils to destroy foreign organisms is matched in some circumstances, however, by a capacity for host tissue destruction.

Neutrophil Myelopoiesis and Clearance

The neutrophil majority in the bloodstream is duplicated in the bone marrow, where 60% of hematopoietic capacity may be dedicated to neutrophil production. Daily, 10¹¹ neutrophils are released into the bloodstream. Neutrophil development in the marrow takes about 14 days, originating from the hematopoietic stem cell. Stem cells fated to become neutrophils first differentiate into myeloblasts, which retain the capacity to develop into eosinophils, basophils, and neutrophils. Subsequent differentiation leads to the neutrophilic promyelocyte, a dedicated precursor of the neutrophil, and proceeds through the stages of neutrophilic myelocyte, metamyelocyte, band cell, and mature neutrophil. At the metamyelocyte stage, neutrophil mitosis ceases, whereas neutrophil development and organization of granules continue. Neutrophils are terminally differentiated; they neither divide nor alter their gross phenotype after their release from the marrow.

Given the origin of neutrophils in a pluripotent stem cell, as well as the precise phases of their development, the mechanisms regulating neutrophil differentiation are of considerable interest. Although incompletely understood, studies have emphasized the role of a particular complement of transcription factors and cytokines that seem to direct the early cells toward neutrophil development. Several myeloid factors are necessary for the transcriptional regulation of neutrophils including LEF-1, CCAAT enhancer binding proteins α and ε (C/EBPα and C/EBPε), and GFI-1. Principal among the cytokines regulating granulopoiesis is granulocyte colony-stimulating factor (G-CSF). G-CSF effects include induction of myeloid differentiation,
The proliferation of granulocyte precursors, and release of mature neutrophils from the marrow. Biologic effects of G-CSF are mediated through its receptor (G-CSFR or CD114), a member of the class I cytokine receptor family. Although other hematopoietic cytokines contribute to granulopoiesis in vivo (including granulocyte-macrophage colony-stimulating factor [GM-CSF], interleukin [IL]-6 and IL-3), their individual presence is not essential as demonstrated by knockout experiments.

Once mature, neutrophils exit the bone marrow through the tight-fitting pores of the sinusoidal endothelium and enter the circulation, a process called transcellular migration. Neutrophils released from the marrow have a bloodstream half-life of approximately 6 hours and a tissue half-life only marginally longer. Neutrophil life spans may be modulated by soluble signals: when exposed to stimuli such as tumor necrosis factor (TNF) and Fas (CD95) ligand, neutrophils undergo apoptosis or programmed cell death. The high output and short half-life of neutrophils imply that neutrophil clearance mechanisms must exist. Recently, the SDF-1/CXCR4 signaling system has been implicated in neutrophil clearance. CXC chemokine receptor 4 (CXCR4), a G-protein coupled receptor, is expressed at low levels in the mature neutrophil. As they age, neutrophils alter their phenotype and upregulate CXCR4. This change supports homing to the bone marrow via the chemoattractant stromal-derived factor 1 (SDF-1 or CXCL12). Once back in the marrow, senescent neutrophils are phagocytosed by stromal macrophages. Senescent or apoptotic bloodstream neutrophils are also cleared by liver and spleen macrophages (reticuloendothelial system). Although little is known about the molecular mechanisms underlying neutrophil clearance in the liver and spleen, upregulation of the adhesion molecule P-selectin on Kupffer cells appears to be relevant. It remains a matter of speculation whether tissue neutrophils are cleared primarily via local macrophages.

Neutrophil Morphology and Content

Neutrophil nuclei tend to have more lobes than nuclei of other polymorphs, typically three to five (Figures 11-1 and 11-2). The multilobed nature of this nucleus reflects a condensation of chromatin and suggests that neutrophils might be incapable of transcription. It is now appreciated, however, that neutrophils retain the capacity for constitutive and stimulated protein synthesis, albeit at a limited rate.

Neutrophil granules identifiable by classic histochemical staining comprise two classes (see Figure 11-1). Neutrophil primary granules form first and, by virtue of their staining tendencies (affinity for the basic dye azure A), are also referred to as azurophilic granules. These granules are oval or round and vary in size. They are similar, and functionally equivalent, to the lysosomes of other cells. In contrast, neutrophil secondary granules constitute a population unique to neutrophils, a fact reflected in the alternative nomenclature of specific granules often used to describe these.
structures. Characteristic of azurophilic granules is the presence of myeloperoxidase (MPO), an enzyme that catalyzes the formation of hypochlorous acid from chloride in the presence of superoxide anion (O$_2^-$). The presence of large amounts of this enzyme in azurophilic granules lends collections of neutrophils (pus) their typical greenish yellow color. Consistent with their role as lysosomes, azurophilic granules also contain a variety of proteases and other enzymes including elastase, lysozyme, acid phosphatase, and cathepsins and enzymes directed at nucleic acids and sugars. At the membrane level, however, they differ from true lysosomes because they lack lysosome-associated membrane proteins 1 and 2 (LAMP-1 and LAMP-2) and the mannose-6-phosphate receptor system. In contrast to azurophilic granules, specific granules possess an extensive array of membrane-associated proteins including cytochromes, signaling molecules, and receptors. Specific granules constitute a reservoir of proteins destined for topologically external surfaces of phagocytic vacuoles and the plasma membrane (Table 11-1). One particularly important family of proteinases found in neutrophil specific granules are the matrix metalloproteinases (MMPs) including neutrophil collagenase-2 (MMP-8), gelatinase B (MMP-9), stromelysin (MMP-3), and leukolysin (MMP-25). They are stored as inactive proenzymes and undergo proteolytic activation after granule fusion and interaction with azurophilic granules, gaining the ability to alter integral components of bacterial membranes. Neutrophil MMP function is not limited to bacterial killing because MMPs are also important for extravasation and diapedesis.

Further study has confirmed the existence of two additional classes of vesicles. Gelatinase granules are almost identical in size to specific granules and share some proteins in common with them. As their name implies, however, gelatinase granules are distinguished by their high concentrations of gelatinase, a latent enzyme with the capacity for tissue destruction. Secretory vesicles are smaller and lighter than the other classes and do not seem to contain proteolytic enzymes. Rather, secretory vesicles are noteworthy for an extensive complement of membrane-associated proteins including receptors otherwise identified with the plasma membrane. These and other data suggest that the secretory vesicle is a reservoir of neutrophil plasma membrane and other membrane proteins.

Azurophilic and specific granules further contain antimicrobial proteins and peptides that are a cornerstone of innate immunity. A detailed description of the neutrophil’s armamentarium against foreign invaders is beyond the scope of this chapter, but a few whose mechanisms of action have more recently been elucidated warrant mention. Elastase, mentioned previously, aids in the killing of gram-negative bacteria via degradation of bacterial outer membrane protein A. Elastase-deficient mice are more susceptible to infection with gram-negative (but not gram-positive) organisms than wild-type mice. The defensins, normally located in azurophilic granules, are found in mg/mL concentrations in the phagocytic vacuole (see later) and render target cell membranes permeable. Bactericidal/permeability-inducing protein, also located in azurophilic granules, acts in concert with the defensins; it potently neutralizes endotoxin and is cytotoxic to gram-negative bacteria. Bactericidal/permeability-inducing protein also enhances the activity of secretory phospholipase A$_2$, which has activity against gram-negative and gram-positive bacteria. Lactoferrin (found in specific granules) deprives microorganisms of iron and has antiviral and antibacterial effects. Other granule-associated proteins such as cysteine-rich secretory protein 3 (CRISP3) and ficolin 1 have recently been described, although their function is still unclear.

Neutrophil granule contents might play important roles beyond their direct antimicrobial effects; they also may amplify or dampen the innate and adaptive immune response. Lactoferrin released during phagocytosis may inhibit proliferation of mixed lymphocyte cultures by decreasing release of interleukin (IL)-2, TNF, and IL-1. Proteinase 3 has been found to augment release of active TNF and IL-1 in monocyte/neutrophil co-cultures by releasing the membrane-bound forms of these cytokines. Gelatinase B has been shown to convert latent IL-1 into its active form and to potentiate IL-8 activity by truncating this chemoattractant and increasing its release, consequently amplifying neutrophil influx.
elastase also may play a proinflammatory role by virtue of its ability to cleave and disrupt phosphatidyl serine receptors on macrophages. Apoptotic cells undergo membrane alterations that lead to expression of phosphatidyl serine on their outer membrane surface, and interaction of phosphatidyl serine with its receptor leads to macrophage responses that downregulate inflammation through the generation of transforming growth factor (TGF)-β. By disrupting these interactions, the presence of neutrophil elastase may permit inflammation to continue.

**Neutrophil Activation and Function**

For bloodstream neutrophils to destroy foreign targets in the periphery, they first must sense the presence of such targets at a distance. They must then attach to the activated endothelium through multiple interactions involving adhesion molecules and their receptors (rolling and adhesion). After passing through the endothelium of postcapillary venules (diapedesis), neutrophils migrate to the source of the signal (chemotaxis). Finally, neutrophils must encounter a target, engulf it, and destroy it. Collectively, these processes are referred to as **neutrophil activation**. Because of the potential for tissue destruction, neutrophil activation must be carefully regulated. The internal responses through which a cell translates an encounter with a stimulus into a particular

**Stimuli and Receptors**

Classic chemoattractants include lipid mediators (e.g., leukotriene B₄ [LTB₄], platelet-activating factor) and proteins/peptides (e.g., formylated peptides, the complement split product C5a, and IL-8). Chemoattractants in vivo are formed at sites of inflammation—either produced at the site by inflammatory cells such as LTB₄ or IL-8 or liberated from already synthesized proteins, as in the case of C5a. The ability of formylated peptides such as N-formyl-methionyl-leucyl-phenylalanine to stimulate neutrophils probably represents a particularly ancient arm of the innate immune response because prokaryotic, but not eukaryotic, cells synthesize proteins whose first amino acid is a formylated methionine. Chemoattractants also have the capacity to stimulate most other aspects of neutrophil activation. Their individual potencies for particular responses may differ, however, suggesting that they may serve overlapping but distinct functions in neutrophil activation. CXC chemokines are a recently described group of chemoattractants, characterized by the presence of two N-terminal cysteines (“C”) separated by any other amino acid (“X”) at the carboxy terminus. Many CXC chemokines appear essential

---

**Figure 11-3**  Signaling pathways in neutrophil activation. Engagement of Fc, growth factor, and chemoattractant receptors and adhesion molecules initiate signaling pathways that result in proinflammatory neutrophil responses including cytoskeletal and morphologic changes, activation of adhesion molecules and the superoxide generating system (NADPH oxidase), and regulation of transcription. Some of the well-established pathways participating in these responses are illustrated (see text for details).
for recruitment including IL-8 (CXCL8), KC (CXCL1), and MIP-2 (CXCL2).

Bloodstream activation of neutrophils depends on the presence of specific surface receptors. Most chemoattractant receptors belong to a class known as seven-transmembrane-domain receptors, “serpentine seven” receptors, or G protein-coupled receptors (GPCR); these receptors are composed of a single protein chain whose seven hydrophobic domains span the plasma membrane. Binding of a particular chemoattractant occurs in a pocket on the cytoplasmic face, close to or below the level of the plasma membrane. Receptors for soluble ligands other than chemoattractants have also been identified on neutrophils including receptors for growth factors, colony-stimulating factors, and cytokines, as described. These receptors fall into several families distinct from the serpentine sevens. Growth factor receptors are members of the protein tyrosine kinase receptor family, in which ligand interaction with two identical or related receptors brings them into proximity, causing their phosphorylation and activation. Recent systems biology studies have shown that receptors for a variety of inflammatory stimuli become highly expressed only in mature neutrophils. The most notable examples include CXC and CC chemokine receptors such as IL-8Rα and β, CXCR-4 and CCR-1, 2, and 3; the cytokine receptors for TNF 1 and 2; interferon (IFN)α and -γ and interleukin receptors IL1R, IL4R, IL6R, IL10R, and IL17R. Some nonchoemoattractant ligands do not directly activate neutrophils but may modulate their function. Pretreatment of neutrophils with either insulin or GM-CSF results in amplification of subsequent neutrophil responses to chemoattractants, a process referred to as priming.

**G Proteins**

Ligation of seven-transmembrane-domain receptors results in the interaction of cytoplasmic elements of the receptor with a class of effectors known as heterotrimeric guanosine triphosphate (GTP)-binding proteins (G proteins). G proteins are composed of α, β, and γ subunits, and individual G protein types are distinguished by their particular combination of subunits. In neutrophils, the predominant G proteins are of the Gα family. G protein Gα subunits are modified by the addition of prenyl (polysoprenyl) and carboxy-terminal methyl groups, which serve to anchor them to the plasma membrane. All G proteins share the capacity, localized to their α subunits, to bind GTP and subsequently to hydrolyze it to guanosine diphosphate (GDP). G proteins are active when GTP bound, but inactive in the GDP-bound form. Engagement of the appropriate seven-transmembrane-domain receptor results in the binding of GTP on the α subunit. As a consequence of GTP binding, heterotrimeric G proteins dissociate into α and βγ components, each with specific effector functions.

A monomeric class of low-molecular-weight (20 to 25 kD) GTP-binding proteins (LMW-GBPs) has also been described. Because the first prototypic LMW-GBP to be described was the proto-oncogene Ras, these also are referred to as Ras-related or Ras superfamily proteins, or small GTPases. LMW-GBPs combine, in one molecule, the prenyl and methyl modifications of the α protein γ subunit with the GTP-binding capacity of the α subunit. At least four families of LMW-GBP have been described: the Ras family, whose members play roles in activation as well as cell growth and division; the Rho family, which functions in cytoskeletal rearrangements; and the Rab and Arf families, which are crucial for vesicular and endomembrane trafficking. All four classes of LMW-GBPs are represented in neutrophils.

**Second Messengers**

Second messengers are small, diffusible molecules that are generated in response to stimuli and transmit signals from membrane receptors to downstream effector proteins. In the classic model of neutrophil activation, engagement of receptors results in the activation of phospholipase C, which cleaves phosphatidylinositol triphosphate (PIP3) into diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP3). DAG and IP3 mediate the influx of cytosolic calcium and the activation of protein kinase C (PKC). Other phospholipases present in the neutrophil include phospholipase A2 (cPLA2s), which cleaves phosphatidylcholine or ethanolamine, or both, and is responsible for the generation of arachidonic acid (AA) and phospholipase D, which cleaves phosphatidylcholine into phosphatidic acid and choline. Although all the previously described second messengers have been implicated in neutrophil activation, other lipid mediators may have negative regulatory effects. Sphingosine and ceramide each inhibits neutrophil phagocytosis.

In addition to lipids, other organic and inorganic messenger molecules have been characterized. Intracellular concentrations of cyclic adenosine monophosphate (cAMP), a classic second messenger, increase rapidly in neutrophils exposed to stimuli and inhibitors. cAMP in these settings is likely to provide a negative regulatory (off) signal because direct exposure to cAMP inhibits neutrophil responses, probably through the activation of protein kinase A (PKA). In contrast, increases in cyclic guanosine monophosphate (cGMP) have a modest enhancing effect on some neutrophil responses. Nitric oxide (NO), an important molecule in the regulation of host defense, is also produced in neutrophils, albeit in low levels. Endogenously produced NO in neutrophils is likely to play an important role in signal transduction; several studies have documented the capacity of exogenously added NO to exert a variety of effects including inhibition of reduced nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, actin polymerization, and chemotaxis (see later). Excessive NO production has been implicated in many rheumatic diseases.

**Kinases and Kinase Cascades**

There has been explosive growth in understanding of how kinases—proteins capable of enzymatically adding phosphate groups to target molecules—contribute to signaling in myeloid and nonmyeloid cells. Protein kinase C (PKC), actually a family of kinases, was among the first kinases implicated in neutrophil activation and is activated in response to chemoattractants. The ability of phorbol myristate acetate (PMA), a synthetic activator of PKC, to stimulate neutrophil responses including adhesion and O2− generation supports a role for PKC in neutrophil
Neutrophil Adhesion

One of the earliest, crucial aspects of the inflammatory response is the ability of bloodstream neutrophils to adhere to vascular endothelium preparatory to movement into the tissues (Figure 11-4). Stimulated neutrophils also possess the ability to adhere to each other, a process termed homotypic aggregation, which may bring in vivo bloodstream neutrophils into proximity with neutrophils already adherent to the vessel or concentrate them at a site of inflammation. Extensive investigation has provided significant insight into the mechanisms involved in neutrophil adhesion. Several families of interacting adhesion molecules have been shown to exist on neutrophils and endothelial cells including the selectins, integrins, intercellular adhesion molecules (ICAMs), and sialylated glycoproteins.

The selectins consist of three related molecules (L-selectin on leukocytes, E-selectin on endothelial cells, and P-selectin on activated platelets and endothelial cells) sharing a common structure of two or more complement regulatory domains, an epidermal growth factor–like domain, and a lectin domain. Each binds to a sialylated glycoprotein on the surface of its interacting cell: E-selectin binds to the sialyl Lewis^x^ antigen on neutrophils, P-selectin binds P-selectin glycoprotein-1 on neutrophils, and L-selectin binds P-selectin glycoprotein-1 and GlyCAM-1 on the endothelium. Selectin expression is largely constitutive, but selectin/sialylated glycoprotein interactions are of low affinity and transient. The result of these interactions is that a pool of bloodstream neutrophils is, at any one time, loosely marginated to the vascular surface and moving along it slowly in a rolling, tumbleweed-like motion. Exposure of neutrophils and endothelium to appropriate stimuli (e.g., adrenergic discharge, corticosteroids) leads to shedding of selectins and neutrophil release (stress demargination), with apparent increases in the peripheral neutrophil count.

The integrins are a large family of heterodimeric molecules generated by various combinations of α and β chains. Similar to the selectins, they require divalent cations (Ca^2+^ or Mg^2+^ or both) to engage their ligands. Neutrophils express three β^2^-type integrins, each constructed from a distinct α subcomponent (CD11a, CD11b, or CD11c) and a common β^2^ chain (CD18). Integrins use the ICAMs as their counterligands. CD11b/CD18 (also called Mac-1 or CR3) binds to fibrinogen, factor X, heparin, and the complement component iC3b in addition to ICAM and is most strongly implicated in neutrophil/endothelial and neutrophil/neutrophil interactions. In contrast to the selectins, neutrophil CD11b/CD18 is constitutively expressed but inactive; stimulation of neutrophils by chemotactic agents results in changes in the activation state of CD11b/CD18 and increases its affinity for ICAMs and other ligands. Stimulation of endothelium with cytokines such as IL-1 results in increased expression of ICAM-1 and ICAM-2, providing a coordinate mechanism for the regulation of adhesion. In contrast to selectin-mediated adhesion, integrin/ICAM interactions are high-affinity and persistent. Stimulation of rolling neutrophils results in their tight adhesion to vessel walls and constitutes the first committed step in the movement of neutrophils into the tissues. Additionally, engagement of integrins by their counterligands sends signals into the cell ("outside-in" signaling), regulating selective cell responses such as cytoskeletal reorganization, oxidant production, and degranulation. Outside-in signaling through CD11b/CD18 also coordinates with signaling through Fc receptor FcγRIII (see later) to mediate phagocytosis of particles opsonized by IgG and the complement component iC3b. Crosstalk between neutrophils and endothelial cells has also been shown to be a CD11b/CD18-dependent event: cross-linking of CD18 on neutrophils leads to increased permeability of endothelial cells, probably through the release of neutrophil proteases.

The function of CD11a/CD18 on neutrophils has been controversial, but accumulating data indicate that this
proteases or oxygen radicals. The disruptions are rapidly repaired by an unknown mechanism, probably involving the endothelium.

Chemotaxis in the direction of a gradient is achieved by the extension of membrane ruffles (lamellipodia), followed by anchorage of the ruffles to the substrate and withdrawal of the trailing edge of the cell in the direction of movement. These changes are accomplished primarily through rearrangement of the actin cytoskeleton after engaging chemotactic factor gradient that signals through GPCRs and PI3K. Actin is a 41-kD protein that exists as a soluble, globular monomeric form (G-actin) and as an insoluble linear polymer (F-actin). F-actin may be assembled (extended) at one end (barbed end) and disassembled at the other, under the control of regulatory molecules. During chemotaxis, F-actin formation and extension is concentrated at the leading edge of the neutrophil, permitting extension of the cell membrane (see Figure 11-5). Chemoattractant receptors also concentrate at the leading edge, defining the cell’s directional response to the gradient (headlight phenomenon). As the neutrophil moves along, receptors that were formerly at the leading edge are swept to the tail and internalized.

**Phagocytosis and Degranulation**

Neutrophil phagocytosis of an encountered bacterium or other particle requires direct contact. Neutrophils are generally poor at phagocytosing unmodified targets, particularly encapsulated bacteria (Figure 11-6). Two mechanisms
for triggering phagocytosis have been defined. First, phagocytosis can be activated by direct neutrophil recognition of pathogen-associated molecular patterns, or PAMPs, which are small molecule motifs found on or in bacteria or viruses (but not typically mammalian cells). PAMPs are recognized by Toll-like receptors (TLRs) and other pattern recognition receptors (PRRs). Human neutrophils express all TLRs except for TLR3, and TLR activation stimulates human neutrophil phagocytosis by promoting structural and conformational changes. Second, phagocytosis also depends on opsonization (from the Greek, “to prepare for the table”), the modification of a target via its decoration with immunoglobulin or complement components or both. Neutrophils express two families of receptors for the Fc portion of complexed or aggregated IgG: low-affinity FcγRIIa and high-affinity FcγRIIIb. During some infections, or after in vitro stimulation with interferon or G-CSF, neutrophils also express the high-affinity receptor FcγRI, which binds monomeric IgG.

FcγRIIa binds subclasses of IgG with varying efficiency depending on a polymorphism at amino acid position 131. Confusingly, the “low-responder” allotype (so named because of its weak interaction with mouse IgG1) binds human IgG3 efficiently, whereas the “high-responder” allotype (which binds mouse IgG1 efficiently) does not. FcγRIIIb polymorphisms of neutrophil antigens NA1 and NA2 also determine binding to IgG subclasses. Individuals homozygous for the NA2 allele have a lower capacity to mediate phagocytosis than individuals homozygous for the NA1 allele. These differences have important implications for rheumatic diseases in which immune complexes play an important role (see later).

Phagocytosis is an active process that involves extension of the neutrophil membrane (filopodia and lamellipodia formation) and invagination of the neutrophil at the locus of the target. Engagement of FcγR and complement receptors results in the activation of diverse signaling pathways. Elegant studies by Caron and Hall indicate that FcγR and complement receptors play distinct roles in phagocytosis. Whereas engagement of CR3 (CD11b/CD18) results in actin stress fiber formation and invagination, engagement of FcγRII results primarily in extension of membranes out from and around the target. Signaling by these receptors depends on the activation of distinct members of the Rho family of LMW-GBPs. These observations remain to be specifically confirmed for neutrophils, however. On activation, neutrophils degranulate, a term actually reflecting two distinct processes. Vesicles can fuse with the plasma membrane, spilling their contents into the extracellular space (see Figure 11-1), or they can fuse with the phagocytic vacuole to form a phagolysosome. The former type of degranulation is regulated differentially from the latter and favors mobilization of lighter granules in response to stimuli (secretory vesicles > gelatinase granules > specific granules > azurophilic granules). In the latter type of degranulation (phagolysosome formation), fusion of azurophilic granules with the phagocytic vacuole results in the delivery of proteolytic enzymes, myeloperoxidase, and antibacterial proteins to the site of the ingested bacterium. Fusion of specific granules with the phagocytic vacuole permits the delivery of collagenase and the appropriate localization of cytochrome b558, a requisite for NADPH oxidase (see later). Containment of potentially toxic substrates within the phagolysosome keeps host tissue damage and neutrophil autodestruction in check. As discussed subsequently, however, in several of the rheumatic diseases, neutrophilic activation plays an important role in abetting inflammation and host tissue damage.

Reduced Nicotinamide Adenine Dinucleotide Phosphate Oxidase System

In addition to the collection of proteases and other antibacterial proteins contained in their granules, neutrophils have the capacity to kill bacteria through the generation of toxic oxygen metabolites such as nitric oxide (NO), superoxide anion (O2−) and hydrogen peroxide (H2O2). This process, frequently referred to as the respiratory burst, is extremely potent and requires tight regulation to prevent neutrophil autodestruction. Studies in cell-free systems have established the so-called minimal system required for O2−-generation-NADPH oxidase. The central component of NADPH oxidase is cytochrome b558, which is localized to the membranes of specific granules and consists of two subunits: a 22-kD component (gp22rac, for phagocyte oxidase) and a 91-kD component (gp91phox). This cytochrome lacks independent activity, however. Three cytosolic proteins are also required: a 47-kD and a 67-kD component (p47phox and p67phox) and a LMW-GBP, p21. On neutrophil stimulation, the p47phox and p67phox components translocate to the membranes to form an active complex with the cytochrome. Although p21 also translocates in response to stimuli, the significance of its translocation is more controversial. A fifth protein, p40phox, also has been reported to be associated with p47/p67 in the cytosol. Evidence suggests that p40phox may regulate the oxidase system in a positive and a negative manner (Figure 11-7) and plays an important role in phagocytosis-induced superoxide production via a phox homology (PX) domain that binds to PIP2. Moreover, autosomal recessive mutations in NCF4, the gene encoding p40phox, have recently been associated with a form of chronic granulomatous disease (see Heritable Disorders of Neutrophil Function later).

When assembled and activated, the NADPH oxidase transfers electrons from NADPH to generate O2−:

\[ 2O_2 + \text{NADPH} \xrightarrow{\text{NADPH Oxidase}} O_2^- + \text{NADP}^+ + \text{H}^+ \]

A subsequent, spontaneous dismutase reaction rapidly produces hydrogen peroxide:

\[ O_2^- + 2H^+ \rightarrow H_2O_2 + O_2 \]

Although O2− and H2O2 can kill organisms in vitro, they are short-lived and probably do not account for most of the bacterial killing capacity of the system under normal circumstances. (Many bacteria possess catalase, an enzyme that degrades H2O2.) Rather, the production of H2O2 within the same space into which the myeloperoxidase has been released permits the generation of large quantities of hypochlorous acid (chlorine bleach), a powerful oxidant with potent killing capacity. Hypochlorous acid may interact further with proteins to form chloramines, less potent but longer-lived oxidants. Neutrophil oxidant production plays
Nonphagocytic Bacterial Killing

Novel distinctive mechanisms of bacterial killing by neutrophils can also augment host defense. Neutrophil extracellular traps, or NETs, are extracellular fibers composed of granule proteins and chromatin that bind and kill microorganisms (Figure 11-8).\textsuperscript{42} NETs are released on cellular activation. They entrap bacteria while simultaneously providing a scaffold to promote high local concentrations of antimicrobial components, thus killing microbes extracellularly. Because neutrophils die immediately after activating their NETs, this process has also been termed beneficial suicide.\textsuperscript{43} Interestingly, these same NETs or related structures have recently been suggested to play roles in promoting clotting (including disseminated intravascular coagulation). In those cases, extruded neutrophil chromatin serves as a platform for the extracellular co-localization of neutrophil elastase and the antithrombotic tissue factor pathway inhibitor (TFPI). Inactivation of TFPI by elastase then permits clotting to proceed.\textsuperscript{44}

Neutrophil Production of Proinflammatory Mediators

Arachidonic Acid Metabolites

The capacity of stimulated neutrophils to liberate arachidonic acid from membranes has implications for the propagation of acute inflammation. Although arachidonic acid itself has chemoattractant and neutrophil-stimulatory properties,\textsuperscript{25,45,46} its metabolites are more crucial to regulation of inflammation. Best recognized among these are the leukotrienes. Neutrophils have the capacity to produce LTB\textsubscript{4}, a highly potent lipid mediator for the chemoattraction of other neutrophils. Intermediates of leukotriene production such as 5-hydroxyeicosatetraenoic acid also are produced by neutrophils and may have stimulatory properties.\textsuperscript{25}
An alternative class of lipoxygenase products, the lipoxins, have been characterized. Synthesis of lipoxins requires coordinated activity of neutrophil 5-lipoxygenase and a related enzyme (either 12-lipoxygenase or 15-lipoxygenase) in another cell type—either platelets or endothelial cells (Figure 11-9). In contrast to leukotrienes, lipoxins inhibit neutrophil function and are anti-inflammatory, suggesting that the assembly of a mixed population of inflammatory cells may trigger the synthesis of anti-inflammatory molecules (resolvins), contributing to the subsequent resolution of inflammation (see section on Resolution of Neutrophil Infiltration and Apoptosis).

The cyclooxygenase (COX) (endoxygenase synthase) pathway is the other major pathway of arachidonic acid metabolism. Arachidonic acid metabolized by COX is converted into prostaglandin H, which undergoes further cell type–specific conversion to a variety of other prostaglandins. The prostaglandins of most relevance to inflammation are those of the E series, particularly prostaglandin E2. Prostaglandins of the E series have numerous proinflammatory effects including increased vasodilation, vascular permeability, and pain. However, the direct effects of prostaglandin E on neutrophils seem to be inhibitory, probably through elevations of intracellular cAMP. Although resting neutrophils exhibit little COX activity, persistent neutrophil activation results in upregulation of COX-2, suggesting that neutrophils may contribute prostaglandin E2 to both the proinflammatory brew and the downregulation of their own activity.

**Cytokine Production**

Although the relative amount of cytokine production by neutrophils is small, the large numbers of neutrophils present in infected or inflammatory sites suggest that overall neutrophil cytokine production may play a role in recruiting additional neutrophils to the target area. Among the cytokines produced by neutrophils are IL-8, IL-12, MIP-1α and β (CCL3 and CCL4), growth-related oncogene-α (GRO-α), oncostatin M, MCP-1, and TGF-β. Neutrophils do not produce IL-1, TNF, and IL-6, the classical products of macrophages and synovial cells. The transcriptional program of terminally differentiated neutrophils requires a selective combination of stimulants for the production of chemokines. In the presence of lipopolysaccharide, TNF drives production of IL-8, GRO-α, and MIP-1, whereas IFN-γ is necessary to generate CXCL9 and 10. Other neutrophil-derived molecules have been identified as bridging factors between innate and adaptive immunity. Following either G-CSF or IL-8 stimulation, activated neutrophils release both B-lymphocyte stimulator (BlyS) and TNF-related apoptosis-inducing ligand (TRAIL). While BlyS stimulates B cell proliferation (through the TNF receptors BAFF-R, TACI, and BCMA), TRAIL induces antitumor T cell effects and apoptosis.

More recently, multiple lines of investigation have suggested that neutrophils may also be a source for IL-17, a potent proinflammatory cytokine, which can amplify the neutrophil migration and recruitment. It is still unknown, however, whether neutrophils express IL-23R (required for most IL-17 producing cells) or can induce transcriptional factors that regulate IL-17 production (such as RORγ or STAT3).

TGF-β is a powerful neutrophil chemoattractant, and its recruitment of neutrophils into an inflammatory space may lead to additional neutrophil cytokine production, including further production of TGF-β. TGF-β also has potent anti-inflammatory effects, however, suggesting that neutrophils may also participate in the resolution of inflammation. Activated neutrophils also produce an antagonist to IL-1, the IL-1 receptor antagonist. The efficacy of recombinant IL-1 receptor antagonist (anakinra) in the treatment of rheumatoid arthritis and in autoinflammatory diseases such as Still’s disease emphasizes its clinical importance in downregulating synovial inflammation.

**Resolution of Neutrophil Infiltration and Apoptosis**

Inflammatory responses must eventually be resolved to avoid excessive tissue damage and initiate the healing process. Resolution of inflammation is an active and carefully regulated process. Proresolution signals include lipid mediators such as lipoxins and resolvins, annexin A1 and chemerin-derived peptides, and certain chemokines and cytokines.

Two types of resolvins are described, depending on the lipid from which they are derived. E-resolvin is a product of eicosapentaenoic acid (EPA), whereas D-resolvin derives from docosahexaenoic acid (DHA). Resolvin D1, for instance, acts on monocytes, macrophages and dendritic cells (DCs), attenuating TNF-mediated nuclear factor

---

**Figure 11-9** Generation of the anti-inflammatory lipoxins A4 and B4 depends on the interaction between two different classes of inflammatory cells. Top, Lipoxin generation by neutrophils and platelets. Arachidonic acid (AA) generated by activated neutrophils is converted by neutrophil 5-lipoxygenase (5-LO) into leukotriene A4 (LTA4), LTA4 may be converted by 12-LO in nearby platelets into lipoxin A4 (LXA4) and lipoxin B4 (LXB4). Bottom, Lipoxin generation by epithelial cells and neutrophils. AA generated by epithelial cells may be converted by 15-LO into 15-hydroxyeicosatetraenoic acid (15-HETE). In the setting of inflammation, 5-LO from adjacent neutrophils subsequently may convert 15-HETE into LXA4 and LXB4.
kappa B (NFκB) activation, leading to an active anti-inflammatory pathway. At the same time, Resolvin E1 selectively interacts with the selective LTβ receptor BLT1 on neutrophils to regulate inflammation, suggesting a tightly coordinated feedback loop. Resolvin D is also a potent inhibitor of neutrophil diapedesis and migration, particularly in the brain, where DHA is highly abundant. The most recently described resolvin, macrophage mediator in resolving inflammation 1 (maresin 1), has properties similar to the D-resolvins. Aspirin has been shown to stimulate the generation of biologically active epilipoxins, suggesting a previously unappreciated mechanism for its anti-inflammatory action. Even some prostaglandins such as 15d-PGJ2 also have anti-inflammatory properties.

Following neutrophil activation, annexin A1 (lipocortin) is released in response to chemoattractants, downregulating transmigration and promoting neutrophil apoptosis and clearance. Similar activities have recently been attributed to chemerin-derived peptides. Apoptotic neutrophils at sites of inflammation are taken up by resident macrophages, promoting lipoxin A4 production, which inhibits polymorphonuclear neutrophil migration and promotes cell arrest while decreasing neutrophil activity and cytokine production. An MMP-mediated mechanism of inflammatory resolution has also been identified. Macrophage-derived MMPs such as MMP-1, MMP-3, and MMP-12 cleave several CXC-chemokines, provoking the loss of their neutrophil-recruiting activity and dampening the influx of cells.

Aptoptotic neutrophils themselves can prevent further chemotaxis and migration through negative feedback loops. Dead neutrophils inhibit migration of granulocytes via release of lactoferrin and annexin A1. Apoptotic neutrophils can also suppress granulopoiesis, by inhibiting the proinflammatory consequences of IL-17/IL-23 axis activation. In this model, macrophage and DCs produce IL-23 at the site of insult, which in turn promotes IL-17 production by T cells (Th17, γδ T cells, and natural killer [NK] T cells). IL-17 is a promoter of G-CSF production and a powerful neutrophil chemoattractant. Recruited neutrophils undergo apoptosis and are phagocytosed by macrophages, resulting in a decrease of IL-23. This is followed by a reduction in IL-17 and G-CSF production, which downregulates granulopoiesis. The role of SDF-1/CXCR4 signaling system in neutrophil homeostasis has been described earlier (see Neutrophil Myelopoiesis and Clearance).

An intriguing mechanism through which neutrophils might downregulate the action of protein inflammatory mediators has been defined. Apoptotic neutrophils (present during the resolving phase of inflammation) show increased expression of the chemokine receptor CCR5 on their surface (mediated by D and E resolvins), and this receptor can scavenge, and reduce the soluble concentration of, chemokines such as CCL3 and CCL5. These data emphasize again that neutrophils are not only inflammatory cells but may also play a direct role in the subsequent resolution of inflammation.

Heritable Disorders of Neutrophil Function

A wide variety of acquired conditions result in neutrophil dysfunction, depletion, or both including malignancies (myeloid leukemias), metabolic abnormalities (diabetes), and drugs (corticosteroids, chemotherapy). In addition, many rare, congenital disorders of neutrophils have been identified (Table 11-2). In general, patients with impaired neutrophil function are prone to infection by bacteria (predominantly Staphylococcus aureus, Pseudomonas species, Burkholderia) and fungi (Aspergillus, Candida), but not viruses and parasites. The major sites of infection include skin, mucous membranes, and lungs, but any site may be affected and spreading abscesses are common. Most of these diseases are potentially life threatening in the absence of available effective therapy.

Diseases of Diminished Neutrophil Number

Severe congenital neutropenia (SCN, Kostmann’s syndrome) results from marrow arrest of bone marrow myelopoiesis and leads to neutrophil counts persistently less than 0.5 × 10⁹ cells/L. Monogenic autosomal dominant, autosomal recessive, and sporadic subtypes were initially identified. However, some SCN patients have recently been identified as carriers of mutations in multiple genes. Patients are prone from early infancy to severe bacterial infections including emphysema, pneumonia, otitis, gingivitis, and perirectal infections. Because acute inflammation is lacking, infections tend to spread extensively before coming to attention. Mortality has been high. Therapy consists of antibiotics and long-term therapy with recombinant human G-CSF, which may help maintain normal or near-normal neutrophil counts. SCN patients are also at risk for acute myeloid leukemia (AML) and myelodysplastic syndrome, particularly in patients who do not respond well to G-CSF therapy. A milder form of neutropenia (benign congenital neutropenia) with higher neutrophil counts and fewer infections has also been observed. Another variant is cyclic neutropenia, which causes transient, recurrent neutropenia on a 21-day cycle. Studies suggest that defects in neutrophil elastase affect neutrophil survival in the marrow and may be responsible for severe congenital and cyclic neutropenia. At least 52 different mutations in the ELANE gene, encoding for neutrophil elastase, have been described as the cause in about half of the patients. Mutations in HAX-1 and Glucose-6-phosphatase catalytic subunit 3 (G6P C3) genes account for a smaller proportion of SCN. Other, less frequent deficiencies include the X-linked neutropenia—caused by constitutively active mutations of the WASP gene—and defects in several myeloid transcription factors.

Leukocyte Adhesion Deficiencies

Leukocyte adhesion disorders arise from defects in cell adhesion to extracellular matrix and vascular endothelium. Three distinct entities have been described in humans. Leukocyte adhesion deficiency type 1 (LAD1) results from an autosomal recessive defect in ITGB2, encoding for the CD18 chain of β2 integrins. Consequently, neutrophil β2 integrins fail to form, and bloodstream neutrophils are unable to adhere firmly to vascular endothelium and to transmigrate to sites of infection. Phagocytosis is also impaired. The clinical picture is similar to that of the neutropenias, with recurrent life-threatening infections. Peripher-
### Table 11-2  Heritable Disorders of Neutrophil Function

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Defect</th>
<th>Inheritance</th>
<th>Presentation</th>
<th>Therapy</th>
<th>Typical Prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutropenia</td>
<td>Maturation arrest (&lt;0.5 × 10⁹ PMN/L)</td>
<td>AR (HAX1 mutations)</td>
<td>Bacterial infections (omphalitis, abscesses, gingivitis, UTIs)</td>
<td>RhG-CSF</td>
<td>Improved with treatment</td>
</tr>
<tr>
<td>Severe congenital neutropenia (Kostmann’s syndrome)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benign congenital neutropenia</td>
<td>Multiple etiologies (0.2-2 × 10⁹ PMN/L)</td>
<td>Variable</td>
<td>Mild infections</td>
<td>None</td>
<td>Good</td>
</tr>
<tr>
<td>Cyclic neutropenia</td>
<td>Stem cell defect, elastase gene deficiency (nadir every 21 days)</td>
<td>AD (ELA2 mutations)</td>
<td>Infection during nadirs</td>
<td>RhG-CSF</td>
<td>Improved with treatment</td>
</tr>
</tbody>
</table>

### Adhesion Deficiency

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Defect</th>
<th>Inheritance</th>
<th>Presentation</th>
<th>Therapy</th>
<th>Typical Prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leukocyte adhesion deficiency type 1</td>
<td>Absent or abnormal CD18; deficiency in β2-integrin chain of leukocyte adhesion molecules</td>
<td>AR</td>
<td>Leukocytosis; recurrent infections (skin mucous, membranes, gastrointestinal tract)</td>
<td>Marrow transplant</td>
<td>Fair-poor</td>
</tr>
<tr>
<td>Leukocyte adhesion deficiency type 2</td>
<td>Absent sialyl-Lewis*</td>
<td>AR</td>
<td>Neutrophilia; infection; retardation, short stature</td>
<td>Poor</td>
<td></td>
</tr>
<tr>
<td>Leukocyte adhesion deficiency type 3</td>
<td>Impaired activation of Rap1 GTPase</td>
<td>AR</td>
<td>Leukocytosis; recurrent infections; bleeding tendency</td>
<td>Poor</td>
<td></td>
</tr>
</tbody>
</table>

### Chemotaxis Deficiency

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Defect</th>
<th>Inheritance</th>
<th>Presentation</th>
<th>Therapy</th>
<th>Typical Prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyper-IgE syndrome</td>
<td>Chemotaxis defect</td>
<td>AD</td>
<td>Eczema; recurrent infections; elevated serum IgE levels</td>
<td>Skin care; antibiotics</td>
<td>Good</td>
</tr>
</tbody>
</table>

### Granule Disorders

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Defect</th>
<th>Inheritance</th>
<th>Presentation</th>
<th>Therapy</th>
<th>Typical Prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chédiak-Higashi syndrome</td>
<td>Defective lysosomal trafficking regulator gene</td>
<td>AR</td>
<td>Albinism; infection</td>
<td>Marrow transplant; antibiotics</td>
<td>Poor</td>
</tr>
<tr>
<td>Specific granule deficiency</td>
<td>Abnormal/reduced specific and azurophilic granules (lactoferrin deficiency)</td>
<td>AR?</td>
<td>Infection of skin, mucous membranes, lungs</td>
<td>Poor Fair-good</td>
<td></td>
</tr>
<tr>
<td>Myeloperoxidase deficiency</td>
<td>Myeloperoxidase absent</td>
<td>Variable (mostly AR)</td>
<td>None</td>
<td>Transfusion of HLA identical leukocytes if severe</td>
<td>Excellent</td>
</tr>
<tr>
<td>p14 deficiency</td>
<td>Defective endosomal adapter protein gene</td>
<td>Recessive</td>
<td>Albinism; infection; short stature</td>
<td>None known to date</td>
<td>?</td>
</tr>
</tbody>
</table>

### Oxidase Defects

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Defect</th>
<th>Inheritance</th>
<th>Presentation</th>
<th>Therapy</th>
<th>Typical Prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic granulomatous disease (multiple types)</td>
<td>gp91(phox) absent</td>
<td>X-linked 50%</td>
<td>Early childhood infections, especially skin and mucous membranes, abscesses</td>
<td>Interferon-γ</td>
<td>Improved with treatment</td>
</tr>
<tr>
<td></td>
<td>p22(phox) absent</td>
<td>AR 5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p47(phox) absent</td>
<td>AR 35%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p67(phox) absent</td>
<td>AR 5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p40(phox) absent</td>
<td>AR 5%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AR, autosomal, recessive; PMN, polymorphonuclear neutrophil; RhG-CSF, recombinant human granulocyte colony-stimulating factor; UTIs, urinary tract infections.

Reflecting the intravascular accumulation of cells in the face of their inability to exit the vessels. Complete LAD1 manifests in infancy and is characterized by omphalitis, recurrent life-threatening bacterial and fungal infections, gingivitis, and delayed wound healing. Absence of pus at sites of infection is a hallmark of LAD1. Bone marrow transplantation is the only curative therapy. LAD2 results from an autosomal recessive defect in the glycosylation of sialyl Lewis’ (SLex or CD15s), the neutrophil counterligand for endothelial selectins. Patients with LAD2 have neutrophils that are unable to roll along the endothelium and have symptoms similar to LAD1 but may also have mental retardation, short stature, distinctive facies, and the Bombay (hh) blood type. A third variant, LAD3, has been described in which leukocytes have normal populations of surface integrins but lack the ability to signal these molecules into an active
state.79 Because integrin activation in this condition is also deficient in platelets, patients with LAD3 are at increased risk for infection and bleeding.

**Neutrophil Granule Defects**

The best-known defect in neutrophil granule formation is Chédiak-Higashi syndrome. Chédiak-Higashi syndrome is an autosomal recessive disorder in which granule subtypes—in neutrophils, but also in lymphocytes, melanocytes, Schwann cells, and others—undergo disordered fusion, resulting in giant, dysfunctional granules. The cause seems to relate to a defect in the gene for lysosomal transport protein (Lyst or CHS1).80 Patients with Chédiak-Higashi syndrome present with partial oculocutaneous albinism, neutropenia, frequent infection, mild bleeding diathesis, and neurologic abnormalities. Approximately 85% of patients who survive childhood enter a so-called accelerated phase, a lymphoma-like infiltration of lymphocytes and histiocytes throughout the body, which is generally fatal. Other diseases of neutrophil granules have less ominous prognoses. A novel immunodeficiency syndrome relating to lack of the endosomal adapter protein p14 (encoded by the ROBLD3 gene) has been described. Patients with this syndrome have congenital neutropenia with structurally abnormal neutrophil primary granules and abnormalities of B cells, cytotoxic T cells, and melanocytes. In addition to immunodeficiency, clinical findings include short stature and partial albinism.81

**Oxidase Deficiencies—Chronic Granulomatous Disease**

Chronic granulomatous disease (CGD) resembles other diseases of neutrophil dysfunction in that it results in severe, recurrent infections of the skin and mucous membranes. Osteomyelitis and intra-abdominal abscesses are common. In contrast to the other diseases, infection in patients with CGD generally results in a delayed but quantitatively normal neutrophil response. Because of incapacity to kill organisms, however, the accumulation of neutrophils at a site of infection generally results in granuloma formation, rather than clearance of the target. Cutaneous infections tend to show persistent drainage and scarring. The presence of even partially responsive neutrophils results in a lower frequency of sepsis in these patients relative to patients with absolute neutropenia. CGD is typically a disease of early childhood, although some milder cases may be recognized later in life.

CGD is actually a group of diseases. In each, a genetic defect in a different component of NADPH oxidase results in failure of neutrophils (and other phagocytes) to generate \( \text{O}_2^- \), impairing both intracellular killing and the ability of neutrophils to generate NETS.82,83 X-linked CGD affecting gp91 \( \text{phox} \) (CYBB gene) is the most common form and accounts for about 70% of cases, whereas the autosomal recessive forms are responsible for the rest.83 Treatment for chronic granulomatous disease consists of aggressive antibiotic prophylaxis and therapy, along with long-term therapy with recombinant human interferon-\( \gamma \) (IFN-\( \gamma \)) to improve neutrophil function. Clinical trials of gene therapy of X-linked CGD have yielded encouraging results.83

**Defects of TLR Signaling**

Defects in human TLR signaling including IRAK-4 and MyD88 deficiencies lead to impaired neutrophil function and increased susceptibility to bacterial infections.84,85

**Neutrophil Relevance to Rheumatic Disease**

**Neutrophil-Mediated Tissue Destruction**

Despite sophisticated regulatory mechanisms, tissue destruction by neutrophils is common. Several mechanisms may permit the release of neutrophil proteases and oxygen radicals into the extracellular milieu. First, necrosis or destruction of neutrophils, or both, may liberate cellular contents indiscriminately. Second, studies have revealed that degranulation and \( \text{O}_2^- \) generation may begin before complete closure of the phagocytic vacuole, releasing products either into the external environment (regurgitation during feeding) or against a target surface. Neutrophils may destroy host tissues that have been misdirected to attack. Although serum and joint fluid contain antiproteases and antioxidants, the “protected space” between a neutrophil and a surface (e.g., cartilage) may exclude these factors. Release of hypochlorous acid, myeloperoxidase, and proteases into the extracellular milieu may inactivate the protective compounds and act as an “antiproteinase shield.”86

**Neutrophil Fc Receptor Polymorphisms**

Given that polymorphisms of Fc\( \gamma \)R determine phagocytic capacity of IgG isotypes, it is not surprising that they determine susceptibility to diseases in which autoantibodies play a key role. Phagocytes from individuals with one Fc\( \gamma \)RIIa polymorphism (H131) allele are able to bind and phagocyte IgG; phagocytes from individuals with a different polymorphism (R131) cannot. In white European and African-American populations, patients with lupus nephritis have a higher frequency of the Fc\( \gamma \)RIIa-R131 allele than control groups; their relative inability to clear immune complexes may make them more susceptible to renal disease.86 The significance of different Fc receptor polymorphisms may vary among rheumatologic diseases. Tse and colleagues found no association between Fc\( \gamma \)RIIa polymorphisms and likelihood of developing antineutrophil cytoplasmic antibody (ANCA)-associated vasculitides, but they found an overrepresentation of homozygosity for the Fc\( \gamma \)RIIb NA1 allele among patients with antimiylopeneroxidase antibodies. Few studies have been published on Fc polymorphisms and susceptibility or severity of rheumatoid arthritis. More recent studies found no association among Fc\( \gamma \)RIIa, IIb, or IIb haplotypes and rheumatoid arthritis susceptibility but did find a correlation between patients with extra-articular disease and the homozygous R/R 131 genotype. In contrast, Fc\( \gamma \)RIIIa receptor expression is increased in rheumatoid arthritis.87

**Gout**

Gout may be the quintessential neutrophilic rheumatic disease. Although the initiation of an acute gouty attack involves the phagocytosis of urate crystals by synovial
macrophages and the generation of cytokines such as IL-1 and IL-8, the hallmark of acute gout is the presence of enormous numbers of neutrophils (sometimes > 100,000/mm³) in the affected joint space. Urate crystals in the joint are capable of nonspecifically binding immunoglobulin and fixing complement by the classic and alternative pathways. C5a liberated from the complement fixation process attracts neutrophils to the joint space, where they phagocytose opsonized crystals via receptor-dependent mechanisms, resulting in further activation of the neutrophil and production of LTB₄, IL-8, and other mediators. Naked urate crystals can also directly activate neutrophils. Activation of neutrophils results in the ingress of additional neutrophils.

Neutrophils in the gouty joint may damage joint structures through discharge of contents directly into the joint fluid during crystal phagocytosis or directly against cartilage during attempted phagocytosis of urate crystals embedded in or adherent to cartilage. In addition, interaction of phagocytosed urate crystals with lysosomal membranes results in the dissolution of the latter, spilling lysosomal proteases into the cytoplasm and, eventually, into the extracellular space.¹⁸

**Rheumatoid Arthritis**

Rheumatoid arthritis may be conceptualized as a two-compartment inflammatory disease: In the synovium, lymphocytes, fibroblasts, and macrophages predominate, but the joint space contains a substantial percentage of neutrophils. Although the numbers of neutrophils in the rheumatoid joint tend to be less than those seen in gout, they are still quite large, with 10 billion cells per day cycling through a 30-mL effusion. The classic model suggests that rheumatoid factor–based immune complexes, produced in the pannus and present in the joint space in high concentrations, can fix complement and draw neutrophils into the joint space in high numbers. Once there, in vitro studies have documented the ability of neutrophils to bind to cartilage surfaces embedded with immune complexes and to damage them via incomplete phagocytosis. No adequate in situ demonstration of direct neutrophil attack on cartilage has been offered yet, however. The fact that seronegative arthritides such as psoriatic arthritis lack rheumatoid factor but nonetheless share with rheumatoid arthritis the presence of synovial hyperplasia and a large neutrophilic infiltrate in the joint space suggests that immune complex formation may be important, but not absolutely required, for neutrophil influx. The ability of rheumatoid synovial monocytic leukocytes to secrete IL-1 and IL-8 and other cytokines indicates that pannus itself may play an important role in the attraction of neutrophils out of the bloodstream and into the joint. Although few in number, neutrophils within the pannus have been documented to concentrate at the pannus/cartilage border, suggesting a possible role in pannus-driven marginal erosion.⁹³

In addition to promoting joint destruction, neutrophils in the rheumatoid joint may contribute to the propagation of pannus and of rheumatoid inflammation. As noted earlier, neutrophils themselves can produce proinflammatory cytokines; expression of several such cytokines including oncostatin M, MIP-1α, and IL-8 is increased in rheumatoid neutrophils, especially from rheumatoid arthritis synovial fluid.⁹⁰ Injection of lysates of neutrophil granules into joints in animal models produces a synovitis indistinguishable histologically from rheumatoid synovitis, an effect that can be reproduced by injection with purified active or inactive myeloperoxidase.⁹⁰ Neutrophil proteins also may regulate synovial proliferation through effects on resident cells and immigrant cell populations. Neutrophil protease 3 may enhance the proinflammatory effects of monocyes, by cleaving and releasing active IL-1 and TNF from the surface of the latter. Neutrophil defensins enhance phagocytosis by macrophages and stimulate the activation and degranulation of mast cells, an interesting observation in light of a report that mice deficient in mast cells are resistant to the development of erosive arthritis.⁹¹ Neutrophil proteases also enhance the adherence of rheumatoid synovial fibroblasts to articular cartilage, and neutrophils may regulate synovial vascularization through the production of vascular endothelial growth factor, leading to endothelial proliferation.

Blood and synovial neutrophils from early rheumatoid arthritis patients also show significantly lower levels of apoptosis. This effect might be related to high levels of antiapoptotic cytokines such as IL-2, IL-4, IL-15, GM-CSF, and G-CSF found in joints of early RA patients.⁹² Lactoferrin is present at significant levels in established RA synovium and may delay spontaneous apoptosis of peripheral and synovial neutrophils.⁹³

Recently, studies by Lee and others have demonstrated that neutrophils are necessary for the propagation of rheumatic disease in mouse models of rheumatoid arthritis. These studies have implicated the ability of neutrophils to produce LTB₄, as well as the presence of FcγRIIA and C5a receptors on the neutrophil surface as necessary for arthritis development.⁹⁴,⁹⁵ Intriguingly, several studies have raised the possibility that, under certain conditions of stimulation, neutrophils can serve as antigen-presenting cells. Neutrophils in rheumatoid arthritis synovial fluid synthesize and express large amounts of class II major histocompatibility complex. The importance of neutrophils to the rheumatoid process may be underscored by rheumatoid arthritis animal models in which mice deficient in neutrophils are resistant to the arthritic process.

**Vasculitis**

Neutrophils may be identified, to a greater or lesser degree, in the lesions of virtually all kinds of vasculitis. The mechanisms of neutrophil accumulation may vary, however, with different mechanisms predominating in different conditions. The early observation that infusions of allospecies serum produced acute inflammation in skin and joints (serum sickness), together with the appreciation that subcutaneous rechallenge with previously administered antigen leads to intense local inflammation (Arthus reaction), led to the development of a model in which immune complex deposition in the blood vessels results in complement activation and an influx of neutrophils to the affected site. Because immune complex formation is a hallmark of many primary rheumatic vasculitides (e.g., essential mixed cryoglobulinemia, hypersensitivity vasculitis, Henoch-Schönlein purpura), it is likely that immune complex deposition is crucial to the genesis of these diseases. In several of these vasculitides, neutrophil disruption and
fragmentation—clasis—is a prominent pathologic finding, leading to their designation under the rubric leukocytoclastic vasculitis. In some rheumatic diseases in which vasculitis is a secondary phenomenon, such as rheumatoid arthritis and systemic lupus erythematosus, the role of immune complex deposition is also implicit. It has been suggested that patients with lupus experience transient accumulations of neutrophils (leukoaggregation) in small vessels of the lungs and other tissues, as a result of complement activation within these vessels or in the soluble phase.\(^6\)

Induction of adhesion molecules on endothelial cells or neutrophils themselves or both is an alternative mechanism through which neutrophil accumulation in vessels may be propagated. The Shwartzman phenomenon, in which reinjection of cellular material leads to vascular inflammation via a cytokine-dependent, immune complex–independent mechanism, is a model for this avenue to vasculitis. Adhesion molecule upregulation may be particularly relevant to vasculitides in which immune complex formation is not a hallmark such as giant cell (temporal) arteritis. Detailed analyses of the inflammatory cells involved in giant cell arteritis indicate the presence of T cells producing IL-1β and IL-6 that may act on vascular endothelium.\(^7\) It is likely that many rheumatic diseases employ immune complex–dependent and immune complex–independent mechanisms in the pathogenesis of neutrophil ingress into vascular structures. In addition to the role of immune complexes, Belmont and colleagues\(^8\) have shown the induction of adhesion molecules in patients with systemic lupus erythematosus.

Several vasculitides are noteworthy for the presence, in the serum of affected patients, of antibodies directed at cytoplasmic components of neutrophils (ANCA). ANCA-positive vasculitides are discussed in detail in Chapter 89.

**Neutrophilic Dermatoses and Familial Mediterranean Fever**

Sweet’s syndrome, named after the physician who first described it in 1964, is characterized by fever, neutrophilia, and painful erythematous papules, nodules, and plaques. It can be subdivided into five groups: idiopathic, parainflammatory (associated with inflammatory bowel disease or infection), paraneoplastic (most commonly in the setting of leukemia), pregnancy related, and drug associated (usually after treatment with G-CSF). Most important clinically, it is a diagnosis of exclusion. Sweet’s syndrome frequently appears after an upper respiratory tract infection and has a propensity to involve the face, neck, and upper extremities. When found on the legs, Sweet’s syndrome lesions can appear after an upper respiratory tract infection and has a propensity to involve the face, neck, and upper extremities. It is likely that many rheumatic diseases employ immune complex–dependent and immune complex–independent mechanisms in the pathogenesis of neutrophil ingress into vascular structures. In addition to the role of immune complexes, Belmont and colleagues\(^8\) have shown the induction of adhesion molecules in patients with systemic lupus erythematosus.

Several vasculitides are noteworthy for the presence, in the serum of affected patients, of antibodies directed at cytoplasmic components of neutrophils (ANCA). ANCA-positive vasculitides are discussed in detail in Chapter 89.

**Effects of Antirheumatic Agents on Neutrophil Functions**

Many antirheumatic therapies currently in use have been documented to act at least partly at the level of the neutrophil. Nonsteroidal anti-inflammatory drugs (NSAIDs) are the most frequently used class of antirheumatic agents. By virtue of their ability to inhibit COX activity and prostaglandin production, moderate doses of NSAIDs have diverse effects on inflammation including inhibition of vascular permeability and modulation of pain. At higher, clinically anti-inflammatory concentrations, NSAIDs inhibit chemoattractant-stimulated neutrophil CD11b/CD18-dependent adhesion and degranulation and NADPH oxidase activity.\(^99,100\) It is unlikely, however, that these effects are due solely to COX inhibition because (1) as noted earlier, neutrophils exhibit little COX activity under normal circumstances, and (2) concentrations of NSAIDs required to inhibit neutrophil function exceed the concentrations required to inhibit COX. High-dose NSAIDs seem to have other, pleiotropic effects on neutrophil signaling. Our laboratory has shown the capacity of aspirin and the poor COX inhibitor sodium salicylate to inhibit Erk activation in a manner consistent with inhibition of adhesion, suggesting that salicylates may have unique effects on inflammation.\(^101\)

Similar to nonsteroids, glucocorticoids exert potent effects on neutrophils including inhibition of neutrophil phagocytic activity and adhesive function. The ability of steroids to increase peripheral blood neutrophil populations acutely—an effect known as demargination—is attributable to both a release of neutrophils from the bone marrow and the release (demargination) of neutrophils adherent to vessel walls. In addition, glucocorticoids inhibit phospholipase A\(_2\) and leukotriene and prostaglandin production.
Glucocorticoids may also regulate the expression of COX-2 and stimulate the release of annexin A1. Effects of glucocorticoids on other cells may also reduce neutrophil responses indirectly through the suppression of cytokines at inflammatory sites.

Other anti-inflammatory/immunomodulatory agents also have well-established effects on neutrophils. Methotrexate, widely used in rheumatoid arthritis, has no direct neutrophil effect but is capable of producing indirect effects, probably by virtue of its ability to stimulate the release of adenosine from surrounding cells. Some data suggest that methotrexate-induced adenosine release might inhibit phagocytosis, O2·− production, and adhesion, and that treatment of patients with methotrexate inhibits the capacity of neutrophils to generate LTβ.

Colchicine, a standard agent in the treatment of gout and familial Mediterranean fever, inhibits microtubule formation and has pleiotropic effects on neutrophils including inhibition of adhesion via decrements in selectin expression. Colchicine has been observed to stimulate the expression of pyrin in neutrophils. Because pyrin deficiencies are implicated in familial Mediterranean fever, this observation suggests a previously unappreciated mechanism of action of colchicine in neutrophilic diseases.

Sulfasalazine has been shown to inhibit neutrophil responsiveness to chemoattractants; to inhibit chemotaxis, degranulation, and O2·− production; to decrease LTβ production; and to scavenge oxygen metabolites. Similar to sulfasalazine, gold salts may scavenge toxic metabolites. Gold salts, still in use for rheumatoid arthritis in some parts of the world, also decrease neutrophil collagenase activity and reduce E-selectin expression on endothelium.

The current era of biologic therapies has been ushered in through the introduction of agents designed to block the effects of TNF or IL-1. As noted earlier, IL-1 and TNF directly affect neutrophil function including priming for stimulus-induced responses such as O2·− production, cartilage destruction, and production of cytokines such as IL-8 and LTβ. Nonetheless, studies examining the effects of anti-TNF treatment on neutrophil function measured ex vivo have not indicated extensive action. Treatment of patients with etanercept or adalimumab induced no effect on neutrophil ex vivo responses including chemotaxis, phagocytosis, and superoxide generation (although CD69 levels were reduced). Reduction of neutrophil populations in rheumatoid arthritis joint effusions after anti-TNF therapy is more likely due to alteration of the inflammatory environment, rather than to direct effects on the neutrophils themselves.

References

The references for this chapter can also be found on www.expertconsult.com.
Eosinophils

JOSE U. SCHER • STEVEN B. ABRAMSON • MICHAEL H. PILLINGER

---

**KEY POINTS**

- Eosinophils are myeloid-lineage cells that contain many cytoplasmic granules consisting of proteins such as major basic protein, eosinophil cationic protein, eosinophil-derived neurotoxin, and eosinophil peroxidase.

- Eosinophils are thought to function in the defense of helminths and other parasites. However, evidence for such an antihelminthic role is limited.

- In contrast to neutrophils, eosinophils are not primarily phagocytic cells but are thought to discharge their granule contents adjacent to larger organisms that may be their targets.

- Eosinophilia may be seen in many rheumatic diseases, including Churg-Strauss syndrome, eosinophilic fasciitis, and the idiopathic hypereosinophilic syndromes.

---

The eosinophil line shares many features with the other families of polymorphonuclear granulocytes. In contrast to the neutrophil, however, the eosinophil is primarily a tissue-localized cell. Eosinophils are produced in numbers smaller than neutrophils, and their half-life in the blood is shorter (3 to 8 hours) owing to higher rates of diapedesis. Normal bloodstream levels of eosinophils tend to be low—typically less than 5% of blood leukocytes. When in the tissues, eosinophils are longer-lived than neutrophils, with estimates ranging from 2 to 14 days. Tissue eosinophils are found in greatest concentrations in gastrointestinal mucosa, suggesting that they participate in barrier rather than bloodstream surveillance.

**EOSINOPHIL DEVELOPMENT AND MORPHOLOGY**

Similar to neutrophils, eosinophils follow a classic pattern of granulocyte differentiation, passing through blast, promyelocyte, myelocyte, metamyelocyte, and band stages before reaching maturity. Along the way, eosinophils successively acquire morphologically distinct classes of granules. Factors required for eosinophil differentiation include granulocyte-macrophage colony-stimulating factor (GM-CSF) and interleukin (IL)-3, which also are required for neutrophil differentiation and cannot account for eosinophil commitment. An essential role for IL-5 in eosinophil development has been described, supported by the observation that intravenous administration of IL-5 rapidly results in peripheral eosinophilia. IL-5 may not be completely eosinophil specific, however, because studies in animal models suggest that it is also trophic for B cells. Similar to IL-5, IL-2 can stimulate eosinophilia. The IL-2 effect seems to be mediated through production of IL-5, however. CCL11 (eotaxin-1) also may cause bone marrow release of mature eosinophils and eosinophil precursors via engagement of CCR3 receptors, which are expressed mainly on eosinophils. Cooperation between IL-5 and eotaxins, in particular eotaxin-1, seems to be needed to induce tissue eosinophilia. Knockout mice with targeted deletion of CCR3 show deficiency in gastrointestinal eosinophils. Several transcription factors are involved in the eosinophilic lineage commitment, including the CCAAT/enhancer binding protein family (C/EBP) members, the interferon consensus sequence binding protein (ICSBP), and GATA-1.

When viewed under hematoxylin and eosin staining, eosinophils appear slightly larger than neutrophils (12 to 17 µm). Their nuclei typically are bilobed. Most striking is the presence of large, pink-staining granules. In addition, lipid bodies occasionally may be seen—nonvesicular accumulations of arachidonic acid and other lipids, presumably liberated from plasma membrane. These are not unique, however, and may be detected occasionally in neutrophils as well.

Eosinophils contain at least three distinguishable classes of granules (Table 12-1). Primary granules form first and are analogous to the primary (azurophilic) granules of neutrophils. In contrast to neutrophils, eosinophil primary granules lack myeloperoxidase. Eosinophil primary granules are most numerous in eosinophilic promyelocytes and persist in smaller numbers in mature forms. In mature eosinophils, a lysophospholipase that is present in large quantities (7% to 10% of total eosinophil protein) has been tentatively localized to primary granules, which when released extracellularly precipitates into bipyramidal structures known as Charcot-Leyden crystals. Deposition of these crystals in tissues is taken as evidence of present or past eosinophilia.

The large granules visible in mature eosinophils are specific granules that form during the myelocyte stage. When viewed under scanning electron microscopy, specific granules show a dense crystalline core surrounded by an intermediate-density matrix. Because of their large size and number (>90% of overall granule population), eosinophil specific granules have yielded to isolation and immunochemical examination, and their contents have been at least partially evaluated. Among the contents probably localized to specific granules are lysosomal enzymes (acid and neutral hydrolases, collagenase, cathepsin, and gelatinase), lectins, and components of the oxidase system. Most distinct is the presence of four highly basic proteins that lend the granule its tinctorial properties. Major basic protein (MBP), an 11,000-kD protein with an isoelectric point
value of 11, accounts for more than 50% of the total granule protein and is the major, or possibly sole, component of the crystalline core. Eosinophil cationic protein (ECP), actually a heterogeneous group of several related proteins (18 to 21 kD molecular weight), also is present in large amounts (up to 10% on weight/weight basis).

Eosinophil-derived neurotoxin (EDN) (18 kD molecular weight), the third of the basic granular proteins, is slightly less basic (isoelectric point 8.9) than the aforementioned proteins and is present in smaller quantities. In contrast to MBP and ECP, which have likely roles in host defense, EDN is mainly recognized for its function as a neurotoxin for myelinated neurons, the evolutionary advantage of which is unclear. The Gordon phenomenon, in which injection of eosinophil-laden tissue into an animal produces profound neurologic deficits, is likely due to eosinophil-derived neurotoxin. It has been shown that EDN serves as an endogenous ligand of TLR2, can activate Myd88 in dendritic cells, and shifts adaptive immunity toward a T helper (Th)-2 response, suggesting a pivotal role for eosinophils in the innate-adaptive immune response. Finally, eosinophil specific granules contain large quantities of eosinophil peroxidase, an enzyme distinctly different from neutrophil myeloperoxidase, but probably subsuming the same function of generating hypohalides for cell killing and activation of latent proteinases. ECP; EDN, and eosinophil peroxidase are localized within the primary granule matrix region. A third population of smaller eosinophilic granules has been identified by virtue of its acid phosphatase and arylsulfatase B content and is present mainly in tissue eosinophils.

**EOSINOPHIL ACTIVATION AND DISTRIBUTION**

Similar to neutrophils, eosinophils undergo activation in response to stimuli and are capable of adhesion, chemotaxis, phagocytosis, degranulation, and O$_2^-$ generation. Eosinophils respond to many of the same chemoattractant stimuli as neutrophils, although with different sensitivities. In addition, eosinophils respond to stimuli that do not affect neutrophils, including IL-3, IL-5, regulation upon activation normal T cell expressed and presumably secreted (RANTES), and macrophage inflammatory protein (MIP)-1α. Whether the distribution of these factors is sufficient to explain the tissue distribution of eosinophils relative to other granulocytes is uncertain; however, eosinophils and mast cells secrete IL-3 and IL-5, suggesting their capacity to attract additional eosinophils to sites of atopy. Eosinophils express not only adhesion molecules identified on neutrophils—CD11a/CD18, CD11b/CD18, and L-selectin—but also others, such as the α4β1 integrin VLA-4. IL-4 and IL-13 induce expression of the VLA-4 counterligand, vascular cell adhesion molecule (VCAM)-1, via an eotaxin-1, STAT6-dependent pathway. Oncostatin-M, an IL-6/gp130 family member, also seems to upregulate VCAM-1 in an eotaxin-1, STAT6-dependent manner, and to play a role in eosinophil accumulation in a mouse model.

Eosinophils also differ from neutrophils in their repertoire of immunglobulin receptors. Although eosinophils possess immunoglobulin (Ig)G receptors, these are relatively sparse. Instead, the predominant immunoglobulin receptors on the eosinophil surface are high affinity for IgA, consistent with the role of the eosinophil in barrier defense. Although eosinophils are activated by IgG and IgA, they are most potently activated by secretory IgA, probably owing to the presence of a receptor unique for the secretory component. In contrast to earlier teaching, expression of IgE receptors on eosinophils surfaces is minimal and most likely is of little biologic significance.

**NORMAL EOSINOPHIL FUNCTION**

Although some studies have demonstrated the capacity of eosinophils to phagocytose bacteria, others suggest that these cells phagocytose poorly, and it is likely that antibacterial defense is not a primary eosinophil function. Instead, eosinophils most likely participate in host defense against multicellular, helminthic parasites; eosinophilia typically occurs in response to parasitic but not bacterial infection. Eosinophils can phagocyte small parasitic forms but more characteristically attach, in a polarized manner, to the surface of larger parasites and discharge their granular contents into the protected space between the parasite and the eosinophil. Although O$_2^-$ generation and proteinase release may play a role in this attack, the specific granule-associated basic proteins are probably the major weapon in the anti-parasitic armamentarium of the eosinophil. In vitro studies have shown the capacities of these proteins, particularly ECP and MBP, to kill protozoa. Although ECP is ≈10-fold more potent, the higher concentration of MBP present in the granules suggests that it is the dominant parasite toxin. Parasitic killing by each of the proteins seems to depend on its capacity to disrupt the plasma membrane; in the case of ECP, membrane disruption occurs through the formation of pores or channels. More recently, it has been shown that similar to neutrophils, eosinophils are able to generate extracellular traps with bactericidal properties. In response to bacterial exposure—C5a or CCR3—eosinophils rapidly release mitochondrial DNA and granule proteins. In contrast to neutrophils, eosinophils do not undergo apoptosis upon release of their DNA, and their traps are composed.

### Table 12-1 Eosinophil Granule Contents

<table>
<thead>
<tr>
<th>Arylsulfatase Granules</th>
<th>Primary Granules</th>
<th>Specific Granules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative size</td>
<td>Smallest</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Contents</td>
<td>Arylsulfatase</td>
<td>Lysophospholipase</td>
</tr>
<tr>
<td>Acid phosphatase</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
mainly of ECP and MBP. Eosinophils (similar to neutrophils) have been reported to present antigen to T cells; it is not established whether antigen presentation plays any role in the antiparasitic effect. Although epidemiologic and in vitro evidence supports a role for eosinophils in parasitic defense, in vivo confirmation of such a role is equivocal. In particular, several studies indicate that eosinophil ablation (through IL-5 depletion) has no effect on the course of parasite infection in mice. This might reflect redundancy of antiparasitic defenses.

**EOSINOPHIL RELEVANCE TO INFLAMMATORY AND AUTOIMMUNE DISEASE**

**Asthma**

Although not strictly a rheumatic disease, asthma represents the most common inflammatory/autoimmune disease in which eosinophils predominate. Although the presence of blood and pulmonary eosinophilia in asthma has been appreciated for a century, the role of eosinophils in the pathogenesis of this disease remains a subject of intense study. Eosinophilia in asthma is stimulated by high levels of IL-5 and other cytokines. Eosinophils possess multiple mechanisms through which they can, at least potentially, enhance the asthmatic response. Similar to neutrophils, stimulated eosinophils synthesize leukotriene (LT)A₄ from arachidonic acid. In contrast to neutrophils, however, eosinophil metabolism of LT₄ leads to the production not of LTB₄, but of LTC₄ and LTD₄ (cysteinyl leukotrienes)—both potent bronchoconstrictors.³ Eosinophils themselves are exquisitely sensitive to the effects of cysteinyl leukotrienes, which stimulate eosinophil adhesion, migration, and degranulation and the proliferation of eosinophil progenitors. The cysteinyl leukotriene receptor antagonists (lukasts) are effective in the treatment of asthma and have been shown to have direct effects on eosinophils in vivo and in vitro, including reduction of eosinophil transmigration and reduction of pulmonary and peripheral eosinophilia.⁶ Lukasts seem to have beneficial effects on other diseases characterized by eosinophilia, including cystic fibrosis, eosinophilic gastroenteritis, and atrophic dermatitis.

Platelet-activating factor is produced by stimulated eosinophils and has bronchoconstricting activities. Release of specific granule proteins per se has multiple proinflammatory effects, including (1) epithelial damage secondary to membrane perturbations similar to those seen in parasites, and (2) activation of mast cells with subsequent histamine and leukotriene production. MBP also may act specifically as an antagonist of muscarinic M2 receptors, resulting in enhanced vagal tone and increased bronchospasm.⁷

**Rheumatic Diseases**

Although hypereosinophilia occasionally can be observed in virtually all rheumatic diseases, it is relatively uncommon in most, perhaps owing in part to the widespread use of corticosteroid therapy.⁸ In Churg-Strauss vasculitis, hypereosinophilia is the classic laboratory abnormality accompanying a constellation of pulmonary and renal vasculitis and asthma. In some cases, peripheral eosinophil counts in Churg-Strauss vasculitis have been reported to exceed 50% of total leukocytes. The cluster of asthma and eosinophilia accompanying Churg-Strauss vasculitis suggests that this syndrome may represent an atopic response to a foreign antigen. IgE response varies, however, and asthma may precede the rest of the disease by years. The presence of antineutrophil cytoplasmic antibody (ANCA), an IgG class antibody, is common, suggesting a broader autoimmune response. Systemic steroids with or without cyclophosphamide, followed by steroid-sparing agents remain the mainstream therapeutic options. Anti-IgE therapy with omalizumab and anti–IL-5 antibodies (mepolizumab) have been tried with mixed results.

Eosinophilia-myalgia syndrome was first observed in 1989 in New Mexico and was defined by the Centers for Disease Control and Prevention for surveillance purposes as peripheral eosinophilia and muscle pains unexplained by other illnesses. Rash and skin edema are common findings. Follow-up of cases over time revealed the frequent appearance of fibrosing fasciitis, which, in more severe disease, results in skin retraction, particularly over the veins, where it gives rise to a train-track appearance. Intensive epidemiologic investigation pinpointed the likely cause of the epidemic as consumption of l-tryptophan supplements produced by a single manufacturer, probably owing to trace contaminants. Discontinuation of the sale of the supplement led to resolution of the epidemic, although sporadic tryptophan-independent cases continue to be reported. Eosinophilic fasciitis, a condition first described in 1975, resembles eosinophilia-myalgia syndrome in that it involves fasciitis and eosinophilia but differs in that myalgias are not a prominent feature, and organ involvement is unusual. Clinically, the skin of patients with eosinophilic fasciitis bears some resemblance to the skin of patients with systemic sclerosis, but the distribution is typically on the distal extremities with sparing of the hands and feet. An epidemic similar to eosinophilia-myalgia syndrome was seen in Spain in 1981, related to consumption of adulterated rapeseed oil (toxic oil syndrome). Although in these syndromes it is unclear whether eosinophils act as mediators of fasciitis or merely as reporters of exposure to an atopic antigen, the capacity of eosinophils to produce transforming growth factor (TGF)-β suggests a potential role for these cells in the generation of fibrous tissue. The presence of eosinophils in affected tissues varies, however, with most infiltrates consisting of other leukocytes.

A single study has shown indirect evidence for the presence of increased numbers of eosinophils (Charcot-Leyden crystal deposition) in progressive systemic sclerosis. The relevance of this observation remains to be determined.

**Primary Eosinophilic Syndromes**

Idiopathic hypereosinophilic syndrome (HES) has been defined as (1) persistent eosinophils numbering 1500/mm³ for 6 or more months (or until death); (2) absence of parasites, allergy, or other causes of eosinophilia; and (3) signs and symptoms of organ involvement related directly to eosinophils or eosinophil accumulation. Morbidity is largely associated with eosinophil tissue infiltration, and granuloma formation may occur. Idiopathic hypereosinophilic
syndrome has gained attention with the description of patients with a genetic rearrangement, del(q12q12), that results in fusion of the platelet-derived growth factor receptor-α (PDGFRα) and Fip1-like 1 (FIP1L1) genes, generating a novel, constitutively active tyrosine kinase responsible for the clonal expansion of eosinophils. Targeted therapy with the selective tyrosine kinase inhibitor imatinib has become an effective tool in many cases of hypereosinophilic syndrome associated with the FIP1L1-PDGFRα fusion gene. A few placebo-controlled trials indicated that mepolizumab, an anti–IL-5 monoclonal antibody, also might offer clinical benefit and steroid-sparing effects. In a subset of patients with refractory HES, alemtuzumab, a monoclonal anti-CD52 antibody, lowers eosinophilia and induces remission. To date, corticosteroids remain the primary treatment for idiopathic hypereosinophilic syndrome.

Eosinophilic esophagitis (EE) is a more recently recognized entity defined by the accumulation of eosinophils in the esophagus, which, in contrast to gastroesophageal reflux disease, does not respond to therapy with proton pump inhibitors. Patients with eosinophilic esophagitis are predominantly young men who have high levels of eosinophils in the esophageal mucosa, extensive hyperplasia, a high rate of atopic disease, and normal pH monitoring compared with patients with gastroesophageal reflux disease. The prevalence seems to be increasing, notably among whites from Western countries. Recent studies suggest that human EE is driven by upregulation of eotaxin-3 in esophageal epithelial cells, which acts as a potent eosinophil chemoattractant. Oral fluticasone propionate and mepolizumab have proved effective in initial trials.

Löffler’s syndrome is a self-limiting eosinophilic pneumonitis with peripheral eosinophilia, presumably a hypersensitivity reaction. Allergic bronchopulmonary aspergillosis also represents a hypersensitivity reaction and may be indistinguishable from Löffler’s syndrome. A novel eosinophilic syndrome has been described, consisting of nodules, eosinophilia, rheumatism, dermatitis, and swelling (NERDS); because only a few cases have been reported to date, the clinical identity of this illness awaits validation.

Addison’s Disease

Addison’s disease is a disorder of adrenal failure resulting in underproduction of steroid hormones. Addison’s disease frequently is accompanied by peripheral eosinophilia. In contrast to increases in peripheral neutrophil counts, the ability of glucocorticoids to reverse the eosinophilia of Addison’s disease implicates that class of agents as a key regulator in the downregulation of eosinophil number. Glucocorticoids rapidly reduce eosinophil numbers in most hypereosinophilic and nonhypereosinophilic patients—a fact enshrined in the clinical maxim that detectable levels of eosinophils in a patient on long-term glucocorticoid therapy may be evidence of noncompliance with medication. Whether glucocorticoids have an effect on eosinophil production, release, or survival remains to be determined. Regardless of the mechanism of their effect, use of glucocorticoids to reduce eosinophil count is an important strategy in reducing morbidity from hypereosinophilic diseases.

References


The references for this chapter can also be found on www.expertconsult.com.
**T Lymphocytes**

RALPH C. BUDD • KAREN A. FORTNER

KEY POINTS

- T cells develop primarily in the thymus. The importance of the thymus is underscored by the complete absence of T cells in patients in whom a thymus has failed to develop (e.g., complete DiGeorge syndrome).

- Thymic selection consists of a positive phase in which T cells must recognize self-MHC molecules, and a negative phase in which thymocytes bearing high-affinity TCRs for self-MHC peptide are deleted through apoptosis.

- T cells emerge from the thymus as naïve T cells that are quiescent and, when activated, express low to negligible levels of most cytokines. Once they acquire a memory phenotype (CD45RO⁺), they can produce high levels of cytokines.

- Naïve T cells can spontaneously undergo homeostatic proliferation to self-MHC peptides in peripheral lymphoid tissues to generate a critical number of T cells. This requires IL-7 and IL-15.

- Th1 and Th17 cells accumulate in inflammatory synovium such as rheumatoid arthritis, whereas Th2 cells accumulate at sites of allergic responses such as asthma.

OVERVIEW

The evolutionary pressures that have molded the immune response and promoted a highly diverse repertoire clearly derive from infectious agents. Two different strategies exist. The more primitive innate immune response (see Chapter 18) uses a limited repertoire of nonpolymorphic receptors that recognize structural motifs that are common to many microorganisms. These include small glycolipids and lipopeptides. The alternative strategy of the evolutionarily newer adaptive immune response (see Chapter 19) relies on generating myriad different receptors that can recognize a wide array of foreign compounds from infectious agents. Whereas the innate immune response allows a rapid focused response, adaptive immunity permits a broader, albeit slower, response but critically offers the additional benefit to the host of immune memory.

T lymphocyte development constantly confronts the dilemma of combating infection without provoking a response to the host. The price for generating an increasingly varied population of antigen receptors needed to recognize a wide spectrum of pathogens is the progressive risk of producing self-reactive lymphocytes that can provoke an autoimmune diathesis. To minimize the possibility of self-reactive cells, T lymphocytes are subjected to a rigorous selection process during development in the thymus. In addition, premature activation of mature T cells is prevented by requiring two signals for activation. Finally, the tremendous expansion of T cells that occurs during the response to an infection is resolved by the active induction of cell death. The consequences of inefficient lymphocyte removal at any one of these junctures can be devastating to the health of the organism. This is vividly displayed in both humans and mice where naturally arising mutations in death receptors such as Fas result in massive accumulation of lymphocytes and autoimmune sequelae. These are discussed in more detail in Chapter 27 on cell survival and death.

The activation of T lymphocytes yields a variety of effector functions that are pivotal to combating infections. Cytolytic T cells can kill infected cells through the expression of perforin, which induces holes in cell membranes, or ligands for death receptors such as Fas or tumor necrosis factor (TNF) receptor. Production of T cell cytokines such as interferon-γ (IFN-γ) can inhibit viral replication, whereas other cytokines such as interleukin (IL)-4, IL-5, and IL-21 are critical for optimal B cell growth and immunoglobulin production. However, this same armamentarium, if not tightly regulated, can also precipitate damage to host tissues and provoke autoimmune responses. This is particularly apparent in situations where T cell infiltration can be observed histologically such as in the synovium of inflammatory arthritides, pancreatic islets in type 1 diabetes, and the central nervous system in multiple sclerosis. Damage in these cases need not be the direct result of recognition of target tissues by the T cells. T cells may be activated elsewhere and then migrate to the tissue and damage innocent bystander cells. T cells may also promote autoimmunity through the augmentation of B cell responses.

T CELL DEVELOPMENT

T cells must traverse two stringent hurdles during their development. First, they must successfully rearrange the genes encoding the two chains of the T cell antigen receptor (TCR). Second, T cells must survive thymic selection during which T cells that interact strongly with self-peptides are eliminated. This minimizes the chances of autoreactive T cells escaping to the periphery.

The TCR is an 80- to 90-kD disulfide-linked heterodimer composed of a 48- to 54-kD α chain and a 37- to 42-kD β chain. An alternate TCR composed of γ and δ chains is expressed on 2% to 3% of peripheral blood T cells and is
discussed later. The TCR has an extracellular ligand binding pocket and a short cytoplasmic tail that by itself cannot signal. Consequently it is noncovalently associated with as many as five invariant chains of the CD3 complex that relay information to the intracellular signaling machinery via immunoreceptor tyrosine activation motifs (ITAMs) (see later). Not surprisingly, the structure of the TCR gene is similar to what was first described for immunoglobulin genes in B cells (see details in Chapter 14). Each overcame the problem of how to encode approximately 10 million different T or B cell specificities within the human genome, which contains only 30,000 genes. To economically package this diversity, the process of gene rearrangement and splicing evolved using machinery similar to that which already existed to promote gene translocations. The β and δ chain genes of the TCR contain four segments known as the V (variable), D (diversity), J (joining), and C (constant) regions. The α and γ chains are similar but lack the J component. Each of the segments has several family members (= 50 to 100 V, 15 D, 6 to 60 J, and 1 to 2 C members). An orderly process occurs during TCR gene rearrangement in which a D segment is spliced adjacent to a J segment, which is subsequently spliced to a V segment.

Following transcription, the V(D)J sequence is spliced to a C segment to produce a mature TCR messenger RNA. Arithmetically, this random rearrangement of a single chain of the TCR locus can give rise to a minimum of 50V × 15D × 6J × 2C, or about 9000 possible combinations. At each of the splice sites, which must occur in-frame to be functional, additional nucleotides not encoded by the genome (so-called N-region nucleotides) can be incorporated, adding further diversity to the rearranging gene. The combinations from the two TCR chains, plus N-region diversity, yield at least 10^8 possible combinations. Cutting, rearranging, and splicing are directed by specific enzymes. Mutations in the genes mediating these processes can result in arrest in lymphocyte development. For example, mutation in the gene encoding a DNA-dependent protein kinase required for receptor gene recombination results in a severe combined immunodeficiency (SCID).

Because the developing T cell has two copies of each chromosome, there are two chances to successfully rearrange each of the two TCR chains. As soon as successful rearrangement occurs, further β-chain rearrangements on either the same or the other chromosome are suppressed, a process known as allelic exclusion. This limits the chance of dual TCR expression by an individual T cell. The high percentage of T cells that contain rearrangements of both β-chain genes attests to the inefficiency of this complex event. Rearrangement of the α chain occurs later in thymocyte development in a similar fashion, although without apparent allelic exclusion. This can result in dual TCR expression by a single T cell.

Development of T cells occurs within a microenvironment provided by the thymic epithelial stroma. The thymic anlage is formed from embryonic ectoderm and endoderm and is then colonized by hematopoietic cells, which give rise to dendritic cells, macrophages, and developing T cells. The hematopoietic and epithelial components combine to form two histologically defined compartments: the cortex, which contains immature thymocytes, and the medulla, which contains mature thymocytes (Figure 13-1A). As few as 50 to 100 bone marrow–derived stem cells enter the thymus daily.

The stages of thymocyte development can be defined by the status of rearrangement and expression of the two genes that encode the α and β chains of the TCR and the expression of CD4 and CD8, proceeding in an orderly fashion from CD4^- CD8^- to CD4^- CD8^+ or CD4^+ (Figure 13-1B). CD4 and CD8 define, respectively, the helper and cytolytic subsets of mature T cells.

CD4^- thymocytes can be further subdivided based on their expression of CD25 (the high-affinity IL-2 receptor α chain) and CD44 (the hyaluronate receptor). Development proceeds in this order: CD25^- CD44^- to CD25^+ CD44^- to CD25^+ CD44^+. These subpopulations correspond to discrete stages of thymocyte differentiation. CD25^+ cells express low levels of CD4, and their TCR genes are in germline configuration. These cells downregulate CD4 and upregulate CD25 to give rise to CD25^- CD4^- thymocytes, which now express surface CD2 and low levels of CD3ε. At the next stage (CD25^- CD44^-), there is a brief burst of proliferation followed by upregulation of the recombination-activating enzymes, RAG-1 and RAG-2, and the concomitant rearrangement of the genes of the TCR β chain. A small subpopulation of T cells rearranges and expresses a second pair of TCR genes known as γ and δ. Productive TCR β-chain rearrangement results in downregulation of RAG and a second proliferative burst. Loss of CD25 then yields CD25^- CD44^- thymocytes.

The TCR β chain cannot be stably expressed without an α chain. Because the TCR α chain has not yet rearranged, a surrogate invariant TCR pre-α chain is disulfide linked to the β chain. When associated with components of the CD3 complex, this allows a low-level surface expression of a pre-TCR and progression to the next developmental stage. Failure to successfully rearrange the TCR β chain results in a developmental arrest at the transition from CD25^- CD44^- to CD25^- CD44^+. This occurs in RAG-deficient mice, as well as in mice and humans with SCID.

A number of signaling molecules are required for early T cell development (Figure 13-2). The Ikaros gene encodes a family of transcription factors required for the development of cells of lymphoid origin. Notch-1, a molecule known to regulate cell fate decisions, is also required at the earliest stage of T cell lineage development. Cytokines including IL-7 promote the survival and expansion of the earliest thymocytes. In mice deficient for IL-7, its receptor components IL-7Rα or γc, or the cytokine receptor–associated signaling molecule JAK-3, thymocyte development is inhibited at the CD25^- CD44^- stage. In humans, mutations in γc or JAK-3 result in the most frequent form of SCID. Pre-TCR signaling is required for the CD25^- CD44^- to CD25^- CD44^+ transition. Thus loss of signaling components including Lck, SLP-76, and LAT-1 results in a block at this stage of T cell development. TCR signals are also required for differentiation of CD4^- CD8^- to CD4^+ or CD8^- cells. Humans deficient in ZAP-70 (see later) have CD4^+ but not CD8^- T cells in the thymus and periphery.

CD25^- CD44^- cells upregulate expression of CD4 and CD8 to become CD4^- CD8^+. It is as a CD4^- CD8^- thymocyte that the α chain of the TCR rearranges. Unlike the β chain, allelic exclusion of the α chain is not apparent. Rearrangement of the α chain can occur simultaneously on both
The earliest thymocyte precursors lack expression of CD4 and CD8 (CD4−CD8−). These can be further divided into four subpopulations based on sequential expression of CD44 and CD25. It is at the CD44−CD25+ stage that the TCR-β-chain rearranges. The SCID mutation or deficiencies of the rearrangement enzymes Rag-1 and Rag-2 result in inability to rearrange the β-chain and maturational arrest at this stage. Those thymocytes that successfully rearrange the β-chain express it associated with a surrogate α-chain known as pre-Tα. Concomitant with a proliferative burst, development can then progress to the CD4+CD8+ stage in the cortex where the TCR-α-chain rearranges and pairs with the β-chain to express a mature TCR complex. These cells then undergo thymic positive and negative selection (as diagrammed in Figure 13-3B). Successful completion of this rigorous selection process results in mature CD4+ or CD8+ T cells in the medulla, which eventually emigrate to peripheral lymphoid sites.

B, Schematic two-color flow cytometry showing subpopulations of thymocytes defined by CD4 and CD8 expression in their relative proportions.

chromosomes, and if one attempt is unsuccessful, repeat rearrangements to other Vα segments are possible. Reports exist of dual TCR expression by up to 30% of mature T cells in which the same T cell expresses different α chains paired with the same β chain. However, in most cases of dual TCR α chains, one is downregulated during positive selection by Lck and Cbl, through ubiquitination, endocytosis, and degradation.

Although the structure of immunoglobulin and TCR are quite similar, they recognize fundamentally different antigens. Immunoglobulins recognize intact antigens in isolation, either soluble or membrane bound, and are often sensitive to the tertiary structure. The TCRβ2 recognizes linear stretches of antigen peptide fragments bound within the grooves of either major histocompatibility complex (MHC) class I or class II molecules (Figure 13-3A). Thymic selection molds the repertoire of emerging TCR so that they recognize peptides within the groove of self-MHC molecules, ensuring self-MHC restriction of T cell responses. The MHC structure is described in detail in Chapters 19 through 21. Pockets within the MHC groove bind particular residues along the peptide sequence of 7 to 9 amino acids for MHC class I and 9 to 15 amino acids for MHC class II molecules. As a result, depending on the particular MHC molecule, certain amino acids will make strong contact with the MHC groove while others will contact the TCR.

The contact between the TCR and MHC/peptide has been revealed by crystal structure to be remarkably flat, rather than a deep lock and key structure one might imagine. The TCR axis is tipped about 30 degrees to the long axis of the MHC class I molecule and is slightly more skewed to MHC class II. The affinity of the TCR for antigen/MHC is in the micromolar range. This is less than many antibody-antigen affinities and is several logs less than many enzyme-substrate affinities. This has led to the notion that TCR interactions with antigen/MHC are brief and that successful activation of the T cell requires multiple interactions, resulting in a cumulative signal.

Once the T cell has successfully rearranged and expressed a TCR in association with the CD3 complex, it encounters the second major hurdle in T cell development, thymic selection. Selection has two phases, positive and negative, and the outcome is based largely on the intensity of TCR signaling in response to interactions with MHC/self-peptides expressed on thymic epithelium and dendritic cells. TCR signals that are either too weak (death by neglect) or too intense (negative selection) result in elimination by apoptosis, whereas those with intermediate signaling intensity survive positive selection (Figure 13-3B). Successful positive selection at the CD4+8+ stage is coincident with upregulation of surface TCR, the activation markers CD5 and CD69, and the survival factor Bcl-2. T cells bearing a TCR that
recognizes MHC class I maintain CD8 expression, down-regulate CD4, and become CD4^8+. T cells expressing a TCR that recognizes MHC class II become CD4^8^.

An enigma for thymic selection has been how to present the myriad self-proteins to developing thymocytes so that self-reactive thymocytes are effectively eliminated by negative selection. This includes particularly those antigens with tissue or developmentally restricted expression. A solution was found with the discovery of the autoimmune regulator (AIRE) gene. AIRE is a transcription factor expressed by the medullary epithelium of the thymus that induces transcription of a wide array of organ-specific genes, such as insulin, that might otherwise be sequestered from developing thymocytes. This effectively creates a self-transcriptome within the thymus against which autoreactive T cells can be deleted. Gene knockout mice of AIRE and humans bearing AIRE mutations manifest various autoimmune sequela of a syndrome known as autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED).

Not surprisingly, a variety of signaling molecules activated by TCR engagement are important to thymic selection. Lck, the Ras \( \text{Raf-1} \cdot \text{MEK1} \cdot \text{ERK} \) kinase cascade, the kinase ZAP-70, and the phosphatases CD45 and calcineurin are involved with positive selection. Among these the Ras \( \text{ERK} \) pathway is particularly important because dominant negative variants of these molecules can disrupt positive selection. Conversely, an activator of Ras known as GRP1 assists the positive selection of thymocytes expressing weakly selecting signals. These molecules are discussed in more detail in the section on TCR signaling.

The survivors of these two stringent processes of TCR gene rearrangement and thymic selection represent less than 3% of total immature thymocytes. This is reflected in the presence of a high rate of cell death in developing thymocytes. This can be visualized by the measurement of DNA degradation, a hallmark of apoptosis, as shown in Figure 13-4. The survivors become either CD4^+ helper or CD8^+ cytolytic T cells and reside in the thymic medulla for...
Abnormalities of Human T Cell Development

Given the vast number of events in T cell development, it is not surprising that a multiplicity of causes can underlie human T cell immunodeficiencies. The influence of the thymic stroma on thymocyte ontogeny is underscored in the DiGeorge anomaly in which development of the pharyngeal pouches is disrupted and the thymic rudiment fails to form. This results in the failure of normal T cell development. Less severe T cell deficiencies are associated with a failure to express MHC class I and/or class II (the “bare lymphocyte syndrome”), which are directly involved with interactions required to induce the positive selection of, respectively, CD8+ and CD4+ mature T cells.

Metabolic disorders can affect thymocytes more directly. The absence of functional adenosine deaminase (ADA) and purine nucleoside phosphorylase (PNP) results in the buildup of metabolic by-products that are toxic to developing T and B lymphocytes. This ultimately produces forms of SCID.

The inability to express a number of surface molecules important in TCR and cytokine signaling also has the potential to perturb development. The failure to express TCR-CD3 components (specifically CD3γ and CD3ζ), CD18, and IL-2Rγ have all been noted among patients who exhibit varying degrees of T cell deficiency or dysfunction. All these molecules are involved in signaling of thymocyte development and survival, and their absence clearly has the potential to alter developmental fate.

PERIPHERAL MIGRATION OF T CELLS

The migration of naïve T cells to peripheral lymphoid structures or their infiltration into other tissues requires the coordinate regulation of an array of cell adhesion molecules. T cell recirculation is essential for host surveillance and is carefully regulated by a specific array of homing receptors. Entry from the circulation to tissues occurs via two main anatomic sites: the flat endothelium of the blood vessels and specialized postcapillary venules known as high endothelial venules (HEV). A three-step model has been proposed for lymphocyte migration: rolling, adhesion, and migration. L-selectin expressed by naïve T cells binds via lectin domains to carbohydrate moieties of GlyCAM-1 and CD34 (collectively known as peripheral node addressin) that are expressed on endothelial cells, particularly HEV. The weak binding of CD62L to its ligand mediates a weak adhesion to the vessel wall, which, combined with the force of blood flow, results in rolling of the T cell along the endothelium. The increased cell contact assists the interaction of a second adhesion molecule on lymphocytes, the integrin LFA-1 (CD11a/CD18) with its ligands, ICAM-1 (CD54), and ICAM-2 (CD102). This results in the arrest of rolling and firm attachment. Migration into the extracellular matrix of tissues may involve additional lymphocyte cell surface molecules such as the hyaluronate receptor (CD44) or the integrin α4β7 (CD49d/β7) that binds the mucosal addressin cell adhesion molecule-1 (MAdCAM-1) on endothelium of Peyer’s patches and other endothelial cells.

Other cytokines known as chemokines may contribute to lymphocyte homing. Chemokines are structurally and functionally related to proteins bearing an affinity for heparan sulfate proteoglycan and promote migration of various cell types. The chemokines RANTES, MIP-1α, MIP-1β, MCP-1, and IL-8 are produced by a number of cell types including endothelium, activated T cells, and monocytes and are present at inflammatory sites such as rheumatoid synovium (see Chapter 70).

Once mature T cells have reached the peripheral lymphoid tissues of lymph node and spleen, they undergo a low level of turnover known as homeostatic proliferation in response to self-peptide/MHC complexes, IL-7, and IL-15. This serves to maintain the number of peripheral T cells, which can become accelerated in states of lymphopenia such as following chemotherapy or irradiation. Because homeostatic proliferation is driven by self-MHC/peptides, its
Figure 13-4  T cell antigen receptor (TCR) signal pathways. Schema showing the principal signal pathways resulting from TCR activation and how they impinge on the regulatory region of the interleukin-2 (IL-2) gene. See text for details.

ACTIVATION OF T CELLS

T cell activation initiates intracellular signaling cascades that ultimately result in proliferation, effector function, or death, depending on the developmental stage of the cell. To guard against premature or excessive activation, T cells have a requirement of two independent signals for full activation. Signal 1 is an antigen-specific signal provided by the binding of the TCR to antigenic peptide complexed with MHC. Signal 2 is mediated by either cytokines or the engagement of co-stimulatory molecules such as B7.1 (CD80) and B7.2 (CD86) on the antigen-presenting cell (APC). Receiving only signal 1 without co-stimulation results in T cell unresponsiveness or anergy.

TCR and Tyrosine Kinases

TCR αβ and γδ have short cytoplasmic domains and by themselves are unable to transduce signals. The molecules of the noncovalently associated CD3 complex couple the TCR to intracellular signaling machinery (Figure 13-4). The CD3 complex contains nonpolymorphic members known as CD3ε, CD3γ, CD3δ, and ζ and η chains that are alternatively spliced forms of the same gene and are not genetically linked to the CD3 complex. Although the functional stoichiometry of the TCR complex is not completely defined, current data indicate that each TCR heterodimer is associated with three dimers: CD3 ε, CD3 δ, and ζζ or ζζ δ δ have an immunoglobulin-like extracellular domain, a transmembrane region, and a modest cytoplasmic domain, whereas ζζ contains a longer cytoplasmic tail. The transmembrane domains of ζζ and the CD3 chains contain a negatively charged residue that interacts with positively charged amino acids in the transmembrane domain of the TCR.

None of the proteins in the TCR complex has intrinsic enzymatic activity. Instead, the cytoplasmic domains of the invariant CD3 chains contain conserved activation domains that are required for coupling the TCR to intracellular signaling molecules. These immunoreceptor tyrosine-based activation motifs (ITAMS) contain a minimal functional consensus sequence of paired tyrosines (Y) and leucines (L): (D/E)XXYXXL(X)6-8YXXL. ITAMs are substrates for cytoplasmic protein tyrosine kinases (PTKs), and upon phosphorylation they recruit additional molecules to the TCR complex. Each ζζ chain contains three ITAMs, whereas there is one in each of the CD3 ε, γ, and δ chains. Thus each TCR complex can contain as many as 10 ITAMs.

Activation of PTKs is one of the earliest signaling events following TCR stimulation. Four families of PTKs are known to be involved in TCR signaling: Src, Csk, Tec, and Syk. The Src family members Lck and Fyn have a central role in TCR signaling and are expressed exclusively in lymphoid cells. Src PTKs contain multiple structural domains including (1) N-terminal myristylation and palmitoylation sites, which allow association with the plasma membrane; (2) a Src homology (SH) 3 domain, which associates with proline-rich sequences; (3) an SH2 domain that binds phosphotyrosine-containing proteins; and (4) a
Adaptor Proteins

Phosphorylation of tyrosine residues in ITAMs and PTKs following TCR stimulation creates docking sites for adaptor proteins. Adaptor proteins contain no known enzymatic or transcriptional activities but mediate protein-protein interactions or protein-lipid interactions. They function to bring proteins in proximity to their substrates and regulators, as well as sequester signaling molecules to specific subcellular locations. The protein complexes formed can function as either positive or negative regulators of TCR signaling depending on the molecules they contain.

Two critical adaptor proteins for linking proximal and distal TCR signaling events are SH2-domain-containing leukocyte protein of 76 kDa (SLP-76) and Linker for activation of T cells (LAT) (see Figure 13-4). Loss of these adaptor proteins has profound consequences for T cell development. Mice deficient for LAT or SLP-76 manifest a block in T cell development at the CD4^+CD25^-CD44^ stage. LAT is constitutively localized to lipid rafts and, following TCR stimulation, is phosphorylated on tyrosine residues by ZAP-70. Phosphorylated LAT then recruits SH2-domain-containing proteins including PLCγ1, the p85 subunit of phosphoinositide-3 kinase, IL-2 inducible kinase (Itk), and the adaptors Grb2 and Gads. Because the SH3 domain of Gads is constitutively associated with SLP-76, this brings SLP-76 to the complex where it is phosphorylated by ZAP-70. SLP-76 contains three protein binding motifs: tyrosine phosphorylation sites, a proline-rich region, and an SH2 domain. The N terminus of SLP-76 contains tyrosine residues that associate with the SH2 domains of Vav, the adaptor Nck, and Itk. Vav is a 95-kD protein that acts as a guanine nucleotide exchange factor for the Rho/Rac/cdc42 family of small G proteins. The complex of LAT, SLP-76/Gads, PLCγ1, and associated molecules results in the full activation of PLCγ1 and activation of Ras/Rho GTases and the actin cytoskeleton.

In addition to acting as positive regulators for TCR signaling, adaptors can also mediate negative regulation. As described previously, the activity of the Src family kinases is regulated by the interaction of kinases (Csk) and phosphatases (CD45) specific for inhibitory C-terminal phosphotyrosine. This is determined by the subcellular localization of these regulatory molecules. A second mechanism by which adaptor proteins can negatively regulate TCR stability is through regulation of protein stability. c-Cbl and Cbl-b are members of a conserved family of proteins that contains a highly conserved N-terminal region containing a tyrosine kinase-binding and RING-finger domains. c-Cbl RING finger domain binds the E2 ubiquitin-conjugating enzymes. Active E2 enzymes are brought into proximity with tyrosine kinase-binding proteins resulting in their ubiquination and degradation by the proteasome complex. Syk and ZAP-70 associate with c-Cbl while Vav, ZAP-70, Lck PLCγ1, and the p85 subunit of PI3K associate with Cbl-b.

**Downstream Transcription Factors**

The previously mentioned signaling events couple TCR stimulation to downstream pathways that culminate in changes in gene transcription that are required for proliferation and effector function (see Figure 13-4). One of the best characterized genes induced following T cell activation is the T-cell growth factor IL-2. Transcription of the IL-2 gene is regulated in part by the transcription factors AP-1, nuclear factor of activated T cells (NFAT), and nuclear factor κB (NFκB), all of which are activated following TCR signaling.
stimulation. Proximal signaling events lead to the activation of Ras and PLCγ. Ras initiates a cascade of kinases including Raf-1, MEK, and the MAP kinase ERK, which leads to the production of the transcription factor Fos. Activation of the co-stimulatory molecule CD28 results in the activation of another member of the MAP kinase family, c-Jun N-terminal kinase (JNK), and phosphorylation of the transcription factor c-Jun. c-Jun and Fos associate to form AP-1. PLCγ hydrolyzes membrane inositol phospholipids to generate phosphoinositide second messengers including inositol 1,4,5 triphosphate (IP3) and diacylglycerol. IP3 stimulates the mobilization of calcium from intracellular stores. Diacylglycerol activates protein kinase C (especially PKCθ in T cells) and, along with CARMA, connects with the NFκB pathway.13

Increased intracellular calcium is central to many forms of cellular activation. Calcium activates the calcium/calmodulin-dependent serine phosphatase calcineurin, which dephosphorylates NFAT. Dephosphorylated NFAT translocates to the nucleus and, together with AP-1, forms a trimolecular transcription factor for the IL-2 gene. The immunosuppressive agents cyclosporin-A and FK-506 specifically inhibit the calcium-dependent activation of calcineurin, thereby blocking activation of NFAT and the transcription of NFAT-dependent cytokines such as IL-2, IL-3, IL-4, and GM-CSF. Recently, it has been appreciated that differences in the amplitude and duration of calcium signals mediate different functional outcomes. Although high spikes of calcium are easily measured in lymphocytes during the first 10 minutes following antigen stimulation, sustained low-level calcium spikes over a few hours are necessary for full activation. These latter, more subtle calcium fluxes appear to be controlled by cyclic-ADP-ribose and ryanodine receptors. Selective inhibitors exist for these molecules, opening the potential for new specific blockers of T cell activation.

A surprising discovery was the observation that caspase activity, particularly of caspase-8, is required to initiate T cell proliferation and activation of NFκB. Previously the role of caspases had been confined to apoptosis. However, it is now appreciated that following TCR ligation, caspase-8 is activated and forms a complex that includes the NFκB adaptor proteins CARMA1, Bcl-10, and MALT1.13

**Co-stimulation**

Signal 2 serves to augment the activation of PI3 kinase and AKT, which augments not only growth factor production but also survival and general metabolic signals. The prototype co-stimulatory signal is CD28 interacting with B7-1 (CD80) or B7-2 (CD86). CD28 is a disulfide-linked homodimer constitutively expressed on the surface of T cells. Virtually all murine T cells express CD28, whereas in human T cells, nearly all CD4+ and 50% of CD8+ cells express CD28. The CD28 subset of T cells appears to represent a population that has undergone chronic activation and can manifest suppressive activity. Increased levels of CD28+ T cells have been reported in several inflammatory and infectious conditions including granulomatosis with polyangiitis (formerly Wegener’s granulomatosis), rheumatoid arthritis, cytomegalovirus, and mononucleosis.14,41 The cytoplasmic domain of CD28 has no known enzymatic activity but does contain two SH3 and one SH2 binding sites. CD28 interacts with PI3 kinase and ORB2 and promotes JNK activation, as noted earlier. CD28 ligation alone does not transmit a proliferative response to T cells, but in conjunction with TCR engagement augments IL-2 production at the level of both transcription and translation. It also increases the production of other cytokines including IL-4, IL-5, IL-13, IFN-γ, and TNF, as well as the chemokines IL-8 and RANTES.

The ligands for CD28, CD80 (B7-1), and CD86 (B7-2) are expressed in a restricted distribution on B cells, dendritic cells, monocytes, and activated T cells. CD80 and CD86 have similar structures but share only 25% amino acid homology. They each contain rather short cytoplasmic tails that may signal directly and bind to CD28 with different avidities.

**Immunologic Synapse**

Antigen-specific interaction between the T cell and APC results in the formation of a specialized contact region called the immunologic synapse or supramolecular activation cluster (SMAC) (Figure 13-5). Synapse formation is an active, dynamic process that requires specific antigen to drive synapse formation; TCR/MHC interaction alone is not sufficient. The synapse also overcomes the obstacles to close T cell/APC contact mediated by short molecules (e.g., TCR, MHC, CD4, CD8) caused by interactions of tall molecules (ICAM-1, LFA-1, CD45). Two stages of assembly have been described. During the nascent stage, cell adhesion molecules such as ICAM-1 on APC and LFA-1 on T cells make contact in a central zone, surrounded by an annulus of close contact between MHC and TCR. Within minutes the engaged TCR migrates to the central area, resulting in a mature synapse in which the initial relationships are reversed—the central area (CSMAC) now contains TCR, CD2, CD28, and CD4 and is enriched for Lck, Fyn, and PKCθ. Surrounding the central domain is a peripheral ring (pSMAC) that contains, CD45, LFA-1, and associated talin. T cell activation leads to compartmentalization of activated TCR and TCR signaling molecules to plasma membrane microdomains called “rafts.” Rafts are composed primarily of glycosphingolipids and cholesterol and are enriched in signaling molecules, actin, and actin-binding proteins. Src family kinases, Ras-like G proteins, LAT, and phosphatidylinositol-anchored membrane proteins have all been shown to localize to raft domains.

Full T cell activation requires engagement of a minimum of about 100 to 200 MHC/peptide molecules on an APC, which can serially stimulate 2000 to 8000 TCR. It has been estimated that naïve T cells also require a sustained signal for 15 to 20 hours to commit to proliferation. T cells face a number of obstacles to achieving full activation including the small physical size of the TCR and MHC molecules compared with other cell surface molecules, the low affinity of TCR for MHC/peptide complex, and the low number of MHC molecules present on the APC that contain the antigenic peptide. The immunologic synapse may provide the mechanism for overcoming these barriers and achieving the duration of TCR stimulation necessary to commit the cell to proliferation. The spatial organization of the synapse juxtaposes the membranes of the APC and T cell, assisting
The immune system is constantly confronted by the problem of how to ensure that T cells are activated only under conditions where there is a true need for a response to a foreign pathogen and not merely a self-component. As with many biologic filters, thymic negative selection is not 100% efficient and not all self-reactive T cells are eliminated. Hence a variety of fail-safe mechanisms are engaged to suppress the ability of these errant T cells to undergo premature clonal expansion. This is partly regulated by the requirement for two distinct signals from separate molecules to be coordinately triggered in order for T cell activation and proliferation to proceed. If only one of the signals is received, the T cell will not proliferate and will actually enter a nonresponsive state known as tolerance or anergy.

The anergy that results from the absence of a CD28 co-stimulatory signal manifests at a signal level by a failure to fully couple the TCR signal to the Ras/MAP kinase pathway and consequent AP-1 transcriptional activity. The available MHC/peptide complexes and TCR are concentrated at the site of contact via actin cytoskeleton-mediated transport. The multistep process of mature synapse formation may also enable the T cell to discriminate between the potential antigen-containing MHC/peptide complexes it encounters on the APC cell surface. It has been shown that co-stimulatory signals may contribute to synapse formation by initiating the transport of membrane rafts containing the kinases and adaptor molecules required for TCR signaling to the site of contact.

Although it has been appreciated for some time that chronic infection can lead to an unresponsive state or “exhausted” T cells, the molecular explanation for this was unknown. More recently it was observed that chronically activated CD8\(^+\) T cells contain more mRNA encoding the inhibitory receptor PD1 (programmed death 1) than acutely activated CD8\(^+\) T cells. In parallel, one of the ligands for PD-1, PDL1, was highly expressed by chronically infected splenocytes. Treatment of mice with a blocking antibody to PDL1 caused virus-specific CD8\(^+\) T cells to undergo marked expansion. The fact that many tumors also express PDL1 enhances interest in reversing suppression of immune responses during infection and tumorigenesis. The fact that PD-1-deficient mice develop spontaneous autoimmunity enhances interest in manipulating PD-1 for therapeutic purposes.

### Tolerance and Control of Autoreactive T Cells

The immune system is constantly confronted by the problem of how to ensure that T cells are activated only under conditions where there is a true need for a response to a foreign pathogen and not merely a self-component. As with many biologic filters, thymic negative selection is not 100% efficient and not all self-reactive T cells are eliminated. Hence a variety of fail-safe mechanisms are engaged to suppress the ability of these errant T cells to undergo premature clonal expansion. This is partly regulated by the requirement for two distinct signals from separate molecules to be coordinately triggered in order for T cell activation and proliferation to proceed. If only one of the signals is received, the T cell will not proliferate and will actually enter a nonresponsive state known as tolerance or anergy.

The anergy that results from the absence of a CD28 co-stimulatory signal manifests at a signal level by a failure to fully couple the TCR signal to the Ras/MAP kinase pathway and consequent AP-1 transcriptional activity. An additional method of provoking an incomplete TCR signal
CHAPTER 13 | T LYMPHOCYTES

and unresponsiveness is to make amino acid substitutions in the recognized peptide antigen. These so-called altered peptide ligands (APLs) cause a suboptimal phosphorylation of TCRζ and consequent inefficient recruitment of ZAP-70.

Following the discovery of CD28 as a co-stimulatory molecule, a related structure known as CTLA-4 was found to also bind to CD80 and CD86 with 20-fold higher affinity than CD28. Unlike CD28, CTLA-4 is expressed only transiently following T cell activation and confers an inhibitory signal for T cell proliferation. In this capacity, CTLA-4 functions to limit T cell clonal expansion induced by CD28. The consequences of the loss of this negative regulation are striking. The genetic deletion of the CTLA-4 gene in mice results in enormous uncontrolled T cell expansion and an autoimmune diathesis.

Chronic exposure to certain inflammatory cytokines, most notably TNF, can also induce anergy. It has been appreciated for some time that T cells from rheumatoid synovium manifest profound deficiencies of proliferation and cytokine production. Since TNF is one of the major cytokines detectable in rheumatoid synovial fluid, it was soon appreciated that chronic exposure of T cell clones to TNF for 10 to 12 days suppressed proliferative and cytokine responses to antigen by as much as 70%. Furthermore, a single administration of anti-TNF receptor monoclonal antibody to patients with rheumatoid arthritis rapidly restored the response of peripheral T cells to mitogens and recall antigens. Similar observations have been made in TCR transgenic mice following TNF exposure. The observation that chronic TNF exposure inhibited calcium responses following TCR ligation supports the view that TNF may uncouple TCR signaling. Conceivably other members of the TNF family may invoke similar T cell anergy.

An additional negative regulator for T cells is the B lymphocyte–induced maturation protein 1 (Blimp-1), previously felt to be expressed only in B lymphocytes. Blimp-1-deficient mice manifest augmented levels of peripheral effector T cells and develop severe colitis as early as 6 weeks of age. Blimp-1 messenger RNA expression increases with TCR stimulation, and Blimp-1-deficient T cells proliferated more and produced more IL-2 and IFN-γ following activation.

Another layer of regulation occurs via a phenotypically defined subpopulation of CD4^+CD25^+FoxP3^+ regulatory T cells (Treg) that has the ability to inhibit antigen-induced proliferation. This subset is expressed in the periphery at a low frequency and appears to be at least partly thymic dependent. The latter point may be of interest because it suggests that the absence of regulatory T cells following day 3 thymectomy may be involved with the subsequent development of autoimmune disease in these animals. Indeed, diminished levels of CD4^+CD25^+FoxP3^+ regulatory T cells have been observed in other autoimmune syndromes and the transfer of Treg to autoimmune mice has shown some alleviation of symptoms. The production of TGF-β and IL-10 appear to be critical to the suppressive activity of Treg. This is an active area of research because of the potential therapeutic implications for autoimmune diseases and their possible role in generating IL-17-producing CD4^+ T cells (Th17, see later). In this regard, recent studies to treat type 1 diabetes by increasing Treg number and function using anti-CD3 antibody or IL-2 are promising.

SUBSETS AND FUNCTION OF PERIPHERAL T CELLS

CD4 Helper and CD8 Cytolytic T Cells

αβ^+ T cells can be subdivided into two main subsets based on their recognition of peptides presented by MHC class I or class II molecules and their respective expression of CD8 or CD4. CD4^+ and CD8^+ T cells have different functions and recognize antigens derived from different cellular compartments. The peptides presented by MHC class I molecules are produced by the proteasome and can be derived from either self-proteins or intracellular foreign proteins as might occur during viral infection. MHC class II-bound peptides are derived largely from extracellular infectious agents or self–cell surface proteins that have been engulfed and degraded in the lysosomal complex.

CD4^+ T cells express a variety of cytokines and cell surface molecules that are important to B cell proliferation and immunoglobulin production and CD8^+ T cell function. Following antigen stimulation, CD4^+ T cells differentiate into different classes of effector T cells based on their cytokine profiles including T helper 1 (Th1), Th2, Th17, Tfh cells (described later), and Treg cells (described earlier) (Figure 13-6). The CD4 molecule is structurally related to immunoglobulins and has an affinity for nonpolymorphic residues on the MHC class II molecule. In this capacity CD4 presumably increases the efficiency with which CD4^+ T cells recognize antigen in the context of MHC class II molecules, whose expression is restricted to B cells, macrophages, dendritic cells, and a few other tissues during states of inflammation. In addition, the cytoplasmic tail of CD4 binds to Lck and promotes signaling by the TCR, as described earlier. However, ligation of CD4 before engagement of the TCR renders the T cell susceptible to apoptosis on subsequent engagement of the TCR. This is clinically important in human immunodeficiency virus (HIV) infections in which the gp120 molecule of HIV binds to CD4 and primes the T cell to undergo cell death when later triggered by the TCR. Accelerated apoptosis of CD4^+ T cells has been demonstrated in acquired immunodeficiency syndrome (AIDS) patients.

CD8^+ T cells are efficient killers of pathogen-infected cells. Given the ubiquitous expression of MHC class I molecules, mature cytolytic T cells (CTLs) can recognize viral infections in a wide array of cells, in distinction to the more restricted distribution of class II molecules. CTLs lyse target cells through the production of perforin, which induces holes in cell membranes, and the expression of Fas-ligand and TNF, which induce apoptosis. In this capacity CTLs kill virally infected target cells in an attempt to restrict the spread of infection. Similar to CD4, CD8 manifests an affinity for MHC class I molecules, enhances the signaling of CTL, and also binds Lck by its cytoplasmic tail.

T Cells in the Innate Immune Response

In addition to the broad array of antigens recognized by αβ^+ T cells, there is growing appreciation that the immune system also recognizes pathogen-associated molecular patterns (PAMPs) through a family of Toll-like receptors (TLRs). These TLRs are expressed on a wide variety of cells including macrophages, dendritic cells, and B cells and can recognize pathogen-derived molecules such as lipopolysaccharide (LPS), double-stranded RNA, flagellin, and CpG motifs. The activation of TLRs results in the production of cytokines such as TNF, IL-12, and IFN-γ that can activate other immune cells and contribute to the innate immune response.
**Figure 13-6** T helper subsets. Naïve CD4 T cells can be polarized into producing particular patterns of cytokines depending on the cytokine environment in which they develop, as well as the expression of certain transcription factors. AHR, aryl hydroxylase receptor.

**γδ T Cells**

γδ T cells were identified following serendipitous discovery of rearranged genes while searching for the TCR α-chain gene, rather than a preexisting knowledge of their presence and biologic function. Structurally, the γ-chain locus contains at least 14 Vγ region genes, of which six are pseudo-genes, each capable of rearranging to any of 5 Jγ regions and two Cγ regions. The δ-chain genes are nested within the α-chain gene locus between Vα and Jα. There are about six Vδ regions, two Dδ and two Jδ regions, and a single Cδ gene. Transcription of rearranged γ and δ genes begins before αβ genes and is apparent on days 15 to 17 of mouse thymus development, after which it declines in the adult thymus. In addition to the ordered appearance of TCR-γδ before TCR-αβ, there is also a highly ordered expression of γ and δ V-region genes during early thymic development. This results in successive waves of oligoclonal γδ T cells migrating to the periphery. The reason for this remarkable regimentation remains unclear.

γδ T cells manifest a number of differences from αβ T cells. γδ T cells are often anatomically sequestered to epithelial barriers or sites of inflammation and frequently manifest cytotoxicity toward a broad array of targets. In contrast to αβ T cells, γδ cells can respond to antigen directly, without evidence of MHC restriction, or conversely, react to MHC molecules without peptide. Human γδ T cell clones, particularly those expressing the Vδ2 gene, and derived from peripheral blood of normal individuals, frequently react to nonprotein components of mycobacteria. These have been identified as nucleotide triphosphates, prenyl pyrophosphate, and alkylamines.

The system also contains small subpopulations of T cells that may be specialized to recognize conserved structures that are either uniquely expressed by prokaryotic pathogens or on stressed host cells. These are discussed in detail in Chapter 18. Such common antigenic motifs include bacterial lipoproteins recognized by Toll-like receptor (TLR)-2 and TLR-4, double-stranded RNA from RNA viruses that bind TLR-3, and methylated cytosine residues in bacterial CpG sequences, or anti-DNA/DNA complexes that bind to TLR-9. TLRs expressed by APC can trigger the release of cytokines and co-stimulatory molecules for T cells. Another family of molecules that likely bind bacterial components is CD1. CD1 structurally resembles MHC class I but contains a deeper and more hydrophobic binding pocket that can accommodate certain lipopeptides and glycolipids. By using such molecular strategies to focus on common and nonpolymorphic molecules, the immune response can respond quickly during the early phase of infections. This response is part of the innate immune response. Although it may represent the remnants of an evolutionarily more primitive immune response, it nevertheless provides a vital early defense system. Among T cells this function is provided by γδ and natural killer (NK) T cells.
These molecules are, respectively, subunits in DNA and RNA, substrates in lipid metabolism for the synthesis of farnesyl pyrophosphate, and products of pathogenic organisms. In mammalian cells, farnesyl addition is a critical modification for targeting certain signaling molecules to the cell membrane such as Ras. This process appears critical to cell transformation. These phosphate-containing nonpeptides can be found in both microbial and mammalian cells. This suggests that γδ cells may recognize a class of antigens shared by a number of pathogens, as well as by damaged or transformed mammalian cells, and may provide insight into the role of γδ cells in infection and their accumulation at sites of inflammation. Another subpopulation of γδ T cells expressing the Vγ1 gene is typically found in the intestine and in inflamed synovial fluid and reacts to the MHC class I-like molecules known as MICA and MICB. Unlike classical MHC class I molecules that are expressed ubiquitously and continuously, MICA and MICB expression appears to be restricted to gut epithelium and occurs only during times of stress, similar to that of a heat shock response.

The contribution of γδ T cells to defense against infection has been examined in mice using a number of pathogens including Listeria, Leishmania, Mycobacterium, Plasmodium, and Salmonella. All of these studies have shown a moderately protective role for γδ T cells. The cumulative evidence suggests that γδ T cells may not react directly to components of microorganisms, but rather indirectly via the ability of microbial products to stimulate the innate immune response. An example is the γδ T cells in synovial fluid from Lyme arthritis, which are activated by Borrelia burgdorferi lipitated hexapeptides of the outer surface proteins that stimulate TLR2.

γδ T cells accumulate at inflammatory sites in autoimmune disorders such as rheumatoid arthritis, celiac disease, and sarcoidosis. The reason for this remains an enigma. However, there is evidence that γδ cells can be highly cytolytic toward a variety of tissues including CD4+ T cells, in part due to their high and sustained expression of surface Fas-ligand. Their presence can strongly bias the cytokine profiles of the infiltrating CD4+ cells, in some instances toward Th1 profiles, and in others toward Th2.

**Natural Killer T Cells**

A minor subpopulation of T cells bearing the NK determinant manifest a curiously restricted TCR repertoire. NK T cells are found within the CD4+ and CD48 T cell subsets and, in both mouse and human, express a limited number of TCR-Vβ chains and an invariant α-chain (Vα14 in mice and Vα24 in humans). Furthermore, most NK T cells are restricted in their response to a monomorphic MHC class I-like molecule, CD1d. Crystallographic analysis of CD1d has shown that it contains a deeper groove than traditional MHC molecules and is highly hydrophobic, conferring a preference for binding lipid moieties. Originally, the sea sponge sphingolipid, α-galactosylceramide was the only known CD1d ligand. Now both endogenous and bacterial (Sphingomonas and Borrelia burgdorferi) sources of CD1d-binding sphingolipids have been identified. This may represent another type of innate T cell response whereby bacterial lipids or lipopeptides may be presented to NK T cells to provoke a rapid early immune response.

The potential importance of NK T cells in autoimmune disease stems from their rapid production of high levels of certain cytokines, particularly IL-4 and IFN-γ. In this capacity, the IL-4 response may be important for modulating inflammatory responses dominated by Th1 infiltrates. This has been noted in the nonobese diabetic (NOD) mouse model of diabetes, which has reduced levels of NK T cells. Adaptive transfer of NK T cells into NOD mice blocks the onset of diabetes. A study extended this observation to human type 1 diabetes. The NK T cells of diabetic individuals produced more IFN-γ and less IL-4 than their unaffected siblings. NK T cells have also been reported to be the predominant CD4+ T cell in the airways of asthma patients. Thus this minor population of T cells may play a pivotal role in early innate responses to certain infections and also in the regulation of inflammatory sites.

**Naïve versus Memory T Cells**

CD4+ and CD8 T cells emigrate from the thymus bearing a naïve phenotype. Naïve T cells produce IL-2 but only low levels of other cytokines and as such provide little B cell help. Naïve T cells express high levels of Bcl-2 and can survive for extended periods without antigen but require the presence of MHC molecules. Naïve T cells circulate from the blood to lymphoid tissues of the spleen and lymph nodes, where antigen, APC, T cells, and B cells are concentrated. Particularly important in this environment as APC are dendritic cells, which can migrate from other areas of the body such as the skin and are highly efficient at processing and presenting antigen to T cells (see Chapter 10). Dendritic cells express high and constitutive levels of MHC class II and co-stimulatory molecules B7-1 (CD80) and B7-2 (CD86), which are critical to promoting proliferation of naïve T cells. In this capacity dendritic cells are particularly adept at promoting clonal expansion of antigen-specific T cells. The development of antigen peptide/MHC tetramer technology has led to direct quantitation of antigen-specific CD8 T cells using flow cytometry. Viral-specific CD8 T cells have an incredible ability to expand following infection, from levels that are undetectable to nearly 50% of the CD8 population, representing a nearly 1000-fold increase in only a few days.

During the process of clonal expansion of naïve T cells and their differentiation into effector and eventually memory T cells, up to 100 genes are induced. These are manifest primarily as increased expression of certain surface molecules involved with cell adhesion and migration (CD44, ICAM-1, LFA-1, α4β1 and α4β7 integrins, the chemokine receptor CXCR3); activation (CD45 change from high-molecular-weight CD45RA to lower molecular weight CD45RO isotype); cytokine production (increased production of IFN-γ, IL-3, IL-4, and IL-5); and death receptors (e.g., Fas/CD95) (Table 13-1). More transiently induced are CD69, the survival factor Bcl-xL, and the high affinity growth factor IL-2 receptor α-chain (CD25). Survival of effector T cells to the memory stage is partly dependent on the cytokines IL-7 and IL-15.

The concept of immune memory has existed since the first successful vaccinations by Jenner for smallpox. A useful memory T cell marker in the murine system is CD44, the hyaluronate receptor. Surface CD44 is low on mature
single-positive T cells as they emerge from the thymus, but its expression is upregulated on the first encounter with antigen stimulation in the periphery. Expression of the IL-7 receptor identifies a subset of effector T cells that are destined to become memory T cells. Several other markers have been shown to change on primary antigenic stimulation. Most notable for human T cells is CD45, in which an isoform known as CD45RO is expressed on naïve T cells, whereas CD45RA expression characterizes memory T cells (see Table 13-1). Using these markers it has been possible to identify a variety of differences between naïve and memory T cells. Activation of memory T cells appears to be more efficient than that of naïve T cells and not be absolutely dependent on co-stimulation. Memory T cells are also able to migrate to nonlymphoid tissues such as lung, skin, liver, and joints. Particularly interesting have been recent reports that the metabolic state of an effector T cell may profoundly affect which T cells survive to the memory state. Surprisingly, improved survival of memory T cells was conferred by agents such as rapamycin and metformin, which inhibit anabolic glycolytic metabolism and promote catabolic fatty acid metabolism.

**T Helper Subsets**

CD4 T cells can be further subdivided based on their cytokine profiles. This is the growing list that includes the classic Th1 and Th2 subsets, as well as Th17 cells, follicular helper T cells (Tfh), and regulatory T cells (Treg) (Figure 13-6). Th1 cells participate in cell-mediated inflammatory reactions, activate macrophages, and produce IL-2, TNF, and IFN-γ. Th2 cells produce IL-4, IL-5, IL-6, IL-9, and IL-10. Further subsets of Th2 cells have been described that produce predominantly either IL-9 and IL-10 (Th9 cells, derived with TGF-β and IL-4) or IL-5 (Th5 cells, generated by antigen and IL-33) and are involved with allergic disorders. IL-4 and IL-5 are important B cell growth factors. Th17 cells express a variety of cytokines and their properties is described in Chapter 26. These patterns have been best characterized during chronic infections. In general, a Th1 response helps eradicate intracellular microorganisms such as Leishmania major and Brucella abortus, whereas a Th2 cell response can better control extracellular pathogens such as the helminth Nippostrongylus brasiliensis. The cytokine profiles of Th1 and Th2 cells are mutually inhibitory, such that the Th1 cytokine IFN-γ or IL-12 from APC suppresses Th2 responses and augments Th1 cytokine gene expression, whereas the Th2 cytokines IL-4, or IL-6 from APC, promote the opposite pattern. Polarization of the cytokine environment also occurs at the sites of inflammation in many autoimmune syndromes. Th2 skewing has been observed in models of systemic lupus erythematosus (SLE) in which increased levels of immunoglobulins and autoantibodies are typical, as well as in chronic allergic conditions such as asthma. Frequently, however, the infiltrating lymphocytes exhibit a bias toward Th1 cytokines. This occurs with brain-infiltrating lymphocytes in multiple sclerosis and its animal model, experimental allergic encephalomyelitis (EAE). Th2 cytokines IL-4 and IL-13 promote a variety of autoimmune disorders. Th2 cytokines IL-4 and IL-13 promote a variety of autoimmune disorders. IL-17, IL-23, and IL-22 are important growth factors for Th17 cells. IL-23 also provides a survival signal for Th17 cells and promotes differentiation of Th17 cells from naïve T cells. A more recently described subset of IL-17-producing CD4+ T cells (Th17) has emerged as being critical for promoting a variety of autoimmune disorders. TGF-β, possibly originating from Treg, accompanied by IL-6 (probably from dendritic cells), appears to be pivotal for the appearance of Th17 cells. IL-23 may also be important for the survival, though perhaps not the appearance, of Th17 cells. Injections of IL-23 into skin produced increased IL-17 in the

### Table 13-1 Surface Markers on Naïve and Memory T Cells

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Other Designation</th>
<th>Molecular Weight (kD)</th>
<th>Characteristic</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD58</td>
<td>LFA-3</td>
<td>45-66</td>
<td>Ligand for CD2</td>
<td>Memory + Naïve +</td>
</tr>
<tr>
<td>CD2</td>
<td>T1</td>
<td>50</td>
<td>Alternative activation pathway</td>
<td>Memory + Naïve +</td>
</tr>
<tr>
<td>CD11a/CD18</td>
<td>LFA-1</td>
<td>180-195</td>
<td>Receptor for ICAM-1, ICAM-2, ICAM-3</td>
<td>Memory + Naïve +</td>
</tr>
<tr>
<td>CD29</td>
<td></td>
<td>130</td>
<td>β chain of β1 (VLA) integrins</td>
<td>Memory + Naïve +</td>
</tr>
<tr>
<td>CD45RO</td>
<td></td>
<td>220</td>
<td>Isoform of CD45</td>
<td>Memory + Naïve +</td>
</tr>
<tr>
<td>CD45RA</td>
<td></td>
<td>80-95</td>
<td>Isoform of CD45</td>
<td>Memory + Naïve +</td>
</tr>
<tr>
<td>CD44</td>
<td>Pgp-1</td>
<td>90</td>
<td>Receptor for hyaluronic acid</td>
<td>Memory + Naïve +</td>
</tr>
<tr>
<td>CD54</td>
<td>ICAM-1</td>
<td>120</td>
<td>Counterreceptor for LFA-1</td>
<td>Memory + Naïve +</td>
</tr>
<tr>
<td>CD26</td>
<td></td>
<td>40</td>
<td>Dipeptidyl peptidase IV</td>
<td>Memory + Naïve +</td>
</tr>
<tr>
<td>CD7</td>
<td></td>
<td>Multichain complex</td>
<td>T cell lineage marker</td>
<td>Memory + Naïve +</td>
</tr>
<tr>
<td>CD3</td>
<td></td>
<td></td>
<td>Part of TCR complex</td>
<td>Memory + Naïve +</td>
</tr>
</tbody>
</table>

CD, cluster of differentiation; ICAM, intercellular adhesion molecule; LFA, leukocyte function–associated antigen; TCR, T cell antigen receptor; VLA, very late activation antigen.
epidermis and inflammatory lesions that resembled psoriasis. \textsuperscript{102} Th17 cells are increased in human psoriatic plaques, \textsuperscript{102} in rheumatoid arthritis synovial fluid, and in multiple sclerosis. \textsuperscript{103} An additional CD4 T helper subset is found in B cell follicles of secondary lymphoid organs due to their expression of the B cell follicle homing receptor CXCR5. \textsuperscript{104} These follicular helper T cells (Tfh) assist B cell activation and germinal center formation through expression of CD40L, IL-4, and IL-21. Dysregulation of Tfh function can result in autoantibody production and systemic autoimmunity. \textsuperscript{105} Finally, CD4\textsuperscript{+}FoxP3\textsuperscript{+} regulatory T cells (Treg) have been described earlier.

**Molecular Mimicry**

Perhaps the oldest concept in autoimmune mechanisms is that of molecular mimicry, the notion that the response of the immune system to a foreign substance may provoke cross-reactivity to a self-protein. This is best established in rheumatic heart disease, where a B cell antibody response to a group A streptococcal cell wall component can precipitate a cross-reactivity to cardiac myosin. Similarly for T cells, investigators used the peptide sequence of myelin basic protein (MBP) recognized by specific T cell clones from patients with multiple sclerosis to search a database of infectious agents. Some of the candidate sequences obtained were able to stimulate the MBP-reactive T cell clones. \textsuperscript{106} This suggested for the first time that T cells responding to an infectious agent might manifest cross-reactivity to self-peptides. Another example is one of the outer surface proteins of \textit{B. burgdorferi} known as OspA, which may trigger a cross-reactive T cell response. \textsuperscript{107} A T cell immunodominant peptide of OspA was identified in a subset of HLA-DR4 patients who are more resistant to antibiotic treatment and manifest an antibody, as well as a T cell response, to OspA. \textsuperscript{108} Using a sequence algorithm to identify homologous peptides that bind the DR4 pocket, a sequence in LFA-1 that bound DR4 and stimulated a T cell response from these OspA-reactive patients was identified. \textsuperscript{107} The technology of peptide/MHC tetramers discussed earlier will enable investigators to determine whether a given subpopulation of T cells within an inflammatory synovium might manifest dual specificity for both a foreign pathogen and a cross-reacting self-protein.

**Death of T Cells**

The rapid removal of the effector T cells following clearance of the infection is as important as the initial clonal expansion of responding T cells for the health of the organism. Failure to clear activated lymphocytes increases the risk of cross-reactivity with self-antigens and a sustained autoimmune reaction. To ensure that resolution of an immune response occurs rapidly, a number of processes promote active cell death of clonally expanded T cells. One means to control T cell proliferation is through limited availability of growth factors. On activation, T cells express receptors for various growth cytokines for approximately 7 to 10 days but only produce cytokines for a more limited period. This results in an unstable situation where T cells tend to outgrow the availability of cytokines. T cells expressing IL-2R, for example, in the absence of IL-2 will rapidly undergo programmed cell death. Another method is restimulation of TCR on actively dividing T cells, which triggers a death cascade known as activation-induced cell death (AICD).

The discovery of a family of death receptors expressed by T cells elucidated an additional regulatory process. These molecules are described more extensively in Chapter 27 on cell survival and are only discussed here as they relate to T cell function. The best described of these is Fas (CD95). Both Fas-deficient mice and humans bearing Fas mutations (Canele-Smith syndrome) \textsuperscript{109} manifest a profound lymphadenopathy accompanied by an autoimmune diathesis. This underscores the importance of efficiently removing T cells after their activation. Nearly all cells have some level of surface Fas, whereas expression of its ligand (FasL) is restricted primarily to activated T cells and B cells. Consequently, regulation of Fas-mediated apoptosis is to a large extent under the governance of the immune system. FasL expression has also been reported for certain components of the eye, the Sertoli cells of the testis, and perhaps some tumors. \textsuperscript{110} Expression of FasL by these nonlymphoid cells is thought to prevent immune responses at sites where such inflammation might cause tissue damage. For years immunologists have been aware of these so-called “immune privileged” sites within which immune responses are difficult to initiate.

During T cell activation, expression of FasL is rapidly induced and the ability to kill Fas-sensitive target cells is easily demonstrated. Yet expression of surface FasL protein has been difficult to demonstrate. This may be due to a sensitivity of surface FasL to certain proteinases, which results in its rapid cleavage and release from the cell, in a manner similar to the release of another member of the Fas family, TNF. Resting T cells are not sensitive to Fas-induced death but must first enter the cell cycle for approximately 3 days. During this period, the cellular level of an endogenous inhibitor of Fas known as c-FLIP is downregulated and this presumably allows Fas signaling to progress. \textsuperscript{111} Thus c-FLIP may function to protect resting T cells from unnecessary death and restrict apoptosis to activated T cells in order to limit their expansion.

The sequence of T cell activation followed by cell death is graphically displayed following the administration to mice of bacterially or virally derived compounds called superantigens. Superantigens activate T cells by directly crosslinking MHC class II molecules with particular β-chain V families of the TCR (see Figure 13-3A). The superantigen staphylococcal enterotoxin B (SEB) strongly activates Vβ8\textsuperscript{+} T cells. \textsuperscript{112} This initiates a rapid expansion of Vβ8\textsuperscript{+} T cells over 2 to 3 days followed by an equally rapid loss of these cells, such that by day 7 few Vβ8\textsuperscript{+} T cells remain. A similar process of T cell activation occurs in the human disease toxic shock syndrome in which a related staphylococcal toxin, TSST, stimulates the expansion of Vβ2\textsuperscript{+} T cells. \textsuperscript{113} The devastating illness that results from this profound activation of a large proportion of T cells underscores the need to rapidly eliminate activated T cells. At least some of the damage in toxic shock syndrome likely results from the extensive T cell expression of FasL and TNF, particularly in certain tissues such as the liver. Hepatocytes are exquisitely sensitive to damage by these ligands. Activation of T lymphocytes leads to homing to the liver, and administration of antigen to TCR transgenic mice can yield a syndrome
resembling autoimmune hepatitis.114 Thus certain autoimmune disorders may result from the death or damage of “innocent bystander” cells as a consequence of the migration of activated T cells to an organ and nonspecific damage due to the expression of FasL family members.

Although it was initially assumed that Fas would largely regulate this process, Fas-deficient mice eliminate T cells activated by either traditional antigens or superantigens nearly as efficiently as wild-type mice. Rather, proapoptotic members of the Bcl-2 family, Bim, Bad, and Bax, appear to regulate death in vivo from cytokine withdrawal or following acute foreign antigen stimulation as with certain infections.115 These molecules are related to the cell survival molecule, Bcl-2, but are more truncated, containing only the BH3 domain of Bcl-2, and hence their designation as the “BH3-only” family. They function as sentinels within the cell in that they are attached to various cytoskeletal proteins and organelles and sense cellular damage. If damage occurs, they are released from these sequestered areas and migrate to the mitochondria to inhibit the survival function of Bcl-2.115 By contrast, Fas serves to eliminate T cells following chronic TCR stimulation as occurs in homeostatic proliferation or chronic infections.116,117

**T Cells at Sites of Inflammation**

The observation that T cells infiltrate target organs in autoimmune diseases such as rheumatoid arthritis, type 1 diabetes mellitus, and multiple sclerosis quickly led to an analysis of T cell subsets and, more recently, to a more detailed study of the T cell repertoire based on TCR expression. In many of these disorders, the known HLA class II association is paralleled by a predominant, though by no means exclusive, infiltration of CD4+ T cells bearing a broad TCR repertoire.118 Many of these CD4+ T cells also manifest a Th1-like cytokine profile as discussed earlier. Evidence for the importance of these CD4+ T cells derives from numerous studies showing the efficacy of CD4 depletion in animal models of these disorders.119 A parallel situation has often occurred in humans with these autoimmune diseases who concurrently become infected with HIV. The CD4 depletion occurring during AIDS can actually ameliorate the course of these disorders, suggesting the CD8+ T cell subset of T cells may be important in these disorders.120 Given the previously mentioned effect of the metabolic state of T cells on their survival, it will be highly informative in future studies to examine the metabolic state of T cells at sites of autoimmune tissue inflammation versus that observed in secondary lymphoid tissues during infections. Conceivably the cytokine environment of inflamed tissues may confer a metabolic state that favors T cell survival compared with states of infection.

**References**

116. Fortner KA, Bad RC: The death receptor Fas (CD95/APO-1) mediates the deletion of T lymphocytes undergoing homeostatic proliferation, J Immunol 175:4374–4382, 2005.

The references for this chapter can also be found on www.expertconsult.com.
B Cells

NATALY MANJARREZ ORDUÑO • CHRISTINE GRIMALDI • BETTY DIAMOND

OVERVIEW: B CELLS AND HUMORAL IMMUNITY

Numerous cells comprise the immune system and are required to generate innate and adaptive immune responses. Adaptive responses are characterized by immunologic memory generated during first exposure to antigen, thereby permitting a rapid response to antigen following subsequent exposure.

B cells (Figure 14-1) are lymphocytes that recognize antigens through a molecule called the B cell receptor. The B cell receptor is composed of a surface immunoglobulin molecule, which recognizes the antigen, and two associated proteins, which transduce the signal. On encounter with its antigen, a B cell begins a process of activation leading to antibody secretion and memory formation regulated by interplay with antigen-activated T cells, dendritic cells, and soluble factors, and in some cases follicular dendritic cells. Both T and B lymphocytes can differentiate from naive to memory cells, but only B cells have the capacity to fine tune their antigen receptor structure to increase its specificity and affinity, giving rise to more effective antibodies. Beyond immunoglobulin secretion, B cells regulate the immune response by cytokine secretion and antigen presentation to T cells in the context of class II molecules.

Importantly, much of the knowledge of B cell biology has been generated in mouse models. However, in this chapter, human B cell biology is described whenever possible.

IMMUNOGLOBULINS: STRUCTURE AND FUNCTION

The Ig molecule is critical to all aspects of B cell biology when anchored to the B cell membrane. Termed B cell receptor (BCR), it contributes to B cell maturation and survival and initiates an activation cascade following contact with antigen. At the end of the activation process, B cells can acquire the ability to secrete large amounts of Ig, intended to neutralize the antigen that elicited the response.

Structurally, Igs, also referred to as antibodies, are composed of four polypeptide chains: two identical light (L)-chains with a molecular weight of approximately 25 kDa and two identical heavy (H)-chains of 50 to 65 kDa. Each of the chains contains a folding motif that is highly conserved among proteins of the immune system, the “Ig domain.” These domains constitute the backbone of the Ig molecule and make for the interface along which the polypeptide chains pair (Figure 14-2). The quaternary structure of an immunoglobulin molecule assumes a Y-shaped conformation, which contains two functional moieties: two identical antigen-binding regions or variable regions, which are the arms of the “Y,” and a constant region, which is the base of the “Y.”

This definition of functional moieties derives from early studies analyzing proteolytic fragments of Ig molecules. Cleavage with papain generates two identical fragments that retain antigen-binding capacity and hence are named Fab, as well as a distinct crystallizable fragment Fc that mediates immune effector functions but is unable to interact with antigen.”

The antigen-binding regions are formed by pairing of the variable domain of the L-chain (\(V_L\)) to the variable domain of the H-chain (\(V_H\)). In contrast to the rest of the molecule, there is a great diversity in the amino acid sequence of the variable domains, which allows for a broad repertoire of antibody molecules that can recognize a wide array of antigens. Within the variable region of the Ig molecule are discrete regions, known as complementary determining regions...
The H- and L-chain molecules each consist of the hinge domain (only in IgG, IgA, and IgD), which increases flexibility and C domains, presence of a hinge region to increase flexibility between Fab regions, serum half-life, ability to form polymers, complement activation, and Fc receptor binding vary among isotypes. Characteristics of the different Ig H-chain isotypes are presented in Table 14-1. These isotypes may also differ in the intracellular signaling they initiate when bound by antigen in their membrane-associated form on the B cell. It should be noted that the interplay between antibodies and the cells that bear the Fc receptors extends beyond pathogen clearance and shapes the immune response by mediating activation or inhibition of specific cell types and by mediating cell death.

**Immunoglobulin M**

IgM is the first isotype expressed in developing B cells and the first antibody secreted during a primary immune response. It is found predominantly in serum but is also present in mucosal secretions and breast milk. Because the process that increases antibody affinity for a particular antigen (affinity maturation) has not yet been initiated during the early stages of a primary immune response, IgM antibodies usually exhibit low affinity. Their low affinity is balanced by the fact that most of the secreted IgM exists in pentameric form, generating multiple binding sites.
IgM also exists as a monomer and a hexamer, but only the pentameric form is linked by the polypeptide called joining (J) chain. The J chain allows the active transport of IgM to mucosal secretions.

Many of the biologic functions of IgM are mediated by its ability to activate the classic complement pathway. The complement cascade is composed of a series of enzymes that, on activation, mediate the removal and lysis of invading organisms. Deposition of antibody molecules or complement components on the surface of the antigen assists phagocytosis. Proteins such as antibody and complement that enhance phagocytosis are called opsonins. Once the complement cascade has been activated, monocytes, macrophages, or neutrophils engulf opsonized particles through specific receptors present on phagocytic cells such as CD21, which recognizes fragments of the C3 complement component. Activation of the complement pathway also results in the generation of the membrane attack complex, which is composed of late complement components and directly lyses C3-opsonized pathogens. Because activation of the classic complement pathway requires Fc regions to be spatially close, but also exposed, multimeric IgM is a potent activator of the classic complete pathwasy once it has bound its antigen. For example, hexameric IgM is between 20 and 100 times more potent as an inducer of complement activation than monomeric IgM.

Table 14-1 Properties of Human Immunoglobulin (Ig) Isotypes

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>IgM</th>
<th>IgG</th>
<th>IgA</th>
<th>IgE</th>
<th>IgD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>Pentamer, hexamer</td>
<td>Monomer</td>
<td>Dimer (IgA&lt;sub&gt;1&lt;/sub&gt;), monomer (IgA&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>Monomer</td>
<td>Monomer</td>
</tr>
<tr>
<td>C&lt;sub&gt;H&lt;/sub&gt; domains</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Serum values (mg/mL)</td>
<td>0.7-1.7</td>
<td>9.5-12.5</td>
<td>1.5-2.6</td>
<td>.0003</td>
<td>.04</td>
</tr>
<tr>
<td>Serum half-life (days)</td>
<td>5-10</td>
<td>7-24</td>
<td>11-14</td>
<td>1-5</td>
<td>2-8</td>
</tr>
<tr>
<td>Complement activation (classic)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>FcR-mediated phagocytosis</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Antibody-dependent cell-mediated cytotoxicity</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Placental transfer</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Presence in mucosal secretions</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Main biologic characteristic</td>
<td>Primary antibody response</td>
<td>Secondary antibody responses</td>
<td>Secreted immunoglobulin</td>
<td>Allergy and parasite reactivity</td>
<td>Marker for naïve B cells</td>
</tr>
</tbody>
</table>

IgG is the most common isotype found in serum, comprising about 70% of the circulating antibody. IgG antibodies are usually of higher affinity than IgM antibodies and predominate in a secondary or memory immune response.

In humans there are four subclasses of IgG: IgG1, IgG2, IgG3, and IgG4. IgG1 and IgG3 arise in response to viral and protein antigens. IgG2 is the main antibody present in response to poly saccharide antigens, and IgG4 participates in responses to nematodes and is observed in chronic antigen stimulation.

All IgG subclasses exist as monomers and have a high structural similarity; however, minor differences make for distinct biologic effects. IgG3 and IgG1 are potent activators of the classic complement pathway, and IgG2 can initiate the alternative complement pathway (see Chapter 23).

All IgG subclasses engage specific Fc gamma receptors (FcγRs) present on dendritic cells, macrophages, neutrophils, and NK cells. The FcγRs on phagocytic cells, when cross-linked, mediate the removal of immune complexes from circulation and initiate antibody-dependent, cell-mediated cytotoxicity resulting in the release of granules that contain perforin, a pore-forming protein, and enzymes known as granzymes that induce programmed cell death (apoptosis) of target cells.

Because IgG antibodies are the only ones that cross the placental barrier, they are critical for the survival of newborns. The transport of IgG from the maternal circulation into the fetal blood supply is mediated by the FcRn receptor. FcRn is also responsible for the long half-life of IgG in serum by blocking IgG catabolism.

**Immunoglobulin A**

Despite its relative low concentration in serum, more IgA is produced than all other isotypes combined. Most IgA exists as secretory IgA (SIgA) in mucosal cavities and in milk and colostrum, and only a small fraction is present in serum. Two subclasses of IgA exist in humans: IgA1 and IgA2. IgA1 is mainly produced as a monomer. In contrast, polymeric IgA2 is produced along mucosal surfaces.

Polymeric IgA exists mainly as a dimer and includes a J-chain (the same chain that links pentameric IgM). It is captured by the polymeric immunoglobulin receptor (plgR) that is expressed on the basolateral surface of the epithelial cells and then transcytosed to the apical side. Release of IgA into mucosal secretions requires cleavage of the plgR, a fragment known as the secretory component (SC) remains associated with secreted IgA and protects it from the action of proteases and increases its solubility in mucus, where it neutralizes toxins and inhibits the adherence of secreted IgA-coated microorganisms to the mucosal surface.

People with IgA deficiency have reduced levels of both serum and secreted IgA and are prone to respiratory tract
and diarrheal infections, as well as an increased incidence of autoimmune disorders. FcεR, present in the surface of neutrophils and macrophages, has been suggested to play a regulatory role in the immune system; it is not clear if the autoimmune manifestations in IgA-deficient patients are a consequence of the absence of the regulatory loop played by engagement of FcεR.

**Immunoglobulin E**

IgE is involved in protection against parasitic infections but also triggers immune responses associated with allergic reactions. Only a small amount of IgE is detectable in serum, where it exists as a monomer. Mast cells and basophils express a high-affinity IgE Fc receptor (FcεRI) that binds free IgE. Cross-linking of the FcεRI by antigen binding to the IgE induces degranulation and release of histamine, proteases, lipid mediators such as prostaglandin D₂ and leukotrienes, many of which are associated with anaphylaxis.

**Immunoglobulin D**

The role of IgD in the humoral response has been the subject of multiple speculations. IgD is found predominantly as a membrane immunoglobulin on the surface of mature naive B cells. Soluble IgD is scarce in serum; however, IgD-producing plasma cells are found in tonsils and tissue associated with the respiratory tract. High levels of secreted IgD can be found in individuals with immunodeficiencies.

**Light Chains**

There are two distinct L-chain polypeptides, designated kappa (κ) and lambda (λ). L-chains contain a variable and a single constant domain. Even though there are two L-chain isotypes, there is no known function associated with the L-chain constant region. The κ chain is used more often than the λ chain in human (65%) and mouse (95%) Ig molecules.

**Immunoglobulin Variable Region**

The recognition of a virtually unlimited number of antigens requires a mechanism to generate Ig molecules with similarly broad specificities. The molecular basis of this process has been known for several years. First, the Ig molecule is composed of both heavy and light chains. These chains are encoded within distinct genetic loci residing on separate chromosomes; the H-chain locus is on human chromosome 14, the κ-chain locus is on chromosome 2, and the λ-chain locus is on chromosome 22.

Within each locus, gene segments encode both the variable and constant regions of the Ig molecule. The H-chain variable region is encoded by a variable (V_H), diversity (D_H), and a joining (J_H) segment. The L-chain is encoded by either V_κ or V_λ or J_κ segments; it does not contain D segments. The human H-chain locus contains from 38 to 46 V_H, 23 D_H, and 9 J_H functional genes (these numbers represent a typical haplotype, but vary among individuals). The κ-chain locus contains approximately 31 to 35 V_κ genes and 5 J_κ functional genes; the λ-chain locus contains 29 to 32 V_λ genes and 4 or 5 J_λ functional genes.

**Generation of Immunoglobulin Diversity**

In a developing B cell, different V_H, D_H, and J_H or V_λ and J_λ gene segments are randomly combined to generate a large number of different Ig molecules (Figure 14-3). This process, known as V(D)J recombination, occurs in the primary lymphoid tissue, in the absence of antigen stimulation and must be successful to continue with B cell maturation. Here, we present the molecular process, and later the functional and developmental consequences are discussed (see B Cell Development).

V(D)J recombination happens sequentially, beginning with the joining of one D_H segment to one J_H. Following this, a V_H segment will be targeted to the rearranged D_HJ_H fragment. The absence of an in-frame recombination leads to recombination of the second allele. Light chain recombination also occurs stepwise. First, the kappa locus is rearranged; in absence of a productive κ-chain rearrangement, the lambda locus undergoes recombination.

The recombination machinery is composed of specific enzymes including Recombination-Activating Gene 1 (RAG-1) and the Recombination-Activating Gene 2 (RAG-2). The complex recognizes recombination signal sequences (RSS) that flank the V, D, and J gene segments. These highly conserved sequences are composed of a palindromic heptamer (7 base pairs) followed by deoxyribonucleic acid (DNA) spacers that are 12 or 23 base pairs in length and an AT-rich nonamer. Once the complex recognizes its target, it generates double-stranded DNA (dsDNA) breaks at the RSS sites. Next, the cellular DNA repair complex recognizes and joins the cleaved segments.

The random recombination that occurs among V, D, and J gene segments can generate a diverse Ig repertoire without the need for a large number of germline H- and L-chain genes. During H-chain recombination, nucleotides may be added at V_HD_H and D_HJ_H junctions by the enzyme terminal deoxynucleotidyl transferase (TdT). These non-germline-encoded sequences are known as N additions. As long as these nucleotide changes do not disrupt the reading frame or lead to the incorporation of premature stop codons, the random addition of N sequences increases the diversity of the amino acid sequence. Further, imprecise ligation at the coding junctions may result in the loss of nucleotides, thereby also enhancing diversity.

**B CELL DEVELOPMENT**

The aim of the maturation process is to generate a pool of mature B cells with a diverse repertoire of Ig specificities that can recognize foreign and pathogenic antigens without compromising the integrity of the self. Therefore the process of generation of Ig diversity is coupled with the censoring of autospecificities.

B cells, as all cells in the hematopoietic lineage, begin with the differentiation of noncommitted, undifferentiated CD34⁺ hematopoietic stem cells to lymphopoietic precursors with restricted lineage potential. These cells, known as common lymphoid progenitors (CLP), have the potential to give rise to NK, T, and B cells. Early B cell progenitors begin to express genes for DNA rearrangement, as well as B cell program transcription factors. Multiple transcription factors act in concert, but Ikaros, E2A, EBF, and Pax5 appear to be
the most important in B cell development.\textsuperscript{27} Pax5 is considered the master transcriptional control for B cells because it is induced in early stages of B cell commitment and plays a dual role by repressing genes required for differentiation to the myelomonocytic lineage and activating B cell–specific genes such as Ig genes, CD19, and signaling molecules.\textsuperscript{28}

**Niches of Human B Cell Lymphopoiesis**

The development of undifferentiated hematopoietic stem cells (CD34\textsuperscript{+}) into mature B cells begins in the first weeks of uterine life. By the eighth gestational week, early B cell precursors can be identified in the fetal liver and omentum. From gestational week 34 and through adulthood, the bone marrow is the primary site of lymphopoiesis.\textsuperscript{29}

It has been unequivocally established that there are differences between the B cells that originate during fetal and adult lymphopoiesis in mice, and it is becoming clear that these differences extrapolate to human lymphopoiesis as well. B cell precursors are susceptible to estrogen, and the maturation of maternal B cells is arrested in the pro–B cell stage during pregnancy; in contrast, fetal B cell precursors lack estrogen receptors and consequently are unaffected by exposure to hormones.\textsuperscript{30} B cells originating during prenatal life have a bias in the usage of D\textsubscript{H} and J\textsubscript{H} gene segments, and this, along with reduced expression of the enzyme TdT, leads to a more restricted repertoire with shorter CDR3s.\textsuperscript{31}

Whether during fetal or adult lymphopoiesis, the maturation of B cells from pluripotent precursors is contingent on the presence of stromal cells that provide both contact-dependent and soluble signals. Although the nature of the interactions provided by the stromal cells to create a lymphopoiesis-permissive environment is still largely unknown, they include both survival and proliferative signals. The stromal-derived factor-1 (SDF-1) is required for the homing of the early B cell precursors to sites of lymphopoiesis but also promotes differentiation.\textsuperscript{32}

Interactions between VCAM-1 on the membrane of the stromal cells and its counterpart, VCAM-4, on early B cell progenitors are required for B cell differentiation. The molecules IL-7, IL-3, and the Flt3 ligand promote B cell lymphopoiesis, although IL-7 appears to be dispensable for human B cell development. Matrix molecules in the microenvironment such as heparan sulfate proteoglycan are assumed to “trap” critical soluble factors.\textsuperscript{33}

**B Cell Ontogeny**

The stages of B cell development are defined by the state of Ig gene rearrangement and the expression of intracellular and surface proteins. The nomenclature and classification of particular stages vary slightly among different laboratories working in this field. For simplicity, we have divided the stages of B cell lymphopoiesis into early B cell progenitors, pro-B, pre-B, immature, transitional, and mature naïve cell (Table 14-2).

When a CLP begins to transcribe RNA coding for proteins required for B cell maturation, E2A and EBF, the cell becomes an early B cell progenitor. Once these two transcription factors are expressed, they enable transcription of the proteins involved in the recombination machinery (RAG1/2). The beginning of D to J recombination marks the progress to a pro-B stage.
**Table 14-2 Human B Cell Maturation Markers during B Cell Development**

<table>
<thead>
<tr>
<th>Marker</th>
<th>HSC</th>
<th>Pro-B</th>
<th>Pre-B</th>
<th>Immature</th>
<th>Transitional 1</th>
<th>Transitional 2</th>
<th>Plasma</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD34</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>CD19</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CD10</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CD20</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>CD21</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>CD22</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>CD23</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>CD38</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>CD40</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>CD45</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CD138</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>RAG-1</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>RAG-2</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>TdT</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Igx</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Igβ</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Heavy chain</td>
<td>−</td>
<td>(D&lt;sub&gt;H&lt;/sub&gt;−J&lt;sub&gt;H&lt;/sub&gt;)</td>
<td>+ (V&lt;sub&gt;H&lt;/sub&gt;−D&lt;sub&gt;H&lt;/sub&gt;−J&lt;sub&gt;H&lt;/sub&gt;)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Pre-BCR</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Surface IgM</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Surface IgD</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Light chain</td>
<td>−</td>
<td>−</td>
<td>+ (V&lt;sub&gt;H&lt;/sub&gt;−J&lt;sub&gt;H&lt;/sub&gt;, V&lt;sub&gt;L&lt;/sub&gt;−J&lt;sub&gt;L&lt;/sub&gt;)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
</tbody>
</table>

**Pro-B Cells**

This stage is defined by the rearrangement of the heavy chain gene segments and synthesis of a β-polypeptide. Pro-B cells are dependent on interactions with endothelial cells present in the stroma. The VLA-4 integrin receptor and CD44, which both mediate adhesion to stromal cells, are highly expressed at this stage and are believed to be important for continued development. Pro-B cells also express high levels of Bcl-2, a molecule that protects cells from apoptosis. At the onset of the pro-B cell stage, the variable gene segments of both H- and L-chain loci are in the unarranged germline configuration but accessible to the recombination machinery. A D<sub>H</sub> gene segment on one H-chain chromosome rearranges with a J<sub>H</sub> gene segment residing on the same chromosome, often with the inclusion of nontemplated nucleotides at the junction of these two segments. Next, a V<sub>H</sub> gene rearranges to the D<sub>H</sub>J<sub>H</sub> gene fragment. Completion of V<sub>H</sub>D<sub>H</sub>J<sub>H</sub> gene rearrangement leads to the generation of an H-chain transcript that also contains the IgM constant region (C<sub>μ</sub>), which is the constant region gene most proximal to the variable region genes on the chromosome (see Figure 14-3). The generation of a β-polypeptide and its subsequent expression on the surface of the cell, together with a surrogate light chain formed by the λ5 and Vpre-B polypeptides as well as the Igα/Igβ dimer, a complex known as the pre-B cell receptor (pre-BCR), marks the end of this phase of gene recombination. This constitutes a critical developmental checkpoint and the entrance to the next developmental stage known as the pre-B cell. The requirement for a pre-BCR complex ensures that B cells without a productive H-chain will not undergo further differentiation.

The pre-BCR transduces a signal that the VDJ rearrangement was successful and halts recombination of the second H-chain allele. This process, known as allelic exclusion, ensures that all Ig molecules generated within a single B cell are identical and have the same antigenic specificity. If no α-chain is generated, rearrangement is initiated on the other chromosome. If the second rearrangement also results in a nonproductive H-chain molecule, the absence of pre-BCR-mediated signal induces apoptosis. The odds of generating a productive rearrangement are one in three, and consequently, approximately 50% of the cells that begin recombination will be unable to proceed along a developmental pathway.

**Pre-B Cells**

The pre-B cell stage is characterized by light chain recombination. Initiation of this stage requires the presence of the pre-BCR and functional signal transduction machinery.

At the onset of the pre-B cell stage, the expression of the pre-BCR induces a proliferative burst that generates daughter cells with the same heavy chain and potential for multiple specificities within daughter cells, each of which may produce a different light chain. It is unknown if expression of the pre-BCR and tonic signaling is enough to trigger these events or if a ligand must engage the pre-BCR. Either way, targeted disruption of genes encoding the pre-BCR complex such as the IgM transmembrane constant region domain, λ5, or the Igα and Igβ accessory molecules results in a profound decrease in developing B cells. Also, defects in the adapter molecule BLNK, λ5, or the tyrosine kinase Btk lead to a serious impairment in pre-B cell maturation.

The expression of the pre-BCR is transient. After the proliferative burst, the µ-heavy chain is present only in the cytoplasm as the pre-B cell rearranges an L-chain. The general rearrangement process is similar to V(D)J rearrangement and is dependent on Rag1/2 expression. Because TdT is not expressed in this stage, L-chains do not usually contain N sequences at the V<sub>L</sub>J<sub>L</sub> junction. At the end of this process, pairing of the newly minted light chain with the µ-heavy chain leads to the surface expression of an IgM molecule, complexed with Igα and Igβ, to form the B cell.
receptor (BCR). It is believed that surface expression of the BCR on immature B cells transduces signals that enforce allelic exclusion at the L-chain locus and downregulate expression of the RAG genes. This immature B cell has completed the gene rearrangement process and is now subject to repertoire selection.39

**Immature B Cells**

Once B cells express surface IgM in the bone marrow, they are subject to repertoire censoring. During this stage, cross-linking of the BCR by antigen leads to the activation of one of several tolerance mechanisms. These mechanisms include deletion, receptor editing, and anergy and diminish the fraction of autoreactive cells present in the mature repertoire (see later discussion on negative selection).

During the maturation process in the bone marrow, the cells become less dependent on interactions with the stroma and move toward the bone cavity. Once they express IgM, they begin to move into blood, where they are called transitional cells.

**Peripheral B Cell Subsets**

As B cells mature and their dependence on stromal cells decreases, they leave the bone marrow and finish their maturation in the spleen before homing to other lymphoid tissue such as the lymph nodes, tonsils, and Peyer's patches of the intestine. It is in these secondary lymphoid organs where mature B cells interact with foreign antigen and specific humoral immune responses are activated.

**Transitional B Cells**

Once immature B cells egress from the bone marrow, they are called transitional cells. These cells are the earliest B cells found in the periphery in healthy subjects and move to the spleen to finish their maturation.

Transitional cells are the last B cell subpopulation that expresses the developmental marker CD24. At this stage B cells begin to express surface IgD, which harbors the same specificity as the IgM because the IgD heavy chain is encoded by the same VDJ fragments as IgM but expresses the C4d instead of the C4b domain. It is the expression of IgD that separates transitional cells into two different maturation stages. Transitional I (T1) B cells that do not express IgD are the recent bone marrow emigrants, and transitional 2 (T2) B cells that begin to express IgD represent the subsequent maturation stage.40 The existence and functional characteristics of a third transitional stage (T3) are still debated.

Transitional cells constitute a stage subject to multiple regulatory processes. First, transitional cells must compete with naïve B cells already present in the periphery for a developmental niche. Transitional B cells are extremely dependent on a B cell survival factor known as B cell–activating factor of the tumor necrosis factor family (BAFF or BlyS). In its absence, B cell development does not progress beyond the T1 stage.41 Second, T1 transitional cells are still highly susceptible to tolerance induction following BCR cross-linking. In T1 cells cross-linking of the BCR ex vivo has been shown to lead to cell death, whereas T2 cells respond to BCR cross-linking by proliferation and differentiation to the mature naïve B cell stage.

**BAFF Family of Cytokines**

The cytokine milieu surrounding the B cell is diverse and spatially and temporally regulated. Although B cells are modulated by multiple cytokines, in recent years two members of the TNF family, BAFF and APRIL, have emerged as key survival factors, particularly at two regulatory points in development and differentiation: the transition from an immature to a naïve B cell in the periphery and the survival of the newly produced plasma cells. BAFF (B cell–activating factor, BlyS) and APRIL (A proliferation-inducing ligand) are proteins produced by cells that take part in the innate response such as macrophages and dendritic cells, as well as stromal cells, and are present as membrane-bound proteins or soluble trimers. They have three known receptors (BAFF-R, TACI, and BCMA) that are present on the membrane of B cells from the T2 stage to their final differentiation to plasma cells. BAFF binds the three receptors, whereas APRIL binds only TACI and BCMA. BAFF induces survival and activation of B cells when bound to BAFF-R, whereas BAFF signaling through TACI decreases the size of the B cell pool. APRIL does not participate in B cell homeostasis but seems critical to the survival of plasmablasts in the bone marrow.42

Enhanced survival and activation of autoreactive B cells have been demonstrated in mice that overexpress BAFF.43 An increase in serum levels of BAFF has been observed in some patients with lupus, rheumatoid arthritis, and Sjögren's syndrome and is thought to contribute to pathogenesis because autoreactive B cells have a survival advantage in the presence of excess BAFF44 (Figure 14-4).

**Naïve B Cells**

The final stages of maturation that occur in the spleen and give rise to the naïve B cell subset have not been fully elucidated, but the prevailing theory is that T2 B cells give rise to the circulating mature naïve cell population. In the mouse, two populations of phenotypically and functionally different naïve B cells are recognized: follicular and marginal zone B cells. Human naïve B cells comprise 60% to 70% of the circulating B cell repertoire and populate the spleen and lymph nodes. They include the equivalent of mouse follicular B cells and represent the circulating, non–antigen exposed B cell subpopulation characterized by surface IgM and IgD expression, lack of CD27, and the presence of the membrane transporter ABC.45 There is also a population in blood of IgM+, CD27+ B cells that have been likened to marginal zone B cells, which do not recirculate in the mouse.46

**Marginal Zone B Cells**

Marginal zone (MZ) B cells are a population of noncirculating mature B cells located in the marginal zone of the rodent spleen. In rodents, MZ B cells present clear phenotypic and functional differences to the cells present in the follicles, responding to blood-borne pathogens and to repetitive...
antigenic structures such as the ones present on polysaccharide antigens.

In the human spleen, the structural definition of the marginal zone does not correspond exactly to the area surrounding B cell follicles. However, there is a population with the functional characteristics of mouse MZ B cells: low activation thresholds, highly responsive to polysaccharide antigens and with a clear surface phenotype, can be differentiated. These cells, which are sometimes named MZ-like or unswitched memory, are not restricted to the human spleen but are found circulating in the peripheral blood, as well as in the lymph nodes, tonsils, and Peyer’s patches.46,47 They are defined as IgM+, IgD+, CD27+, CD21+, and CD1c+.47

Given that these cells possess the “memory” marker CD27, it has been suggested that they have experienced antigenic exposure; however, the presence of MZ-like B cells in subjects with X-linked agamaglobulinemia (a CD40L deficiency) suggests that even if antigen exposure has occurred, T cell help has not.

Interestingly, MZ-like B cells, although already present at birth, do not seem to be fully functional at that time; infants up to age 2 are particularly susceptible to infections by capsulated bacteria. This might be due to the functional immaturity of the cells or to the lack of development of the antigen-trapping microstructure.

B1 Cells

In mice, B1 cells represent a minor population of B cells that reside predominantly in the pleural and peritoneal cavities. They are named B1 because it is assumed that is the first population of B cells to develop during intrauterine life.

Functionally, B1 cells have been characterized as self-renewing cells that possess a limited BCR repertoire and respond with low affinity to a broad array of antigens, mainly phospholipids and carbohydrate structures in the bacterial cell wall. Despite their low numbers, these cells secrete most of the natural antibodies of the organism (antibodies that appear without evidence of previous immunization) and are assumed to be the precursors of most of the plasma cells that are home to the intestinal lamina propria. Phenotypically, these cells are defined as IgMhigh and IgDlow. About 70% of them express the marker CD5.

In humans, the definition of B1 cells is still unresolved. The CD5 marker has been used extensively as a surrogate marker for B1 cells with mixed success, given that activated human B cells also upregulate the expression of CD5.48

SITES OF B CELL HOMING AND ACTIVATION

After the immature B cell stage, B cells home to secondary lymphoid organs, which contain the microenvironment and architecture necessary for the retention and activation of B cells. These organs include the spleen and lymph nodes, as well as lymphoid structures in mucosal tissue (e.g., Peyer’s patches, appendix, tonsils). The secondary lymphoid tissue is adapted to trap circulating antigen and expose the B cells to it and to provide interactions with T cells and other co-stimulatory cells. Peripheral lymphoid tissue contains
specialized antigen-presenting cells known as dendritic cells. In the Peyer's patches of the intestines, foreign antigen is taken up in specialized epithelial cells known as M cells. Even though peripheral lymphoid tissues vary in structure and cellular organization, they all possess antigen-presenting cells and B cell–containing follicles surrounded by T cell–rich zones.49 As explained in the following section, antigen, T cells, and dendritic cells are required for B cell activation and differentiation into Ig-secreting plasma cells or memory B cells.

B1 cells typically home to the peritoneal and pleural cavities and, to a lesser extent, the spleen. Naïve B cells enter the peripheral circulation by passing through the endothelial lining of the sinusoids of secondary lymphoid tissue and recirculate throughout the follicles of secondary lymphoid tissues. Because naïve B cells require antigen-specific T cell help for activation, the localization of this subset near a T cell zone facilitates chance encounters between antigen-specific B cells and cognate T cells. MZ-like cells respond to antigen without cognate T cell help and in mice are well positioned to capture blood-borne antigens owing to interactions between adhesion molecules and chemokine receptors expressed in part by specialized macrophages within the marginal zone that sequester MZ B cells within the marginal sinuses.

**Circulation and Homing**

The entry, retention, and recirculation of B cells through secondary lymphoid organs depend on both adhesion molecules and chemokine receptors.50,51 First, expression of LFA-1 and VLA-4 is required for entry into the lymphoid tissue. Then the chemokine receptors CXCR5 and CCR7 direct localization within the tissue. The CXCR5 molecule is expressed on all mature B cells and mediates B cell migration to follicles in response to the chemokine CXCL13, which is produced by follicular stromal cells. These cells, in turn, are regulated by lymphotxin-α made by B cells. In the follicle, the B cells scan for antigen, making contact with potential antigen-bearing cells such as follicular dendritic cells (FDC), subcapsular macrophages, and dendritic cells. If the B cell does not encounter a cognate antigen, it will exit the lymphoid organ through the efferent lymphatics in response to the molecule sphingosine-1-phosphate (S1P). Neutralization of S1P leads to sequestration of B cells in lymphoid organs.52

On antigenic encounter the B cells are retained in the lymphoid organ due to the upregulation of CCR7. The ligands for CCR7 (CCL19 and CCL21) mediate the organization of the T cell zone and attract antigen-activated B cells to this area, where cognate T cell–B cell interactions occur.

In contrast, it is assumed that MZ-like B cells respond to antigen without the help of cognate T cells. In mice, MZ B cells are present exclusively in the spleen and do not recirculate; in humans they are found in the spleen and tonsils, as well as in blood. MZ B cells localize to the outer layers of the follicles, making them among the first cells to encounter blood-borne antigens.

The interplay of chemokine expression and the induction of chemokine receptors play an important role in the germinal center response. The chemokine CXCL12 retains centroblasts in the dark zone during the process of somatic hypermutation and isotype class switching. CXCL13 regulates migration to the light zone, where survival and selection events are mediated by interactions with CXCR5-expressing follicular helper T cells and FDCs.53 Later, CXCL12 promotes the migration of plasmablasts to the bone marrow, where they undergo further development into long-lived plasma cells.

**Mucosa-Associated Compartments**

Within the mucosal tissue, the sites of induction of immune responses are distinct from the site where the effector cells reside. There are two main sites for the induction of an immune response. The first is the mucosa-associated lymphoid tissue (MALT) that includes Peyer's patches, nasopharynx-associated tissue, and isolated lymphoid follicles, where exogenous antigen is displayed by specialized M cells that transport antigen to the follicle. The second site of induction includes mucosa-draining lymphoid nodes such as the mesenteric and cervical lymph nodes.53

B cells reach these sites through the systemic circulation. Once they are stimulated by antigen and induced to differentiate, they home to the effector sites in the intestinal and respiratory lamina propria, where they differentiate into plasma cells and produce antibody mainly of the IgA isotype. An interesting characteristic of the plasma cells induced in the mucosal compartments is their selective homing to mucosal effector sites. Nasal activation leads to IgA-secreting cells with high levels of CCR10 and α4β1-integrins that home to the respiratory and genitourinary tracts in response to their ligands, CCL28 and VCAM-1. Migration to the intestinal lamina propria, in contrast, seems to be dependent on orally induced activation and subsequent expression on B cells of the chemokine receptor CCR9 and integrins that bind to CCL25 and MADCAM1/VCAM-1, respectively.54

**B CELL ACTIVATION AND DIFFERENTIATION**

Engagement of surface Ig by antigen triggers a series of cellular events that regulate B cell proliferation and differentiation. Receptor cross-linking leads to relocation of the BCR to microdomains known as lipid rafts, which leads to the rapid activation of proximal mediators of the BCR signal transduction pathway.55 This results in the activation of second messengers such as phospholipase C, phosphatidylinositol 3-kinase, and Ras pathways. Induction of these pathways ultimately transmits signals to the nucleus, initiating new gene expression. Depending on the type of signal delivered and the stage of maturation of the B cell, B cells can undergo either differentiation into memory B cells and plasma cells, or apoptosis. Along with surface Ig, several other membrane receptors modulate antigen-induced signal transduction.

**B CELL RECEPTOR SIGNALING**

B cell activation is triggered by the binding of specific antigen to the BCR. The end result of this process will
depend on the characteristics of the antigen, the B cell subpopulation activated, and the co-stimulatory signals provided by the antigen itself, T cells, and the microenvironment.

The BCR complex is composed of surface Ig, noncovalently bound to a dimer formed by the molecules Igα and Igβ. The surface Ig on the naïve B cell includes both IgM and IgD. The role of surface Ig is to recognize foreign antigen; the Igα and Igβ molecules transduce the signal through a particular amino acid sequence known as an immunoreceptor tyrosine-based activation motif (ITAM). This sequence contains two tyrosine residues that can be phosphorylated on activation. After phosphorylation, the ITAM acts as a docking site to recruit other signaling molecules.

In resting B cells, the BCRs are disorganized in the cell membrane, but after BCR cross-linking, they aggregate and translocate to specific regions of the cell membrane named lipid rafts. The signal transduction events that occur after BCR cross-linking are mediated by the subsequent recruitment and activation of intracellular kinases including Lyn, Fyn, Btk, and Syk. The most proximal event following BCR cross-linking is the activation of Lyn, which results in the activation of the phosphatase CD45. CD54 removes the inhibitory phosphates on the ITAMs of Igα and Igβ, and the activation of Lyn leads to the activation of Syk and Btk. There is evidence that ligation of CD19, an activating co-receptor of the BCR, leads to recruitment and activation of Vav, phosphatidylinositol 3-kinase, Fyn, Lyn, and Lck. Subsequently, the tyrosine kinases Syk and Btk are activated by tyrosine phosphorylation. The phosphorylation of Syk triggers the activation of phospholipase C, phosphatidylinositol 3-kinase, and Ras pathways. The activation of Syk appears to be absolutely critical for BCR-mediated signal transduction because Syk-deficient cell lines exhibit a loss of BCR-induced signaling. Btk also appears to be required for the activation of second messenger pathways. In some patients with X-linked agammaglobulinemia, a mutation in the Btk gene results in impaired BCR signaling at the pre-B cell stage. As a consequence, these patients have a greatly reduced number of mature B cells and generate poor antibody responses. In mice, however, a mutation in Btk leads to a disease known as X-linked immunodeficiency. B cell development is impaired at the transitional T2 stage, and B cells that do go on to maturity are unable to respond to certain T cell–independent antigens.

Following recruitment and activation of the intracellular kinases, downstream pathways are initiated. Btk, Syk, and the adapter molecule BLNK are required for the activation of phospholipase C gamma. This leads to breakdown of phosphatidylinositol 4-phosphate to DAG and IP3 to trigger calcium release from intracellular stores and the subsequent translocation of nuclear factor of activated T cells (NFAT) to the nucleus. In addition, Btk activates Ras, which leads to nuclear translocation of the transcription factor activator protein-1 (AP-1). BCR cross-linking also activates MAP kinases.

Induction of these pathways ultimately transmits signals to the nucleus, where signals are integrated to regulate gene expression. The main transcriptional activator related to B cell activation is NFκB, a family of transcriptional factors consisting of homodimers or heterodimers of different subunits (Figure 14-5). NFκB regulates cellular processes leading to activation, differentiation into memory B cells, and plasma cells or apoptosis.

**Surface Co-receptors and Intracellular Regulators**

Along with surface Ig, many molecules can modulate BCR signal transduction, either enhancing or diminishing the signal transduced by antigen. These include, but are not limited to, the B cell co-receptor complex (CD19/CD21/CD81/Leu-13), CD45, SHP-1, SHP-2, SHIP, CD22,FcγRIIB1, CD5, CD72, PIR-B, and PD-1 (see Figure 14-5). The integration of all these signals sets the threshold for activation of the BCR signal.

The main positive regulator of B cell activation is the B cell co-receptor complex, composed of CD19, CD21, CD81, and an interferon-inducible molecule called Leu-13. After antigen cross-linking of surface Ig, specific tyrosine residues contained within the CD19 cytoplasmic domain rapidly become phosphorylated. Although the natural ligand for CD19 is not known, in vitro studies have demonstrated that ligation of CD19 with anti-CD19 antibody lowers the threshold required for BCR-mediated B cell activation and enhances the proliferative effect of anti-IgM treatment on B cells. A role for CD19 in B cell activation has been clearly defined in mice that either lack or overexpress CD19. The CD19 molecule is required for T cell–dependent and T cell–independent B cell responses and for germinal center formation. The CD21 molecule serves as a receptor for cleavage fragments of the C3 component of complement. Cross-linking of the BCR and CD21 with complement-coated antigen trigger has been proposed to trigger the activation of the cytoplasmic tail of CD19. Mice deficient in CD21 also have impaired responses to T cell–dependent and T cell–independent antigens and a defect in germinal center formation. The functions of the remaining two components of the B cell co-receptor complex, CD81 and Leu-13, have not been characterized, although it has been suggested that these molecules may mediate homotypic cell adhesion.

In contrast, CD22 that also associates with the BCR is primarily a negative regulator of BCR activation. Although CD22 contains an ITAM motif and is able to recruit Src tyrosine kinases to its cytoplasmic domain, it also contains a specific motif known as the immunoreceptor tyrosine-based inhibition motif (ITIM). Similar to the ITAM, the ITIM contains a critical tyrosine residue. After activation of Lyn, it is believed that the ITIM of CD22 is phosphorylated by Lyn, leading to the recruitment of intracellular phosphatases such as SHP-1 that downregulate the activation cascade.

**FcγRIIB**

Simultaneous ligation of both the BCR and FcγRIIB1 sends an inhibitory signal to diminish antigen activation of naïve and memory B cells. This inhibitory signal is activated by the presence of immune complexes and provides a negative feedback mechanism to attenuate an antigen-induced antibody response. After co-ligation of FcγRIIB1 and the BCR,
that CD72 plays a negative role in B cell activation, presumably through recruitment of SHP-1. The B cell of CD72-deficient mice is similar to that of viable moth-eaten mice with expansion of B1 cells and B cells that are hyperresponsive to BCR cross-linking and are more resistant to BCR-mediated apoptosis.

There are several putative ligands for CD72 including CD5 and CD100. PIR Paired Ig-like receptor (PIR)-A and PIR-B are expressed in a pair-wise fashion, as the names imply. These receptors are believed to have opposing functions, with PIR-A inducing an activation signal and PIR-B inducing an inhibition signal. The ligands for PIR-A and PIR-B remain to be elucidated. Although little is known about the role of PIR-A in B cell activation, recent data demonstrate that PIR-B plays a role in downregulating B cell responses. The cytoplasmic tail of PIR-B possesses several ITIMs that recruit inhibitory phosphatases. Mice that harbor a targeted disruption of the PIR-B gene exhibit a phenotype similar to mice that are deficient in the other ITIM-bearing inhibitory receptors such as an expansion of B1 cells and B cell hyperresponsiveness.

it is believed that Lyn phosphorylates FcγRIIB1. SHIP then associates with the FcγRIIB1 and mediates the dephosphorylation of CD19, thereby terminating BCR signaling.

**CD5**

The role of CD5 in B1a cell function is not well understood. After BCR cross-linking, CD5 is thought to mediate signals that induce apoptosis and block proliferation. Cross-linking of CD5 with an anti-CD5 monoclonal antibody results in apoptosis. There is some evidence that CD5 recruits the inhibitory phosphatase SHP-1 to its cytoplasmic domain. However, unlike CD22 and FcγRIIB1, CD5 does not contain a strong ITIM consensus sequence and may recruit SHP-1 indirectly. The ligand-binding region of CD5 remains to be elucidated, but recent evidence suggests that CD5 may be a ligand for another negative regulator of BCR signaling, CD72.

**CD72**

CD72 is a transmembrane receptor that is expressed as a homodimer. The cytoplasmic tail of CD72 contains ITIMs. Mice with a targeted disruption of the CD72 gene reveal Figure 14-5  Molecules that regulate the activation state of B cells. Co-ligation of surface immunoglobulin results in tyrosine phosphorylation at specific tyrosine residues present in the immunoreceptor tyrosine activation motif (ITAM) of Igα and Igβ cytoplasmic domains. This occurs after the removal of the inhibitory tyrosine residues of the B cell receptor (BCR)–associated cytoplasmic kinases such as Blk, Fyn, and Lyn, which is mediated by CD45. The phosphorylated ITAMs recruit and activate the Syk and Btk kinases, which in turn activate a series of second messenger pathways (PLC, PKC, and Ras) that result in the upregulation of genes required for B cell activation and survival. Co-ligation of the pre-BCR complex (CD19, CD21, CD81, and Leu-13) results in phosphorylation of tyrosine residues residing in the cytoplasmic domain of CD19. Cytoplasmic kinases including Vav, Fyn, and Lyn become activated and enhance the signaling mediated by the BCR. Following the activation of distal mediators of BCR signaling such as PLC, PKC, and Ras, molecules of the MAPK pathway become activated and translocate to the nucleus to regulate gene expression. Signals mediated by CD22, PIR-B, CD72, PD-1, FcγRIIB1, and CD5 deliver negative signals that block the activation of distal molecules. Following phosphorylation of the immunoreceptor tyrosine inhibition motif (ITIM), present in the cytoplasmic tail of these molecules, the phosphatases SHP-1, SHP-2, and SHIP are recruited and activated.
**PD-1**

PD-1 is an inhibitory molecule expressed predominantly on activated B and T cells. The ligand-binding domain binds PD-1L, and the cytoplasmic tail of PD-1 contains ITIMs that recruit SHP-2 to attenuate BCR signals. The B cells of PD-1–deficient mice are hyperresponsive to BCR signaling, and these mice display an augmented response to T cell–independent type II antigens. On certain genetic backgrounds, PD-1 deficiency leads to an autoimmune phenotype.69

**Protein Tyrosine Phosphatases**

In general, intracellular signaling is regulated by a balance of phosphorylation and dephosphorylation. Protein tyrosine phosphatases (PTP) play a major role; among these, SHP-1, a tyrosine phosphatase, is a potent negative regulator of BCR signaling. SHP-1 is found in association with transmembrane proteins such as CD22, FcγRIIB1, CD5, CD72, and PIR-B. SHP-1 antagonizes BCR signaling by inactivating tyrosine kinases associated with signaling.66 The function of SHP-1 has been extensively studied in mice that bear a naturally occurring mutation in the SHP-1 gene. Mice with this genetic defect are known as moth-eaten mice because of the appearance of their fur. These mice have a decreased number of conventional B cells and an expansion of B1 cells, producing autoreactive IgM antibodies.70 This defect in B cell regulation underscores the importance of SHP-1 in limiting the extent of BCR signaling. In contrast, the intracellular tyrosine phosphatase SHP-2, although structurally similar to SHP-1, seems to play a positive regulatory role, as it augments ERK responses.71 The important role played by PTP in disease is demonstrated by the discovery that a single mutation in the PTPN22, a protein that regulates Src family kinases, augments the risk of multiple autoimmune diseases.72

Another important regulator, SHIP is an inositol phosphatase that inhibits B cell activation by hydrolyzing the 5′ phosphate form of phosphatidylinositol 3,4,5-trisphosphate, a critical component of numerous signaling pathways. Similar to SHP-1 and SHP-2, SHIP is recruited to an ITIM on FcγRIIB following BCR cross-linking. Mice deficient in SHIP display splenomegaly and elevated levels of serum antibody.73

**Signal Transduction in Immature versus Mature B Cells**

An important event in BCR signaling is the recruitment of signaling components to lipid rafts, which are lipid-rich microdomains of the membrane that serve to bring together requisite signal molecules to facilitate integrated functions in three-dimension. In the resting state, the BCR is excluded from lipid rafts, but following antigen engagement, the BCR relocates into rafts and the BCR signaling cascade ensues because of the clustering of signaling components. In addition to activating signals, however, inhibitory components also localize to lipid rafts. As discussed later, BCR ligation mediates negative selection during the immature and transitional B cell stages and B cell activation during the mature B cell stage. The reason for these two distinct outcomes is not clear because the same signal components are present; however, differences in membrane cholesterol limit recruitment of the BCR to lipid rafts in immature B cells.74 There is also evidence that the signal strength and duration and stage-specific expression patterns of signaling elements differ between immature and mature B cells.75

**B Cell Activation**

B cells develop without bias for a particular antigenic specificity, ensuring that a diverse repertoire of different Ig molecules is produced. Despite the expression of antiapoptotic molecules such as Bcl-2, naïve B cells are short-lived unless they are activated by antigen and accessory cells such as dendritic cells and T cells. Antigen-activated B cells undergo clonal expansion; B cells that do not interact with antigen are destined to undergo programmed cell death in a matter of days or weeks. The B1 and B2 cell subsets are regulated by different activation mechanisms and are involved in different immune responses (Table 14-3).

**B1 Cell Activation**

B1 cells present in the pleural and peritoneal cavities respond to T cell–independent antigens. There are two classes of T cell–independent antigens: type I, which includes lipopolysaccharide, and type II, which includes large multivalent antigens with repetitive epitopes, often

---

### Table 14-3  Markers of Human Mature B Cell Subsets

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Naive</th>
<th>Unswitched Memory</th>
<th>Switched Memory</th>
<th>B1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface IgM</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Surface IgD</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>CD5</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–/+</td>
</tr>
<tr>
<td>CD21</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CD23</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>CD11b/CD18</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Bone marrow progenitors</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Self-renewal capacity</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Response to T cell–independent antigens</td>
<td>+/−</td>
<td>+/−</td>
<td>+/−</td>
<td>+/−</td>
</tr>
<tr>
<td>Predominant isotype</td>
<td>IgM</td>
<td>IgM</td>
<td>IgG</td>
<td>IgM</td>
</tr>
<tr>
<td>Anatomic locations</td>
<td>Peritoneum, pleura, spleen</td>
<td>Peritoneum, pleura, spleen</td>
<td>Spleen, lymph nodes, Peyer’s patches, tonsils</td>
<td>Spleen, tonsils</td>
</tr>
</tbody>
</table>

---
found on the surface of bacteria. T cell–independent antigens can directly activate B cells, resulting in the secretion of antibody. Soluble factors such as IL-5 and IL-10 also appear to be involved in the maintenance and activation of B1 cells.

B1 cells do not require interaction with antigen-specific T cells. However, activated T cells and macrophages may augment B1 cell activation, enhance Ig production, and influence isotype class switching such that B1 cells may also produce IgA and IgG.

Marginal Zone B Cell Activation

If antigen reaches the lymphoid organ from the blood and is recognized by an MZ-like B cell, B cell activation is independent of T cell help. The absence of a requirement for T cell help is a function of the highly repetitive structures of MZ B cell antigens that causes BCR cross-linking. MZ B cells are situated in the marginal sinuses, where specialized macrophages trap and remove antigen from the circulation. Soluble factors such as BAFF and T cell–derived cytokines are important for MZ B cell activation. Following activation by antigen, MZ B cells differentiate rapidly into short-lived antibody-secreting plasma cells. MZ B cells secrete predominantly IgM and, to a lesser extent, IgG antibodies. These antibodies display low to intermediate affinity, and there is no induction of affinity maturation events.

Naïve B Cell Activation

Engagement of the BCR by antigen signals the B cells to engulf the antigen, process it intracellularly, and express peptide fragments bound to MHC class II molecules on its surface. After the expression of peptide fragments bound to MHC class II molecules, antigen-activated B cells present peptide to primed helper T cells. B cells and T cells interact through T cell receptor recognition of a peptide-MHC complex on the B cell and through engagement of co-stimulatory molecules B7 and CD40 on the B cell by CD28 and CD40L, respectively, on T cells (Figure 14-6). The B7 molecule is not constitutively expressed on B cells but is induced after antigen uptake by B cells. B cell activation during a T cell–dependent immune response also depends on engagement of the CD40 receptor expressed on B cells with the CD40 ligand (CD40L) expressed on T cells. The importance of signal transduction events mediated by CD40-CD40L engagement is evident from studies of patients with X-linked hyper-IgM syndrome, an immunodeficiency disease resulting from a defect in CD40L. These individuals do not mount strong immune responses to T cell–dependent antigens; they have high concentrations of circulating IgM but only trace amounts of IgG and no affinity maturation of the antibody response. Helper T cells secrete cytokines such as IL-2, IL-3, IL-4, IL-5, IL-10, IL-17, and IFN-γ and provide co-stimulatory signals that are important for B cell maturation and differentiation.

On encountering antigen, a naïve B cell with high affinity for the antigen will activate and differentiate to a plasma cell that will secrete IgM or IgG antibodies without somatic mutations. If the activated B cell, however, harbors a receptor with low to intermediate affinity for antigen, the cell will migrate first to the T cell area of the follicle and will begin the process of creating a germinal center.

Germinal Centers

As B cells progress to form germinal centers (GCs), they interact with cognate T cells that are a specialized subpopulation of helper cells known as T follicular helpers (Tfh). These Tfh cells express CXCR5 and, therefore, can migrate to the B cell areas, where, along with the activated B cells, they form discrete structures in the primary follicles, known as germinal centers. These are the sites where class switch recombination (CSR), affinity maturation through somatic hypermutation (SHM), and differentiation into memory B cells or long-lived plasma cells occur.

Germinal centers can be subdivided into separate regions where the different stages of B cell differentiation take place (Figure 14-7). The dark zone is the initial site of rapid proliferation, and B cells within the dark zone are called centroblasts. They are derived from a relatively small number of antigen-activated B cells (Table 14-4). Expression of the antiapoptotic Bcl-2 protein is low in these cells, whereas expression of the proapoptotic Fas protein is upregulated. Low levels of Bcl-2 expression render developing B cells sensitive to apoptosis, but these cells can be rescued by antigen and CD40-CD40L interactions provided by antigen-specific helper T cells. When these cells migrate to the light zone, they are termed centrocytes and encounter a dense network of follicular dendritic cells (FDCs) and Tfh. B cells that do not express moderate to high affinity BCRs are excluded from the light zone.

The process of somatic mutation is activated during the centroblast stage. During this process, a nucleotide base-pair...
change is introduced in the DNA sequence of Ig genes, at approximately $10^3$ per base pair per cell division, in the variable region of H- and L-chain genes. The mechanism for somatic hypermutation (SHM) is complex and requires specific hotspot sequences and the enzyme Activation-induced Cytidine Deaminase (AID). The process of SHM is responsible for the affinity maturation of the antibody response. B cell clones that express surface Ig with an increased affinity for antigen are selectively expanded during the affinity maturation process, whereas B cells that express somatically mutated Igs with low affinity for antigen or novel binding to self-antigens are targeted for apoptosis or inactivation. The importance of somatic mutation and affinity maturation during an immune response is underscored by the fact that patients with mutations in the AID gene are immunocompromised.

AID is critical for not only somatic hypermutation but also CSR. Although expression of IgM and IgD in B cells occurs through alternative splicing of a long transcript that contains coding regions for both the $\mu$ and $\delta$-chain, expression of any other Ig isotype requires the excision of all the heavy chain genes between the recombined VDJ region and the isotype to be expressed. Isotype switching requires pre-activation of the particular heavy chain locus to be involved in the recombination event; this is controlled by the cytokine milieu in which the cell is activated. For both processes, AID induces the deamination of cytidine, creating dU:dG pairs (instead of dC:dG) that activate the cellular DNA repair machinery and ultimately create circumstances that favor both CSR and SHM. For CSR, CD40-CD40L interactions and cytokines are also required.

As centroblasts become centrocytes, they require survival signals from FDCs to overcome their low level of Bcl-2 expression and high level for Fas expression. FDCs are stromal-derived cells that trap antigen by collecting antigen-antibody complexes bound to Fe receptor (FcR) in bodies known as iccosomes (immune complex-coated bodies) on the cell surface. Iccosomes deliver an antigen-specific signal to B cells through the BCR (Figure 14-8). Centrocytes with specificity for antigen on FDCs are saved from apoptosis by upregulation of the Bcl-2 molecule. Engagement of the complement receptors CR1 and CR2 (CD21 and CD35, respectively) on B cells by components of the C3 complement protein (iC3b, C3dg, and C3d) bound to FDCs may mediate a secondary co-stimulatory signal. If centrocytes do not receive these positive selection signals, they rapidly die through a Fas-dependent pathway. Should they receive survival signals, they continue to differentiate into memory B cells or plasma cells.

### Ectopic Lymphoid Structures

Despite being the prototypical structure for the induction of humoral immune responses, germinal centers are not found exclusively in lymphoid tissue. Ectopic lymphoid

---

**Table 14-4  Markers of Antigen-Activated B Cells in Secondary Lymphoid Tissue**

<table>
<thead>
<tr>
<th>Marker</th>
<th>Naive</th>
<th>Centroblast</th>
<th>Centrocyte</th>
<th>Memory</th>
<th>Plasma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface IgD</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Surface IgM, IgG, IgA, or IgE</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>CD10</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CD20</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>CD38</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>CD77</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Presence of somatic mutation</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Isotype class switch</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Bcl-2</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Fas</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>AID</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Blimp-1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>?</td>
<td>+</td>
</tr>
</tbody>
</table>

*Bcl-2 is expressed in centrocytes only after interaction with follicular dendritic cells.*
structures (eLS) with GC-like characteristics are present in sites of chronic inflammation such as the synovium in rheumatoid arthritis, pancreatic islands in type 1 diabetes, and salivary glands of Sjögren’s syndrome patients. 

These GC-like structures seem to develop as a consequence of chronic inflammation, which leads to the release of soluble mediators such as the chemokines CCL21 and CXCL12, which recruit lymphocytes. These cells, once activated, secrete cytokines such as lymphotoxin that act in a paracrine manner and contribute to the organization of a GC-like structure that includes a dark and light zone with local induction of AID. In contrast to the GC of the secondary lymphoid organs, these structures are not encapsulated. The B cells in these structures, therefore, are continuously exposed both to the local antigens that might be absent from the lymphoid organs and to the inflammatory microenvironment that may facilitate bypass of the regulatory points in B cell differentiation, hence contributing to a potential autoimmune bias in these sites. Although there is no evidence that these structures are the underlying cause of any disease, in some diseases they may contribute to tissue pathology and augment the pool of autoreactive plasma and memory cells. Not all ectopic lymphoid foci share characteristics with the GC. Which structures are more damaging to tissue is not known.

REPERTOIRE POST ANTIGEN ACTIVATION

The hallmarks of the germinal center process, affinity maturation and class switch recombination, find their functional expression in the two cell types to which B cells differentiate in the late stages of the germinal center: memory B cells and plasma cells.

Memory B Cells

Postterminal center memory B cells express Ig genes that have undergone isotype class switching and possess somatic mutations, and in humans they are distinguished by the presence of the marker CD27. The CD40-CD40L interaction contributes to directing GC B cells to mature into long-lived memory B cells. The exact life span of memory B cells is unknown. It has been postulated that these B cells either persist throughout the lifetime of the host or are renewed constantly through either nonspecific or antigen-specific stimulation.

Memory B cells circulate throughout the body in a quiescent state until specific antigen is re-encountered and triggers a potent secondary immune response. Memory cells respond to antigen much faster, require lower amounts of antigen, and can even be induced in its absence by soluble mediators such as IL-2 or IL-15, in part because the BCR is already localized to lipid rafts. Subsequently, just like naïve B cells, memory B cells ingest antigen and express peptide–MHC class II fragments. After antigen presentation of peptide to helper T cells, memory B cells undergo expansion and may differentiate to plasma cells.

Plasma Cells

The B cell differentiation cascade ends with the generation of a plasma cell. At the molecular level, the differentiation program is directed by the transcriptional repressor known as B lymphocyte-induced maturation transcription factor (Blimp-1). Blimp-1 induces a program in which plasma cells lose expression of several markers, as they no longer express surface Ig, MHC molecules, or CD20. They initiate increased protein synthesis and secretion and consequently exhibit a large cytoplasm with a well-developed endoplasmic reticulum, devoted to antibody production.

B cells differentiating into plasma cells exit the lymphoid follicles and migrate to extrafollicular regions of secondary lymphoid tissue or to the bone marrow, where the final stage of plasma cell maturation occurs. B cells in the GC are induced to become plasma cells by IL-5, IL-6, and IL-21. Plasma cells have a variable life span that may be days for short-lived plasma cells that originate extrafollicular responses to years for long-lived plasma cells that arise from the GC response and home to the bone marrow.

The longevity of plasma cells in the bone marrow depends on the presence of a niche that provides the survival factors cxcl12, APRIL, and TNF. Cross-linking the FcγRIII expressed by the plasma cells induces apoptosis, a phenomenon that has been credited with pruning the plasma cell repertoire.

Trafficking of Postimmune Cells

Both memory and plasma cells express CXCR4 that directs homing to the bone marrow but also allows localization in lymphoid tissues. Therefore CXCR4-expressing plasma cells can be found surrounding the GC or B cell follicles. A subset of memory cells expresses CXCR3 and therefore migrates toward inflammatory chemokines such as CXCL9 and CXCL10.

ANTIGEN-INDEPENDENT ACTIVATION

Although it is a central paradigm that B cells require cognate antigen for activation, it has also been known
achieved through signals provided by the BCR, co-receptors, inflammatory mediators, and metabolic by-products that are integrated by the cell according to its developmental stage. There is no absolute recognition of self. BCR cross-linking during development or in absence of co-stimulatory signals activates tolerance mechanisms in immature and transitional B cells. The same mechanisms that induce tolerance to self-antigen can induce tolerance to pathogens, and because some autoantigens are sequestered in sites of immune privilege, cells specific to them are not subject to tolerance mechanisms.

The tolerance mechanism activated depends on the stage of the B cell and the strength of the BCR signal delivered by antigen. The strength of the signal depends on the degree of cross-linking of the BCR, which in turns depends on the concentration of self-antigen and the affinity of the antibody for self-antigen. With little receptor cross-linking, because of low antigen concentration or low affinity for antigen, there is no BCR signaling, and the autoreactive B cell is not tolerized. Three mechanisms mediate tolerance induction: receptor editing, anergy, and deletion. When mechanisms that regulate autoreactive B cells fail, the breakdown of self-tolerance can lead to the development of autoimmune disease.

During B cell development, autoreactive B cells are generated in the bone marrow after V(D)J rearrangement and expression of surface Ig on immature B cells. Because of the vast number of different antibody molecules that can be formed through the random recombination of H- and L-chain variable region genes and the random association of H and L chains, all individuals generate autoreactive B cells. The process that prevents the maturation of naïve autoreactive B cells is known as central tolerance, and it occurs with such efficiency that despite the fact that more than 75% of immature B cells bear receptors with some degree of autoreactivity, less than 20% of naïve B cells do so.95

**Mucosal T–Independent Class Switch Recombination (CSR)**

In the mucosa, IgA specific to commensal bacteria is induced through machinery that is T cell independent and occurs outside of organized lymphoid tissue.93 Because this phenomenon is independent of T cells, other interactions must provide the required signals for CSR. One of the candidate molecules is BAFF (BlyS) that is expressed by dendritic cells in mucosal tissue. BAFF has been shown to induce CSR in B cells in the presence of IL-10 or TGF-β, both of which are present in the mucosal microenvironment.94

**REPERTOIRE SELECTION**

The ability to discriminate between foreign and self-antigens is as important as the capacity to mount a protective immune response. Any molecule derived from the intracellular or extracellular components of the host can be considered self-antigen; the censoring of the responses against these antigens occurs throughout B cell maturation and includes multiple mechanisms in the B cells themselves and in cells that cooperate in their activation and differentiation.

**Tolerance**

The discrimination between self and foreign structures is achieved through signals provided by the BCR, co-receptors,
cells marked for deletion are subject to apoptosis. This is a highly regulated event mediated primarily through the activation of a series of endogenous proteases. In B cells the Fas pathway and the Bcl-2 pathway play important roles in regulating apoptosis.

Fas (also known as CD95 or Apo-1), a member of the tumor necrosis factor receptor gene family, and Fas ligand are transmembrane proteins expressed on a variety of cell types. Because Fas ligand is a homotrimeric molecule, it can bind three Fas molecules. Clustering of Fas on the cell surface, which occurs when Fas molecules bind Fas ligand, activates apoptosis. It appears that when B cells engage CD40L expressed on helper T cells in the absence of BCR ligation, Fas signaling induces apoptosis. Mutations in Fas (lpr) or Fas ligand (gld) in mice result in a systemic lupus erythematosus (SLE)-like syndrome characterized by the production of pathogenic autoantibodies and lymphadenopathy. In humans, similar mutations lead to lymphadenopathy and antierythrocyte antibodies, but anti-DNA antibodies and glomerulonephritis are not expressed in these individuals. The Bcl-2 gene family is composed of molecules that either protect against or induce apoptosis in many cell types. Relative levels of these molecules dictate cell fate. For example, excess Bcl-2 or Bcl-XL promotes cell survival, whereas excess Bax or Bim induces cell death. Bcl-2 and Bcl-XL are upregulated at critical points during B cell development but can be easily counterbalanced on BCR cross-linking. The fact that certain mouse strains that overexpress Bcl-2 in B cells produce autoantibodies highlights the importance of apoptosis in tolerance.

**Anergy**

Anergy is a hyporesponsive state considered to be induced in immature B cells when they undergo a modest degree of BCR cross-linking. Anergic B cells downregulate surface Ig receptors and display a desensitization of the BCR, blocking activation of downstream signaling. Anergic B cells are short-lived. Goodnow and colleagues performed classic studies on B cell tolerance induction in mice engineered to express an anti–hen egg lysozyme (HEL) antibody, along with soluble HEL, to act as a self-antigen. In the anti-HEL transgenic mouse model, B cells that encounter soluble, monovalent HEL are anergized. These B cells populate secondary lymphoid tissue but do not secrete anti-HEL antibody and are not recruited into B cell follicles. This phenomenon is known as follicular exclusion.

Although anergy implies that the cells are not activated through BCR engagement, they can be activated by non-antigen-specific T cell co-stimulation, lipopolysaccharide, or IL-4. Exposure of anergic B cells in vivo to multivalent antigen in the presence of activated helper T cells may also lead to their activation. Consequently, it has been suggested that anergic B cells may serve as a potential source of autoantibody and may be activated in inflammatory conditions. Recently, it has been shown that B cells can be blocked from activation if they are chronically exposed to antigen and IL-6. If IL-6 is removed from the microenvironment, those chronically activated B cells will secrete antibody.

**B Cells as Immune Regulators**

B cells produce cytokines in response to their environment. Recently, the paracrine and autocrine role of these cytokines has become of great interest because clinical trials directed at B cells in multiple autoimmune diseases have shown improvement, which cannot be completely explained by a decrease of antibody titers. For example, it has been suggested that the therapeutic efficacy of B cell depletion reflects IL-10 secretion by transitional cells reconstituting the B cell compartment. Recently, it has been shown that in healthy subjects, some transitional B cells secrete IL-10 in response to CD40 engagement, whereas the equivalent population in SLE patients fails to do so.

**REGULATION BY SMALL MOLECULES**

Beyond the classic activators and regulators, the molecules described below play a particularly important role in the biology of B cells and are highlighted given their potential as biomarkers and therapeutic agents.

**Vitamin D**

Vitamin D is acquired from the diet or synthesized in the skin, followed by a conversion into a biologic product in the liver and kidney. The active metabolite, 1,25-dihydroxyvitamin D, has been shown to decrease activation and proliferation of B cells, as well as differentiation to plasma cells. Circulating levels of vitamin D tend to be decreased in patients with autoimmune disease; whether this contributes to the disease process is not known.

**Estrogens**

The role of estrogens in B cell–mediated autoimmune diseases has long been suggested by the female gender predilection within autoimmune diseases. This may reflect a variety of effector mechanisms. However, estrogens have been shown to modify the B cell repertoire, allowing survival of autoreactive B cells, and to alter the peripheral compartments in mice.

**Leptin**

Although its first described role was as an endocrine hormone with a primary role in the control of metabolism, leptin was later shown to exhibit immune regulatory effects. For example, a murine model of experimentally induced arthritis is attenuated in leptin receptor–deficient mice. More recently, it has been shown that leptin promotes B cell survival and proliferation, through induction of Bcl-2 and cyclin D1.

**B CELL–MEDIATED AUTOIMMUNITY**

B cell–mediated autoimmunity is the consequence of the production of self-reactive antibodies. We have detailed multiple mechanisms operating throughout B cell maturation and differentiation that are designed to avoid autoreactivity. The failure of only one tolerance checkpoint rarely leads to autoimmune disease; it may, however, increase...
the level of circulating autoantibodies, without clinical disease.

The generation of a B cell–mediated autoimmune disease must involve (1) the generation of B cells bearing autoreactive BCRs; (2) failure of mechanisms that in the normal event will abrogate their maturation to short- or long-lived plasma cells; and (3) tissue effects mediated by the autoantibody that leads to clinical disease.

**Origin of Autoreactive B Cells**

Theoretically, autoreactive cells can arise early in the repertoire from any B cell subpopulation. In mice, B1 cells that bear BCRs with low affinities but high polyreactivity produce autoantibodies, but these autoantibodies help remove cellular debris and their absence is associated with pathogenic autoreactivity. In addition, evidence indicates that MZ B cells secrete autoantibodies. These may also provide a physiologic rather than pathologic function.

**Autoreactivity in the Preimmune B Cell Repertoire**

Studies of the reactivity of human B cells have shown that in the healthy peripheral B cell compartment, about 20% of the naive B cells bear some degree of autoreactivity; however, little of those can be considered potentially pathogenic given their low affinity for autoantigen. In subjects with SLE, the frequency of autoreactive cells is as high as 50% in the naive B cell population; the frequency is greatest when disease is active and diminishes during periods of disease quiescence, demonstrating that an inflammatory milieu may alter B cell selection.

**Autoreactivity in the Postimmune B Cell Repertoire**

It has been suggested that most of the autoreactivity in patients with autoimmune disease is derived from class-switched antibodies that display extensive somatic mutation. Back mutation to the germline sequence often leads to a loss of autoreactivity. This observation suggests that the germinal center does not possess a fail-proof mechanism to effectively purge mutated autoreactive cells, although post-GC receptor editing and deletion of autoreactive B cells has been shown to occur. The understanding of the tolerance mechanisms for autoreactive B cells that achieve autoreactivity through somatic mutation is still incomplete.

**MOLECULAR TRIGGERS OF AUTOIMMUNITY**

Several prevailing theories attempt to explain the activation and expansion of B cells that should normally be silenced. Autoimmunity is thought to arise by a combination of environmental factors such as infectious agents that initiate an autoimmune response and genetic defects that alter B cell regulation. Proposed models for autoimmunity include (1) cross-reactivity of foreign antigen with self-antigen, (2) inappropriate co-stimulation, and (3) altered thresholds for BCR signaling.

Much of our understanding of the breakdown of self-tolerance and the progression of autoimmunity comes from the examination of mouse models. Autoimmune mouse models can be divided into two broad categories: induced autoimmunity and spontaneously occurring autoimmunity. Even though the progression of autoimmunity in humans is thought to be a highly complex process that involves multiple genetic and environmental factors, these animal models have provided much information about the molecular events that lead to a loss of self-tolerance.

**Molecular Mimicry**

One proposed model for the initiation of autoreactivity is that cross-reactive anti-self, anti-foreign B cells escape central tolerance because self-antigen is present at too low a concentration to trigger tolerance induction or because the affinity of the antibody for autoantigen is below the signaling threshold. These B cells become activated in the periphery by foreign pathogens resembling self-antigen and produce antibodies that bind both foreign and self-antigen. This cross-reactivity is known as molecular mimicry—a popular model to explain the induction of many autoimmune disorders. Once the pathogen is cleared, the autoantibody response should be terminated because antigen-specific T cell help is no longer present. In the case of autoimmune-prone individuals, it is proposed that intrinsic B cell defects prevent the downregulation of autoantibody production, even after foreign antigen clearance. Several data support molecular mimicry as a trigger for B cell–mediated autoimmunity in some instances: Antibodies to infectious agents have been identified that cross-react with self-molecules associated with specific autoimmune diseases (Table 14-5). Intriguing examples include the cross-reactivity observed between the M protein of group A Streptococcus and cardiac myosin in rheumatic heart disease and the cross-reactivity between Campylobacter and aquaporin.

Because both nonautoimmune and autoimmune-prone individuals have the capacity to generate autoantibodies, it is unlikely that cross-reactivity between foreign and self-molecules is a major mechanism of disease.
self-antigens is solely responsible for breakdown of tolerance. A plausible explanation is that foreign antigen acts as a molecular trigger to initiate an immune response to self-molecules, and a defect in the mechanism that regulates B cell activation leads to the propagation of an autoimmune response.

In general, an initial immune response is generated against a dominant set of epitopes, followed by a later response to secondary or "cryptic" epitopes, a process known as epitope spreading. Epitope spreading is an important aspect of a protective immune response because the ability to recognize multiple antigenic determinants increases the efficiency of the neutralization and removal of pathogens. When an autoimmune response has been triggered, epitope spreading can lead to the production of additional autoantibodies with specificity for multiple self-antigens. There are several proposed mechanisms by which epitope spreading triggers a cascade of T and B cell activation. For instance, antigen-presenting cells may present a foreign peptide that mimics a self-peptide to T cells (Figure 14-9A). Such cross-reactive T cells become activated and provide co-stimulation to autoreactive B cells that recognize self-antigen. This results in the production of autoantibodies specific for the antigen recognized by the T cell. After internalization of the self-antigen by the autoreactive B cells, the autoantigen is processed and new cryptic epitopes of the self-antigen are presented to T cells. A B cell binding to the self-antigen internalizes not only that self-antigen but also any complex of molecules that includes the self-antigen. The B cell may, therefore, present cryptic epitopes of many self-antigens and activate autoreactive T cells representing multiple autospecificities. In the periphery, T cells are present that have not been tolerated necessarily to these (cryptic) epitopes and thus are activated by self-peptide. These activated T cells in turn help provide co-stimulation and activate other autoreactive B cells.

Alternatively, cross-reactive B cells may be activated first after exposure to foreign antigen and T cell help (Figure 14-9B). These B cells internalize self-antigen and present cryptic peptides to T cells that have not been tolerized, leading to activation of autoreactive T cells and initiation of the cascade. Thus molecular mimicry and epitope spreading could lead to the activation of T cells and B cells specific for multiple autoantigens as long as the autoantigens form a complex in vivo.

**Supraoptimal B Cell Co-stimulation**

It is evident that co-stimulatory signals provided by T cells play a critical role in B cell activation. Therefore inappropriate co-stimulation may lead to the propagation of an immune response directed against a self-antigen. The interaction between B7 on B cells and CD28 on T cells is crucial for the activation of antigen-specific T cells and B cells. When a genetically engineered protein that inhibits B7-CD28 interactions is administered to autoimmune mice, progression of disease is blocked. Reciprocally, autoreactive B cells present in mice that constitutively overexpress B7 are not sensitive to Fas killing and mice display high serum autoantibody titers. Overexpression of CD40 or CD40L may also activate autoreactivity. In vitro studies have demonstrated that CD40-CD40L ligation in the presence of IL-4 activates anergic cells. It has been suggested that CD40L may be overexpressed in lymphoid cells of patients with SLE.

Roquin, a member of the ubiquitin ligase family, regulates the function of follicular helper T cells. Roquin belongs to a family of RING-type ubiquitin ligases involved in the post-translational regulation of gene expression and represses the expression of ICOS, a co-stimulatory molecule that plays an important role in follicular helper T cell function. These cells provide strong co-stimulation in the germinal center. Mice harboring a mutation in the roquin gene display high-affinity dsDNA antibodies owing to increased numbers of germinal center B cells and follicular helper T cells.

Interferon regulatory factor-4 binding protein (IBP) has also been shown to regulate T cell co-stimulatory signals. IBP is a regulator of Rho GTPases and is recruited to the immunologic synapse following T cell receptor cross-linking to mediate the reorganization of the cytoskeleton. Mice deficient in IBP exhibit an autoimmune phenotype characterized by the production of dsDNA antibodies and glomerulonephritis. IBP plays an important role in the survival and effector function of memory T cells and underscores a novel role for Rho GTPases in regulating interactions between T cells and autoreactive B cells.

Toll-like receptors (TLRs) belong to a family of pattern recognition receptors that initiate innate immune responses to various components of pathogens. TLR7, which recognizes RNA, and TLR9, which recognizes unmethylated CpG-containing nucleic acid sequences, are expressed on B cells and have been implicated in autoimmunity. Numerous studies suggest that engagement of the BCR and TLR with immune complexes containing nuclear antigens triggers the activation of antinuclear B cells, implying that TLR7 and TLR9 can function to enhance the activation of autoreactive B cells under some circumstances.

**B Cell Signaling Thresholds**

The effects of altering the threshold for BCR signaling have been demonstrated in several mouse models. In transgenic mice that overexpress the BCR co-receptor complex component CD19, normally anergic B cells are activated and secrete autoantibody. These results suggest that a decrease in the minimal requirement for antigen engagement of the BCR can lead to inappropriate activation of autoreactive B cells. Viable moth-eaten mice also develop an autoimmune syndrome due to a naturally occurring deficiency in the SHP-1 phosphatase, a potent negative regulator of BCR signaling. In these mice, B1 cells are responsible for the production of IgM anti-DNA antibodies. Transgenic mice deficient in other signaling molecules that alter threshold activation such as CD22 and Lyn also produce autoantibodies. Thus changes in thresholds for antigen-induced B cell activation can lead to the activation of autoreactive B cells.

**SUMMARY**

The generation of a diverse repertoire of antibody molecules provides an important line of defense against microbial infections. The immune system is exquisitely controlled at
Figure 14-9  Epitope spreading. **A**, Epitope spreading by activation of cross-reactive T cells. I, Following antigen presentation of a foreign peptide that is recognized by cross-reactive T cells, co-stimulatory signals are delivered to B cells with surface immunoglobulin receptors that recognize a self-antigen as part of a complex of self-molecules. II, The complex is engulfed by a self-reactive B cell, and antibodies specific for self-antigen are generated. III, Self-reactive B cells process the self-molecules and present cryptic peptide–MHC II complexes on the cell surface. IV, If these cryptic peptides are recognized by nontolerized autoreactive T cells, B cells specific for these cryptic peptides are activated and the autoantibody response spreads to other components of the self-antigen complex. **B**, Epitope spreading by activation of autoreactive B cells. I, A foreign antigen that mimics a self-molecule can mediate the endocytosis of a self-molecule that is included in a self-antigen complex. The self-molecules of the complex are processed and expressed on the cell surface of the B cell as cryptic peptide–MHC II complexes. II, Cryptic peptides are recognized by nontolerized autoreactive T cells. III, These T cells provide co-stimulation to B cells that recognize cryptic peptides, resulting in the production of additional self-reactive antibodies.
multiple levels to allow the maturation of B cells that produce protective antibodies while attempting to avoid the production of autoantibodies (Figure 14-10). Only a small percentage of B cell precursors generated completes the maturation pathway. During the pro-B and pre-B cell stages of development, B cells with aberrantly rearranged H- or L-chain genes are eliminated. As the remaining precursor cells transit into the immature B cell stage, they are subject to negative selection; that is, immature B cells with autoreactivity are either deleted or inactivated, whereas nonautoreactive B cells are released into the periphery. B cells that are stimulated by foreign antigen are selectively expanded and undergo further Ig gene diversification in peripheral lymphoid tissue. During this stage of development, B cells that express high-affinity Ig receptors undergo positive selection, whereas B cells with a diminished affinity or those that have acquired autoreactivity are eliminated. B cells that pass through these critical developmental checkpoints differentiate into long-lived memory B cells or plasma cells. The underlying causes of B cell–associated autoimmunity are not understood, but just as there are multiple checkpoints for the survival or activation of autoreactive B cells, it seems likely that multiple defects in the regulatory mechanisms that control B cell maturation and differentiation contribute to autoimmune disease.

References

8. Froelich CJ, Hanna WL, Poitier GC, et al: Granzyme B/perforin-mediated apoptosis of Jurkat cells results in cleavage of poly (ADP-ribose) polymerase to the 89-kDa apoptotic fragment and

Figure 14-10 Selection checkpoints during B cell maturation. Autoreactive B cells can be censored at multiple developmental checkpoints: (1) Following surface expression of surface immunoglobulin, immature B cells that encounter autoantigen in the bone marrow are subject to negative selection. (2) B cells that are not eliminated in the bone marrow may undergo negative selection during the transitional B cell stage. Transitional B cells that emerge from this development stage give rise to follicular (FO) or marginal zone (MZ) B cells. Follicular B cells activated by antigen and the help of cognate T cells progress to the germinal center (GC) B cell stage. (3) Germinal center B cells that acquire high affinity for autoantigen by the process of somatic hypermutation may be eliminated in the germinal center to block their further maturation into long-lived plasma cells or memory cells. (4) There is evidence that autoreactive plasmablasts may also be subject to negative selection. Long-lived plasma cells that emerge from the selection process home primarily to the bone marrow, and memory B cells circulate throughout the periphery. HC, heavy chain; LC, light chain.


The references for this chapter can also be found on www.expertconsult.com.
Fibroblasts and Fibroblast-like Synoviocytes

ANDREW FILER • CHRISTOPHER D. BUCKLEY

KEY POINTS

Fibroblasts are programmed epigenetically to determine the unique structure and function of different organs and tissues. However, these unique features might contribute to organ-specific disease.

Tissue fibroblasts may be recruited from a number of sources and cell types including the bone marrow, blood, and local stromal cells and act as organ-specific innate immune sentinel cells.

Under inflammatory conditions, fibroblasts become key immune system players recruiting and modulating the behavior and survival of infiltrating immune cells.

Fibroblasts can be programmed epigenetically through exposure to inflammatory and environmental hits such that they inappropriately prolong inflammation, which becomes persistent.

Within the synovium, this persistent abnormal behavior results in continued damage to vital joint structures such as cartilage and bone, which, if untreated, will result in deformity and functional impairment.

In vivo models allow us to observe and modulate fibroblast behavior; these models have recently revealed the possibility of diseased fibroblasts autonomously moving to different sites in the body.

WHAT IS A FIBROBLAST?

The architecture of organs and tissues is closely adapted to their function in order to provide microenvironments in which specialized functions may be carried out efficiently. The nature and character of such microenvironments are primarily defined by the stromal cells that reside within the tissues. The most abundant cell types of the stroma are fibroblasts, which are responsible for the synthesis and remodeling of extracellular matrix (ECM) components. In addition, their ability to produce and respond to growth factors and cytokines allows reciprocal interactions with adjacent epithelial and endothelial structures and with infiltrating leukocytes. Fibroblasts also act as integrators of microenvironmental stimuli such as oxygen tension and pH. As a consequence, fibroblasts play a critical role during tissue development and homeostasis and are often described as having a “landscaping” function.

Fibroblast Identity and Microenvironments

Tissue-resident macrophages in the liver (Kupffer cells) and lung (alveolar macrophages) perform very different functions compared with macrophages in the brain (glial cells) or skin (Langerhans cells), yet they are all members of the monocyte/macrophage family. Until recently fibroblasts had been thought of as ubiquitous, generic cells with a common phenotype even within different tissues. However, we now know that fibroblasts from different organs are more like their macrophage counterparts, with unique morphology and repertoires of ECM proteins, cytokines, co-stimulatory molecules, and chemokines specialized to the different microenvironments in which they are found. This also extends to their function as “immune sentinel” cells, expressing innate immune system pattern recognition receptors such as Toll-like receptors (TLRs), which trigger a proinflammatory response when ligated by bacterial or viral determinants. When fibroblast transcriptional profiles are examined using microarray techniques, fibroblasts hold a strong memory of their anatomic position and function in the body. Early studies demonstrated that fibroblast transcriptomes (the global picture of transcribed genes measured using microarrays) could be clustered into peripheral (synovial joint or skin fibroblasts) versus lymphoid (tonsil or lymph node) groups according to their organ of origin, with the potential to shift their transcriptional profiles by treatment with inflammatory mediators such as tumor necrosis factor (TNF), IL-4, or interferon-γ. More extensive analysis of expression profiles from primary human fibroblasts by Rinn and colleagues has shown large-scale differences related to three broad anatomic divisions: anterior-posterior, proximal-distal, and dermal-nondermal. Genes involved in pattern forming, cell-signaling, and matrix remodeling were found to predominantly account for these divisions. The gene expression profile of adult fibroblasts may therefore play a significant role in assigning positional identity within an organism. More recently, it has become clear that these stable changes in gene transcription are brought about through epigenetic activation and silencing of the HOX family of landscaping genes. Such epigenetic patterning, whereby covalent modifications are made to regulatory regions of DNA, or to the histones around which the DNA is wrapped in order to control access of transcriptional complexes, is a prototype for the stable changes that are also seen in fibroblasts. Epigenetic modifications result in stable changes in gene expression that persist over cellular generations in the absence of mutation of the primary DNA sequence, and which therefore drive the persistence of disease, as is described in more detail in Chapter 22.

Embryologic Origins

The problem of distinguishing fibroblasts of differing origin or maturity has historically been difficult due to a lack of specific cell surface markers. Whereas cluster of
differentiation (CD) markers have revolutionized the isolation and study of leukocyte subsets, there have been relatively few, poor-quality discriminatory markers allowing the identification of fibroblast subpopulations. Fibroblasts have therefore traditionally been identified by their spindle-shaped morphology (Figure 15-1), elaboration of ECM, and lack of positive markers for endothelium, epithelial, and hemopoietic cells.

However, there is growing evidence that fibroblasts are not a homogeneous population but exist as subsets of cells, much like tissue macrophages and dendritic cells. It is likely that connective tissue contains a mixture of distinct fibroblast lineages with mature fibroblasts existing side by side with more immature fibroblasts that are capable of differentiating into other connective tissue cells. Recent studies have begun to identify novel markers that demarcate distinct subpopulations of stromal cells during development and which have the potential to act as markers for different subpopulations of fibroblasts, each with different roles. Such markers include smooth muscle actin, which marks out a population of secretory, activated cells termed myofibroblasts, and more recently discovered markers such as CD248 and gp38 (podoplanin) (Table 15-13-14 and see text later). Fibroblasts have been defined in terms of their embryologic origins and lineage relationships and are generally considered to be mesenchymal in origin. However, cell populations that appear to blur the distinction between hemopoietic and nonhemopoietic populations have now been identified. In addition, other unexpected shifts in lineage have been reported including differentiation from neural stem cells into myeloid and lymphoid hemopoietic lineage. Classification by such lineages is therefore becoming increasingly untenable.

**Origins of Fibroblasts in Tissue**

Both inflammation and wound healing are characterized by the formation of new tissue. However, recent findings suggest that the new cells that form the remodeled tissues are not necessarily, as was hitherto assumed, derived from the proliferation of cells that are resident in the adjacent noninjured tissue. This is an important issue because in both RA and fibrotic pathologic conditions, fibroblasts accumulate in excessive numbers despite apparently low proliferative rates. The principle origin for fibroblasts is from primary mesenchymal cells and that on appropriate stimulation fibroblasts can proliferate locally to generate new fibroblasts. In fact, though an increase in fibroblast numbers caused by local proliferation does occur, fibroblasts may arise from other sources (Figure 15-2). The first of these is local epi-

---

**Figure 15-1**  Fibroblast phenotype. **A**, Staining and differential interference contrast microscopy of live fibroblast cells in culture illustrating typical morphology and marked differences between synovial fibroblasts of the rheumatoid arthritis joint and skin fibroblasts. Red stain (fibronectin) demonstrates matrix production. Blue stain indicates nuclei. **B**, Stromal cell status is confirmed by fluorescence microscopy of cells showing collagen synthetic enzymes (prolyl-4-hydroxylase) within and matrix production (fibronectin) on the surface of skin fibroblasts.
Table 15-1  Synovial Stromal Markers and Their Geographic and Functional Significance

<table>
<thead>
<tr>
<th>Marker</th>
<th>Associated Cell Type</th>
<th>Synovial Location</th>
<th>Functional Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD55</td>
<td>Fibroblast-like synoviocyte</td>
<td>Lining layer</td>
<td>Receptor/ligand for synovial macrophage CD97</td>
</tr>
<tr>
<td>VCAM-1</td>
<td>Fibroblast-like synoviocyte</td>
<td>Lining layer</td>
<td>Activated lining layer fibroblasts; adhesion molecule</td>
</tr>
<tr>
<td>Alpha–smooth muscle actin (α-SMA)  CD248/endoosialin</td>
<td>Myofibroblast</td>
<td>Variable, minority subpopulation</td>
<td>Secretory, profibrotic fibroblast</td>
</tr>
<tr>
<td>gp38/podoplanin</td>
<td>Pericyte</td>
<td>Lining and sublining cells</td>
<td>Acute inflammation, cancer and vasculogenesis</td>
</tr>
<tr>
<td>SBS/prolyl-4-hydroxylase S100A4/FSP-1/Mts-1</td>
<td>Broad fibroblast marker in vivo</td>
<td></td>
<td>Structural, proangiogenic lymph node role; promotes motility in cancer</td>
</tr>
<tr>
<td>Fibroblast activation protein (FAP)</td>
<td>Associated with αSMA-positive fibroblasts</td>
<td>Lining layer</td>
<td>Marks collagen synthetic machinery</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cancer, invasiveness roles via motility and impaired apoptosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Role in cancer fibroblasts, protective if ectoenzyme blocked in rheumatoid arthritis</td>
</tr>
</tbody>
</table>

Figure 15-2  Routes of differentiation to tissue fibroblasts. In response to wounding or inflammation, increased numbers of fibroblasts are produced within tissue. 1. Fibroblasts can proliferate locally to generate new fibroblasts. 2. The transition of epithelial to stromal cell populations has been shown to occur in cancer and diseases of the lung, kidney, and possibly the synovium. 3. Fibrocytes arise from the monocyte population in blood and then differentiate toward fibroblasts in tissue. 4. Blood-borne mesenchymal progenitor cells (MPC) may be recruited to tissues and undergo local differentiation to tissue fibroblasts. EMT, epithelial to mesenchymal transition.

Thelial to mesenchymal transition (EMT). This is an essential, physiologically important developmental mechanism for diversifying cells in the formation of complex tissues. However, fibroblasts also appear to be derived by this process in adult tissue following epithelial stress such as inflammation or tissue injury. EMT both disaggregates epithelial cells and reshapes them for movement. The epithelium loses polarity as defined by the loss of adherens junctions, tight junctions, desmosomes, and cytokeratin intermediate filaments. Epithelial cells also rearrange their F-actin stress fibers and express filopodia and lamellipodia. A combination of cytokines and matrix metalloproteinases (MMPs) associated with digestion of the basement membrane is believed to be secreted and important in the process. The transition of epithelial to mesenchymal cell populations has been shown to occur in cancer and in diseases of the lung and kidney in which the process has been implicated in fibrotic disease. Early evidence suggests that a similar process may occur within the RA synovium.
An alternative explanation for the accumulation of stromal cells in chronic inflammatory conditions such as rheumatoid arthritis (RA) lies in the possibility of blood-borne precursors. In the mid 1990s it was shown that vascular precursors (angioblasts) could be found circulating in the blood of normal individuals and that they could be recruited to sites of vasculogenesis in a rabbit ischemic hind limb model. This demonstrated that circulating mesenchymal precursors exist outside the hemopoietic system. Subsequent work has confirmed the presence of circulating cells of a mesenchymal phenotype in human subjects. These cells bear a remarkable resemblance to the synovial fibroblasts found in the joints of patients with RA, which accumulate in large quantities in the joint lining despite little evidence of proliferation. Interestingly, Marinova-Mutafchieva and colleagues showed that an influx of such cells preceeded inflammation in a mouse collagen-induced arthritis model, suggesting that there may be a role for blood-borne stromal cell precursors in the initiation of inflammatory diseases. Furthermore, evidence now exists to show that synovial fibroblasts themselves may migrate in the bloodstream, at least between distant sections of human cartilage in SCID mice, raising an intriguing parallel to cancer and the radical concept of RA as a metastatic disease of the stroma.

Another circulating precursor cell that could account for the accumulation of fibroblasts in disease is the fibrocyte. Fibrocytes appear to comprise 0.1% to 0.5% of nonerythrocytic cells in peripheral blood and have been shown to rapidly enter sites of tissue injury and contribute to tissue remodeling in models of inflammatory lung disease. They are adherent cells with a spindle-shaped morphology that express MHC class II and type I collagen and which arise from within the CD14+ (monocyte) fraction of peripheral blood. Fibrocytes are capable of matrix elaboration and differentiate along fibroblast lineages under the influence of cytokines, particularly transforming growth factor-β (TGF-β). The mere fact that a cell type apparently arising from within the monocyte lineage may become a “mesenchymal” stromal cell such as a fibroblast implies a further degree of plasticity and blurring of the apparently clear dividing line previously thought to exist between hemopoietic and non-hemopoietic lineages.

**Fibroblasts versus Mesenchymal Progenitor Cells**

The potential role of circulating mesenchymal cell precursors (variously termed mesenchymal stem cells (MSCs), mesenchymal stromal cells, or mesenchymal progenitor cells (MPCs)) as sources of tissue fibroblasts is highlighted by the remarkable capacity of these cells to differentiate into other members of the connective tissue family including cartilage, bone, adipocyte, and smooth muscle cells. This ability was initially demonstrated in bone marrow stromal cells, RA synovial fibroblasts, and circulating mesenchymal cells. Therefore it was suggested to define a characteristic mesenchymal phenotype on the basis of the hypothesis that the rheumatoid synovium could become populated by a large proportion of circulating mesenchymal progenitor cells exported from the bone marrow. However, the property of trilineage differentiation (“pluripotentiality”) has now been shown to be a property of many adult tissue fibroblasts, though varying somewhat between fibroblasts from different tissues, implying a hitherto unsuspected degree of plasticity in the body’s stromal cell populations. The two previously separate fields of mesenchymal precursor cell biology and largely disease-centered fibroblast biology have therefore rapidly converged. However, the concept of bone marrow stromal precursors remains interesting; in chimeric murine models with bone marrow GFP expression, arthritic joints contained significantly more GFP-positive+ cells than non-arthritic joints, supporting a bone marrow origin for expanded fibroblast populations.

**Physiologic Characteristics and Functions of Fibroblasts**

**Production of ECM Components**

Ensuring the homeostasis of the ECM is one of the primary functions of fibroblasts. In order to do this, fibroblasts must be capable of both producing and degrading ECM, as well as adhering to and interacting with existing matrix components. Fibroblasts produce a number of ECM molecules, both fibrous proteins and polysaccharide gel components such as collagens, fibronectins, vitronectin, and proteoglycans, which are then assembled into a three-dimensional network. This provides a framework through which other cell types, which use varying strategies to navigate through the ECM, can move and also provides a substrate for the deposition of haptotactic (tissue rather than fluid-based) gradients of chemokines and stores of growth factors to direct cellular movement and behavior in a regional fashion. The types of ECM molecules produced by individual populations of fibroblasts differ from tissue to tissue, reflecting the diversity of fibroblasts in different organs. For example, dermal fibroblasts produce significant amounts of type VII collagen, which adheres the epidermal and dermal layers in the skin. Fibroblasts in other organs such as the lung and kidney produce mainly interstitial, fibrillar collagens (particularly types I and III).

In the synovial membrane, fibroblasts also have a barrier function, in that they provide the joint cavity and the adjacent cartilage with lubricating molecules such as hyaluronic acid and with plasma-derived nutrients. Anatomically, the intimal synovial membrane is an unusual structure in that barrier function is maintained in the absence of a laminin-rich basement membrane, as is seen in epithelial structures. In addition to lacking a basement membrane, cellular contacts between the fibroblast-like synoviocytes also lack tight junctions and desmosomes. However, a strong homophilic adhesion between synoviocytes is mediated by the adhesion molecule cadherin-11 (see later), which is largely responsible for fibroblast organization into synovial tissue. In disease, fibroblasts have to migrate to sites of tissue injury or remodeling and interact with ECM molecules through specific surface receptors. Through such receptors, fibroblasts must sense changes in both the structure and the cellular composition of connective tissues. They respond dynamically by adjusting the production of ECM components and cross-linking them into the appropriate matrix.
Attachment to and Interaction with Extracellular Matrix

**Integrins**

Integrins are key mediators of both cell-to-matrix and cell-to-cell adhesive interactions. They are expressed as transmembrane heterodimers containing one α- and one β-subunit, of which at least 25 αβ combinations are known (Table 15-2). αβ1 and αβ2 integrins are the main adhesion molecules responsible for the attachment of fibroblasts to collagen, while other β1 integrins such as α2β1 and α3β1 integrins mediate attachment of fibroblasts to fibronectin and its spliced variants. In addition, α1 integrins are responsible for attachment to vitronectin.

**Syndecans**

In addition to conventional integrin-to-ligand binding, additional accessory molecules allow for the integration of adhesive contacts and local growth factor signaling. Syndecans are a family of four single transmembrane domain proteins that carry three to five heparan sulfate and chondroitin sulfate chains, allowing for interaction with a large variety of ligands including fibroblast growth factors, VEGF, TGF-β, and ECM molecules such as fibronectin. Syndecans are expressed on fibroblasts in a tissue-specific and development-dependent manner. Data from syndecan knockout mice indicate that syndecan-4 is involved in wound healing, and the response of syndecan-4-deficient fibroblasts to fibronectin attachment is significantly altered.

**Immunoglobulin Superfamily Receptors**

The immunoglobulin superfamily is a diverse group of transmembrane glycoproteins defined by the presence of one or more immunoglobulin-like repeats of 60 to 100 amino acids with a single disulfide bond. While including numerous adaptive immune system genes (immunoglobulins, T cell receptor, major histocompatibility complex), adhesion proteins such as ICAMs 1 to 3, VCAM-1, and MadCAM mediate both cell-to-cell interactions and adhesive interactions with integrins (see Table 15-2).

**Cadherins**

Cadherins mediate homotypic, calcium-dependent adhesive interactions with the same cadherin species expressed by neighboring cells. Classical cadherins possess five extracellular domains, a single-pass transmembrane domain, and a highly conserved cytoplasmic tail. The cytoplasmic tail interacts with β-catenin, which in turn binds α-catenin, forming a linkage between the cadherin-catenin complex and the actin cytoskeleton. Tightly regulated expression of cadherins is essential to embryogenesis but is also critical for tissue morphogenesis and tissue-specific cell differentiation. Cadherins also modulate cell proliferation and invasion through activation of intracellular signal transduction pathways, modulation of matrix metalloproteinase production, and association with growth factor receptors.

Importantly, interaction with adhesion molecules not only regulates adhesion and motility but also directly influences activation status, apoptosis, and proinflammatory and anti-inflammatory responses in fibroblasts and other cells. The engagement of cell adhesion molecules such as integrin receptors on the surface of fibroblasts results in the formation of focal adhesion complexes, which activate intracellular signaling cascades regulating cell proliferation and survival, the secretion of certain cytokines and chemokines, and matrix deposition and resorption. In particular, integrin-to-fibronectin engagement induces matrix metalloproteinase expression, linking adhesion-to-matrix remodeling.

### Table 15-2  Cell Adhesion Molecules (CAMs) and Their Receptor/Ligand Molecules

<table>
<thead>
<tr>
<th>Family</th>
<th>CAM</th>
<th>Alternative Names</th>
<th>Ligands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrins</td>
<td>αβ1</td>
<td>VLA-1</td>
<td>Laminin, collagen</td>
</tr>
<tr>
<td></td>
<td>αβ2</td>
<td>VLA-2</td>
<td>Laminin, collagen</td>
</tr>
<tr>
<td></td>
<td>αβ3</td>
<td>VLA-3</td>
<td>Laminin, collagen</td>
</tr>
<tr>
<td></td>
<td>αβ4</td>
<td>VLA-4, CD49d/CD29</td>
<td>VCAM-1, CS1 fibronectin</td>
</tr>
<tr>
<td></td>
<td>αβ5</td>
<td>VLA-5</td>
<td>Fibronectin</td>
</tr>
<tr>
<td></td>
<td>αβ6</td>
<td>VLA-6</td>
<td>Laminin</td>
</tr>
<tr>
<td></td>
<td>αβ7</td>
<td>LFA-1, CD11a/CD18</td>
<td>ICAM-1, ICAM-2, ICAM-3, JAM-A</td>
</tr>
<tr>
<td></td>
<td>αβ8</td>
<td>Mac-1, CR3, CD11b/CD18</td>
<td>ICAM-2, IC3b, fibrinogen, factor X</td>
</tr>
<tr>
<td></td>
<td>αβ9</td>
<td>P150, 95, CD11c/CD18</td>
<td>IC3b, fibrinogen</td>
</tr>
<tr>
<td></td>
<td>αβ10</td>
<td>CD49d</td>
<td>E-cadherin</td>
</tr>
<tr>
<td></td>
<td>αβ11</td>
<td>CD52/CD61, vitronectin receptor</td>
<td>Fibronecin, VCAM-1, MadCAM-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vitronecin, fibronectin, osteopontin, thrombospondin-1, tenascin</td>
</tr>
<tr>
<td>Ig Superfamily</td>
<td>ICAM-1</td>
<td>CD54</td>
<td>LFA-1, Mac-1</td>
</tr>
<tr>
<td></td>
<td>ICAM-2</td>
<td></td>
<td>LFA-1</td>
</tr>
<tr>
<td></td>
<td>ICAM-3</td>
<td></td>
<td>LFA-1</td>
</tr>
<tr>
<td></td>
<td>VCAM-1</td>
<td></td>
<td>αβ1, αβ2, αβ3, αβ4, αβ5, αβ6, αβ7, αβ11, L-selectin</td>
</tr>
<tr>
<td></td>
<td>MadCAM-1</td>
<td></td>
<td>E-cadherin</td>
</tr>
<tr>
<td>Cadherins</td>
<td>E-cadherin</td>
<td>Cadherin-1</td>
<td>N-cadherin</td>
</tr>
<tr>
<td></td>
<td>N-cadherin</td>
<td>Cadherin-2</td>
<td>Cadherin-11</td>
</tr>
<tr>
<td></td>
<td>Cadherin-11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ICAM, intercellular adhesion molecule; JAM, junctional adhesion molecule; LFA, lymphocyte function-associated antigen; MadCAM, mucosal addressin cell adhesion molecule; VCAM, vascular cell adhesion molecule; VLA, very late antigen.
Degradation of Extracellular Matrix by Fibroblasts

Remodeling of the ECM requires fibroblasts to express an extensive repertoire of matrix-degrading enzymes with varying specificity. Although these are crucial to tissue maintenance and repair, inappropriate overexpression of such enzymes is a key factor in the joint damage, particularly to cartilage, which occurs in inflammatory disease. Such enzymes fall into a number of families including matrix (Figure 15-3).

(Figure 15-3). Among the signaling molecules that transmit signals from the integrins to the cell interior, focal adhesion kinase (FAK) plays a central role. FAK, a tyrosine kinase, is recruited into newly established focal contacts and, in turn, recruits other adapter proteins such as p130Cas and Grb2. This leads to the activation of phosphatidylinositol 3-kinase (PI3K) and Src-kinase and promotes the initiation of a variety of signaling cascades, culminating in the ERK MAPKinesases and activation of transcription factors. Such pathways can also be activated through FAK-independent signaling events, such as through growth factor receptor ligation. The exact mechanisms by which different signals cooperate to mediate a specific response of fibroblasts and how this translates into distinct pathologies are not yet fully defined.

**Degradation of Extracellular Matrix by Fibroblasts**

Remodeling of the ECM requires fibroblasts to express an extensive repertoire of matrix-degrading enzymes with varying specificity. Although these are crucial to tissue maintenance and repair, inappropriate overexpression of such enzymes is a key factor in the joint damage, particularly to cartilage, which occurs in inflammatory disease. Such enzymes fall into a number of families including matrix...
metalloproteinases (MMPs), tissue inhibitors of metalloproteinases (TIMPs), cathepsins, and aggrecanases. These are covered in detail in Chapter 8.

With the exception of MMP-2 and the MT-MMPs, which are constitutively expressed by fibroblasts, MMP expression is regulated by extracellular signals via transcriptional activation in fibroblasts. Three major groups of inducers can be differentiated: proinflammatory cytokines, growth factors, and matrix molecules. Among the cytokines, IL-1 is perhaps the most potent inducer of a variety of MMPs including MMP-1, MMP-3, MMP-8, MMP-13, and MMP-14. Fibroblast growth factor (FGF) and platelet-derived growth factor (PDGF) are also known inducers of MMPs in fibroblasts because they potentiate the effect of IL-1 on MMP expression. All MMP promoter regions except MMP-2 contain activator protein-1 (AP-1) binding sites; however, there is good evidence that all MAPKinase families (ERK, JNK, and p38 pathways) (see Figure 15-3), in addition to activators of NFκB, STAT, and ETS transcription factors participate in MMP regulation.\(^{15-19}\) Matrix proteins (collagen, fibronectin) and especially their degradation products also activate MMP expression in fibroblasts, providing the possibility for site-specific MMP activation in regions of matrix breakdown.\(^{40}\)

**Fibroblasts as Innate Immune Sentinels**

Classically, macrophages have been studied as sources of inflammatory cytokines and chemokines in response to innate immune stimuli and portrayed as immune sentinel cells accordingly. However, when activated by substances released during tissue injury or the products of invading microorganisms, fibroblasts are capable of elaborating a broad repertoire of inflammatory mediators, which fully justifies their classification as immune sentinel cells. Through expression of TLRs 2, 3, and 4, fibroblasts respond to bacterial products such as LPS by activating the classical NFκB and AP-1 inflammatory pathways, generating chemokines capable of recruiting inflammatory cells, and generating metalloproteinases capable of degrading matrix.\(^{41-43}\) However, TLR expression may be increased by proinflammatory cytokines TNF and IL-1β within the local microenvironment\(^{44}\) and may also be activated by endogenous cellular debris such as necrotic cells in synovial fluid, leading to widespread fibroblast activation in disease.\(^{45}\) As immune sentinels, fibroblasts are capable of bridging the innate and adaptive immune responses through expression of the molecule CD40. This molecule was initially assumed to be restricted in its expression to antigen-presenting cells such as macrophages and dendritic cells. However, it is widely expressed by fibroblasts within discrete tissues. Engagement of CD40 by its ligand CD40L expressed on a restricted population of immune cells including activated T lymphocytes is critical for the further induction of proinflammatory cytokines and chemokines during an immune response, as well as for antibody production by CD40-expressing B lymphocytes. Fibroblasts also need to be able to respond to more generic danger signals. The intracellular apparatus for response to danger signals such as high levels of urate has recently been identified as the NOD-like receptor family made up of NOD (nucleotide-binding oligomerization domain) and NALP (NACHT domain, leucine-rich-repeat [LRR] domain, and pyrin domain [PYD]-containing protein) receptors. A high local level of urate released by dying cells triggers formation of the active NALP3 inflammasome complex, which results in release of IL-1.\(^{66}\) Expression of high levels of NOD-2 and NALP3 (cryopyrin) are seen in the RA synovium and can be induced in fibroblasts by TLR ligands and/or TNF\(^{67,68}\), although their relationship with as yet unidentified ligand molecules remains unclear; the NOD-2 ligand muramyl dipeptide is, however, present in the bowel, where it has a role in driving inflammation in Crohn’s disease.

**Role of Specialized Fibroblast Subsets within Tissue Microenvironments**

Combining surface markers with consistent function has been the key to decades of development in the field of leukocyte biology. By comparison, stromal cell biologists have had remarkably few such stable markers. However, this situation is now gradually changing and certain areas of developmental biology have spearheaded identification of putative markers (such as CD248) through approaches such as immunization of animals with human fibroblasts and digesting and identifying stromal cell subpopulations in tractable organ systems. One such example is the murine thymic stroma, in which subsets with both geographic and functional consistency have been identified. For instance, Link and colleagues\(^{49}\) identified CD45\(^+\), gp38\(^+\), positive cells, which were not lymphatic endothelium (CD31\(^−\)) in the thymus as T-zone fibroblastic reticular cells. This population of cells geographically restricted to the T zone provides a limited pool of essential homeostatic survival factors, IL-7 and CCL19, for T lymphocytes, serving a key niche function for which adaptive immune cells must compete.\(^{49}\)

A further subpopulation of specialized fibroblast-like cells of mesenchymal origin is the pericyte. These cells ensheath small blood vessels (arterioles, capillaries, and venules) and are involved in vasculogenesis, matrix stabilization, and immunologic defense. Pericytes have been hypothesized to represent the extralymphoid source of mesenchymal progenitor cells and express markers consistent with mesenchymal stem cells. Their further definition with newer stromal cell markers will be able to establish a mesenchymal progenitor cell niche.\(^{50}\)

**Fibroblast-like Synoviocytes in the Normal Synovium**

The normal synovium provides an excellent prototypic model of fibroblast subsets defined by known markers, some of which are responsive to disease. In health the synovium is a delicate, thin structure attaching bone and the joint capsule. It is divided into two layers. One is a two- to three-cell thick lining layer, which is formed in roughly equal proportions of CD68\(^−\), phagocytic type A macrophage like synoviocytes and type B mesenchymal, fibroblast-like synoviocytes (FLS). This layer must subserve a barrier function, and FLS must secrete lubricative substances including hyaluronic acid and lubricin, as well as secreting the lining layer matrix. The second layer is the sublining layer, which is composed of less densely packed fibroblasts and
macrophages in a loose tissue matrix along with blood vessel networks. FLS in the lining layer are associated with a number of cellular markers (see Table 15-1) including CD55 (decay accelerating factor [DAF]), VCAM-1 (which outside T cell–to-integrin interactions is generally only expressed by bone marrow fibroblasts providing support for the B cell developmental niche51), uridine diphosphoglucose dehydrogenase (UDPGD) reflecting the ability to synthesize hyaluronan, and the novel marker gp38.52 Sublining FLS are instead marked by the nonspecific cellular marker CD90 (Thy-1), which also recognizes endothelium, and by the recently discovered marker CD248, which marks both pericytes and stromal fibroblasts. Gp38 marks cells in the sublining region including lymphatic endothelium and pericytes (Figure 15-4). As mentioned earlier, the unique lining layer barrier function is supported not by a basement membrane and conventional tight junctions but by homotypic interactions between cadherin-11 molecules.53 Randomly assorted cells expressing classical cadherins such as cadherin-11 will sort themselves in a cadherin-specific manner, emphasizing their importance in the generation and maintenance of organ integrity. Cadherin-11 mediates selective association of mesenchymal rather than epithelial tissues, a function that is carried forward after embryogenesis in structures such as the joint lung and testis.54 Cadherin-11 knockout mice exhibit a hypoplastic synovial lining that lacks the normal numbers of synovial lining cells and is deficient in ECM quantity.55 Adhesion between type A and type B synoviocytes is maintained by ICAM-1:β2 integrin and VCAM-1:αvβ3 integrin interactions.

By virtue of their role in defining the geography of specialized tissues, fibroblasts and other stromal cells exist in living organisms within three-dimensional environments, whereas the vast majority of experiments performed using fibroblasts in the laboratory are still conducted within two-dimensional environments. Furthermore, fibroblasts are frequently grown in nonphysiologic stimuli such as serum, to which fibroblasts would not normally be exposed unless tissue damage were to occur. It has been shown that behavior is significantly different when cells are cultured in artificial three-dimensional environments.56 It is therefore all the more remarkable that fibroblasts cultured using conventional two-dimensional techniques retain characteristics such as positional memory and unique cytokine profiles. Recent work has addressed the issue of three-dimensional synovial models. In so-called “micromass cultures,” FLS but not dermal fibroblasts within matrigel spheres reproduced a lining layer structure with production of lubricin, co-culture within a lining layer–type structure of cells of monocyte origin, and expansion of the membrane on stimulation with proinflammatory stimuli such as TNF, some cells remained in a “sublining” zone of low density as well.57 FLS therefore have the ability to self-organize in a tissue organoid, which recapitulates some of the key features

Figure 15-4  Microscopic appearance of the synovium and stromal cell markers. The microscopic structure of hematoxylin and eosin-stained synovium is illustrated in D, indicating lining and sublining layers. This geographic structure is reflected in serial frozen sections of rheumatoid arthritis synovium stained for stromal markers (A-C, E-G). For reference, a nuclear stain is shown in H. A, CD248 stains only sublining fibroblasts. B, Prolyl-4-hydroxylase stains most populations of synovial fibroblasts. C, VCAM-1 (CD106) characteristically stains only the lining layer. E, CD90 (Thy-1) stains predominantly sublining layers but also strongly stains endothelial cells, outlining synovial vasculature. F, GP38 marks lining layer cells and a proportion of sublining cells. G, CD68 highlights macrophage-like synovial cells in the lining layer and resident tissue macrophages in the sublining layer.
of the synovium. This is further evidence of the robustness of epigenetic programming, which determines site and organ specialization.

**FIBROBLASTS IN RHEUMATIC DISEASES**

**Role of Fibroblasts in Persistent Inflammation**

Inflammatory reactions proceed against the backdrop of specialized stromal microenvironments. The response to tissue damage involves a carefully choreographed series of interactions among diverse cellular, humoral, and connective tissue elements. In order for an inflammatory lesion to resolve, dead or redundant cells that were recruited and expanded during the active phases of the response must be removed. In addition, resident fibroblasts attempt to repair damaged tissue. It is becoming increasingly clear that fibroblasts are not only passive players in immune responses but also active players in determining the switches that govern progression from acute to chronic inflammation, as well as those governing resolution or the progression to chronic, persistent inflammation. The “switch to resolution” is an important signal that permits tissue repair to take place and enables immune cells to return to draining lymphoid tissues (lymph nodes) in order for immunologic memory to become established. However, in chronic immune-mediated inflammatory diseases such as RA, fibroblasts contribute to the inappropriate recruitment and retention of leukocytes in a site- or organ-dependent manner, leading to tissue- and site-specific initiation and subsequent relapse of chronic persistent inflammatory disease, effectively a “switch to persistence.”

It is now recognized that fibroblasts themselves may undergo fundamental changes while responding to such environmental stimuli. It is known that during wound healing and under profibrotic conditions, some fibroblast-like cells are transformed into myofibroblasts, which are distinct from tissue fibroblasts in terms of both their phenotype and their behavior. The mechanisms underlying such persistent phenotypic change, which is maintained through cellular generations, are highly likely to involve epigenetic modifications of gene promoters and their closely related histones (see Chapter 22). This has been recently shown in both human and murine renal fibrotic disease, where hypermethylation of the promoter region of a ras oncogene inhibitor led to gene silencing, ras pathway activation, and hence persistent fibrogenesis. Such fibrotic transformation of fibroblasts is also characteristic of systemic sclerosis, a generalized fibrotic disorder that affects the skin and various internal organs such as the lungs, heart, and gastrointestinal tract (see Chapter 83). The overproduction of ECM components, particularly type I, III, VI, and VII collagen, by skin fibroblasts is a hallmark of this disease and is closely linked to the disease-specific activation of these fibroblasts. This pattern of activation includes not only a distinct profile of ECM overproduction but also altered responses to both inflammatory mediators and immune cells. Although the phenotype of fibroblasts in RA is not fundamentally profibrotic in this sense, the hallmark of these cells, both in vitro and in vivo, is also a persistently imprinted phenotype that is maintained even in the absence of continuous stimulation by inflammatory triggers or leukocytes.

**FIBROBLAST-LIKE SYNOVIOCYTES IN RHEUMATOID ARTHRITIS**

Inflammatory arthritis such as RA, the two compartments of the synovium undergo radical change. The lining layer undergoes dramatic hyperplasia, sometimes reaching 10 to 15 cells in depth with both type A and type B cell populations expanded and becoming merged with the sublining. At the articular borders of the synovium, the thickened synovial lining layer may become a mass of “pannus” tissue rich in FLS and osteoclasts, which aggressively invade the adjacent articular cartilage and subchondral bone. The sublining layer undergoes expansion as well, with sometimes huge infiltrates of inflammatory cells including macrophages, mast cells, T cells, B cells, and plasma cells in addition to dendritic cells. T and B lineage cells may remain in diffuse infiltrates or may coalesce into aggregates of cells varying from simple perivascular “cuffs” a few cells in diameter to structures resembling B cell follicles in up to 20% of samples. This increased activity is supported by further ECM production and neoangiogenesis, although the inflamed synovium remains in a state of relative hypoxia.

As mentioned earlier, cadherin-11 serves a vital role in preserving the integrity of the synovial lining layer, and cadherin-11 knockout mice display a hypoplastic lining layer. However, both knockout mice when exposed to the K/BxN serum transfer model and cultured fibroblasts with mutant cadherin-11 constructs demonstrate impaired invasiveness into cartilage, with a 50% reduction in overall inflammation in the mouse model. Cadherin-11 expression is also much stronger in RA than osteoarthritis (OA) or normal synovium. This unique structural molecule may therefore emerge as a therapeutic target.

**Persistent Activated Fibroblast Phenotype in the Rheumatoid Arthritis Synovium**

Increased expression of cadherin-11 is but one facet of the persistent, activated phenotype of rheumatoid FLS, which remains stable even after culturing in vitro for many months. These cells play a direct role in tissue damage through secretion of multiple matrix metalloproteinases and cathepsins, which degrade cartilage and bone tissues in the joint. In vitro functional assays such as the matrigel invasion assay produce intriguing results, in which the degree of invasion with a given in vitro cultured fibroblast sample correlates with the degree of radiographic progression seen in the joints of the patient from whose samples the fibroblasts were initially cultured. The most compelling evidence for a persistent phenotype is the attachment to and invasion of fibronectin-rich matrix such as human cartilage in the absence of functioning leukocyte immune cells in the SCID mouse model of arthritis. Here, fibroblasts in a tissue construct with human cartilage are implanted under the kidney capsule or skin of immune-deficient SCID or Rag mice. Multiple-passage cultured rheumatoid, but not osteoarthritic or normal FLS, invade and destroy the co-implanted human cartilage. This model has been used to explore the in vivo mechanisms governing invasiveness. Targeting MMP-1 and cathepsin L using ribozymes inhibits cartilage destruction. The effectiveness of glucocorticoids and the relative efficacy of different formulations of methotrexate in
preventing erosions have also been examined. Of most recent interest, fibroblasts implanted with cartilage will migrate to a contralateral cell-free implant and that subcutaneous, intraperitoneal, and intravenously injected fibroblasts will also migrate to sections of human cartilage, suggesting a tropism to damaged cartilage tissue.

This raises the question of which cell populations are grown from the synovium when tissue is digested and adherent cells are cultured in vitro: lining layer cells, sublining cells, or a mixture of both? This is a challenging question to answer from a methodologic perspective. However, we do know from transcriptomic approaches that what is cultured remains more stable than might be expected, with little transcriptional divergence over the first two to four passages and the level of differentially expressed genes between parallel cultures rising to over 10% only after passage 7.

These models demonstrate the remarkably stable and disease-specific phenotype of cultured RA synovial fibroblasts, which includes high basal and stimulated expression of signature cytokines such as IL-6 and chemokines (discussed later). RA synovial fibroblasts also express characteristic adhesion and immune-modulating molecules such as VCAM-1, galectin-3, and a specific repertoire of TLRs, which initiate innate immune cellular responses. A satisfactory molecular explanation for the stable phenotype of RA synovial fibroblasts has until recently evaded the field. However, epigenetic changes including DNA methylation, histone modifications such as acetylation, and microRNA expression have now been suggested to underlie the observed persistent changes in fibroblast gene transcription and posttranscriptional repression (see Chapter 22). Further characteristic aspects of the RA FLS phenotype and their biology are discussed extensively in Chapter 69.

**Interactions of Fibroblasts with Leukocytes**

**Recruitment of Inflammatory Infiltrates into the Joint**

Stromal elements such as synovial fibroblasts are subject to a proinflammatory cytokine network within the inflamed synovium. Direct-contact interactions with other infiltrating cells such as T lymphocytes lead to high levels of expression of many inflammatory chemokines (see Figure 15-3). Neutrophil-attracting chemokines are expressed at high levels by stimulated fibroblasts and include CXCL8 (IL-8); CXCL5 (epithelial-cell-derived neutrophil attractant 78, ENA-78); and CXCL1 (growth-related oncogene alpha, GROα). Monocytes and T cells are recruited by a range of chemokines found at high levels in the synovium; CXCL10 (IP-10) and CXCL9 (Mig) are highly expressed in synovial tissue and fluid. CXCL16 is also highly expressed in the RA synovium and acts as a potent chemoattractant for T cells. CCL2 (MCP-1) is found in synovial fluid and known to be produced by synovial fibroblasts; it is considered to be a pivotal chemokine for the recruitment of monocytes. CCL3 (Mip-1α), CCL4 (Mip-1β), and CCL5 (RANTES) are chemotactic for monocytes and lymphocytes and are known products of synovial fibroblasts. CCL20 (MIP-3α) is also overexpressed in the synovium and has a similar chemotactic profile via its specific receptor, CCR6. CX3CL1 (Fractalkine) is also widely expressed in the rheumatoid synovium. A number of chemokine receptors have been shown to differ between peripheral blood and synovial leukocytes, suggesting that they are enriched in the synovium either though their selective recruitment by endothelial-expressed chemokines or following upregulation by the microenvironment after their recruitment.

**Fibroblast Support for Leukocyte Survival**

Stromal cell support for the survival of leukocyte populations fulfills a physiologic role in certain organs within the body. The selective recruitment and support of hemopoietic subsets is an essential physiologic function of stromal cells in specific microenvironments. For instance, immature B lymphocytes are completely dependent on factors such as IL-6 produced by bone marrow stromal cells. Although the bone marrow niche plays a critical role in the early development of all hemopoietic leukocyte populations, it also acts as an active reservoir for terminally differentiated leukocyte subpopulations including CD4 and CD8 T cells and neutrophils. The bone marrow stromal microenvironment therefore maintains not only the selective survival, differentiation, and proliferation of all lineages of immature hemopoietic cells but also in some cases the survival of their mature counterparts. The stromal microenvironment plays a crucial role in the maintenance of such survival niches, which are not generic, but highly specific to certain organs and tissues, resulting in site-specific differences in the ability of different stromal cells to support the differential accumulation of leukocyte subsets.

In the case of an inflammatory response, successful resolution requires the removal of the vast majority of immune cells that were recruited and expanded during the active phase of the inflammation. A number of studies have shown that during the resolution phase of viral infections, the initial increase in T cell numbers in peripheral blood that is seen within the first few days is followed by a wave of apoptosis occurring in the activated T cells. This situation is mirrored within tissues, where apoptosis induced by the molecule Fas occurs at the peak of the inflammatory response and may be responsible for limiting the extent of the immune response. In contrast, the resolution phase appears to be principally triggered by cytokine-deprivation-induced apoptosis, during which leukocytes compete for a shrinking pool of survival factors provided by the microenvironment, leading to programmed death of those cells, which are surplus to requirements.

In RA the resolution phase of inflammation becomes disordered. Recent studies have shown that a failure of synovial T cells to undergo apoptosis contributes to the persistence of the inflammatory infiltrate. The T cell survival pathway shares all the essential hallmarks of a stromal cell, cytokine-mediated mechanism (high Bcl-Xl, low Bcl-2, and lack of cell proliferation). Type I interferons (interferons α and β), produced by synovial fibroblasts and macrophages, have been identified as one of the principal factors responsible for prolonged T cell survival in the rheumatoid joint (Figure 15-5). Interestingly, although type I interferon has been shown to be beneficial in multiple sclerosis (a disease in which tissue scarring and low levels of T cell infiltrates are observed), these results suggest that type I
Cell-cell interactions in the synovium. Synovial fibroblasts interact with multiple cell types in the rheumatoid arthritis (RA) synovium to maintain persistence of inflammation and continued joint destruction. A, Fibroblast-like synoviocytes in the RA synovial lining interact with macrophage-like synoviocytes through both secretion of soluble factors and cell surface receptor interactions to maintain lining layer structure and promote the activation of both cell types. Key soluble interactions include production of IL-1 and TNF by macrophages and IL-6 by fibroblasts. Adhesive interactions consist of integrin-receptor interactions as described in the text and the critical presence of homotypic interaction through cadherin-11.

B, Sublining synovial fibroblasts interact with numerous cell types including mast cells and plasma cells (not shown), T cells, B cells, interstitial macrophages, and endothelial cells, leading to their recruitment, retention, activation, and differentiation. Both cell surface receptor interactions and secreted mediators are important in this process. T cell–fibroblast interactions include T cell recruitment and retention by fibroblast-secreted chemokines such as CXCL12, CCL5, and CX3CL1 (fractalkine). In addition, fibroblasts may activate T cells through antigen presentation, co-stimulatory receptors (e.g., CD40, ICAM-1), and cytokine secretion. Fibroblast cytokines such as IL-6 and IL-15 may be particularly important for differentiation of the Th17 T cell subset, and secretion of IFN-β supports T cell survival. Fibroblasts, in turn, are activated by these cell surface interactions and by T cell cytokines, including IL-17 and IFN-γ. B cells are similarly recruited and retained through fibroblast secretion of chemokines such as CXCL12 and CXCL13 and through cell surface adhesion interactions (e.g., VLA-4 and VCAM-1). Critical survival and differentiation signals are maintained through fibroblast secretion of BAFF (Blys) and APRIL. Neutrophils and monococyte/macrophage lineage cells are also recruited by fibroblast chemokine production. Macrophages, in turn, help activate synovial sublining fibroblasts through the production of cytokines such as IL-1 and TNF. Finally, synovial sublining fibroblasts promote angiogenesis through the production of proangiogenic factors such as VEGF and PDGF and may help direct endothelial recruitment of inflammatory cells through secretion of cytokines such as IL-6. C, Pannus tissue, an extension of the hyperplastic synovial lining consisting of both activated MLS and FLS, actively degrades both cartilage and bone through production of matrix degrading enzymes such as MMPs and cathepsins. In addition, fibroblasts and T cells secrete RANKL, which promotes osteoclast differentiation and activation, leading to bone erosions. Furthermore, production of DKK1 inhibits Wnt signaling pathways, which normally promote anabolic osteoblast activity, preventing repair of bone erosions. CD40L, CD40 ligand; DKK1, dickkopf-1; FGF, fibroblast growth factor; ICAM-1, intercellular adhesion molecule-1; IFN-γ interferon-γ; IL, interleukin; MHC II, major histocompatibility complex type II; MMPs, matrix metalloproteinases; PDGF, platelet-derived growth factor; RANKL, receptor activator of nuclear factor κB ligand; TCR, T cell receptor; TNF, tumor necrosis factor; VCAM-1, vascular cell adhesion molecule-1; VEGF, vascular endothelial growth factor; VLA-4, very late antigen-4.
interferon is not likely to be a successful therapy for RA patients, a prediction that has been borne out in clinical trials.\textsuperscript{84} It is likely that this mechanism of stromal cell–induced leukocyte survival occurs in many chronic inflammatory conditions in which T cells accumulate. Not surprisingly, other leukocyte subpopulations have been shown to derive support from stromal cells. Although fibroblast support for T cell and B cell survival exhibits site-specific properties, neutrophil survival is dependent on prior cytokine activation of fibroblasts and shows no differences between fibroblasts taken from different anatomic sites.\textsuperscript{72} Plasma cells are, of course, rescued from apoptosis within the bone marrow stem cell niche,\textsuperscript{82} but mast cells of the gut are rescued by intestinal fibroblasts,\textsuperscript{83} whereas dermal fibroblasts maintain Langerhans-like cells in the skin.\textsuperscript{84}

**Fibroblast-Mediated Retention of Leukocytes in Tissue**

Although inhibition of T cell death by stromal cells at sites of chronic inflammation contributes to T cell accumulation, it is unlikely to be the only mechanism because lymphocytes should be able to leave the inflamed tissue during the resolution of inflammation, even if their death is inhibited. A number of studies have recently reported that the synovial microenvironment contributes directly to the inappropriate retention of T cells within the joint, by an active chemokine-dependent process. The presence of high levels of inflammatory chemokines, produced by stromal cells, is a characteristic of environments such as the rheumatoid synovium. However, recent data suggest that paradoxically constitutive chemokines, which are involved in the recruitment of lymphocytes to secondary lymphoid tissues, are ectopically expressed in immune-mediated inflammatory diseases. The constitutive chemokine CXCL12 (SDF-1) and its receptor CXCR4 emerged as unexpected but crucial players in the accumulation of T lymphocytes within the rheumatoid synovial microenvironment. This chemokine-receptor pair plays an important role, both in the constitutive traffic of lymphocytes and in the recruitment and retention of hemopoietic cells within the bone marrow. Unexpectedly, CD45RO\textsuperscript{+} T lymphocytes in the rheumatoid synovium were found to express CXCR4 receptors at high levels in the rheumatoid synovium. The CXCR4 ligand CXCL12 was highly expressed on endothelial cells at the sites of T cell accumulation.\textsuperscript{85,86} In addition, stromal cell–derived TGF-\beta is responsible for upregulation of CXCR4 receptors on T cells in the synovium.\textsuperscript{85} Evidence also suggests that the stability of lymphocyte infiltrates is reinforced by a positive feedback loop, whereby tissue CXCL12 promotes CD40 ligand expression on T cells, which in turn stimulates further CXCL12 production by CD40-expressing synovial fibroblasts. Furthermore, levels of CXCL12 secreted by synovial fibroblasts have recently been shown to be controlled in part by T cell–derived IL-17.\textsuperscript{87}

Therefore clear evidence supports the hypothesis that aberrant ectopic expression of constitutive chemokines such as CXCL12, CCL19, and CCL21 by synovial stromal cells contributes to the retention of T cells within the RA synovium.

Other cell constituents of the rheumatoid inflammatory infiltrate may be affected by the CXCL12/CXCR4 axis. Blades and colleagues\textsuperscript{88} have shown increased expression of CXCL12/CXCR4 by monocyte/macrophage cells in RA compared with osteoarthritis. In addition, using implanted human synovial tissue in SCID mice, they demonstrated that monocytes are recruited into transplanted synovial tissue by CXCL12.\textsuperscript{89} Contact-mediated B cell survival induced by synovial fibroblasts has also been shown to depend on CXCL12, BAFF/BLyS, and CD106 (VCAM-1)-dependent mechanisms that are independent of TNF.\textsuperscript{91,92} Overexpression of CXCL12 has also been identified as a distinctive feature of RA, as opposed to osteoarthritis synovia, using cDNA arrays. Data validating these findings in vivo have come from a collagen-induced arthritis model of RA in DBA/1 (interferon-\gamma receptor deficient) mice, where administration of the specific CXCR4 antagonist AMD3100 significantly ameliorated disease severity.\textsuperscript{90} In another murine collagen-induced arthritis model the small molecule CXCR4 antagonist 4F-benzoyl-TN14003 ameliorated clinical severity and suppressed delayed-type hypersensitivity (DTH) responses.\textsuperscript{93} The CXCL12/CXCR4 constitutive chemokine pair therefore seems to play an important role in lymphocyte retention in RA.

These experiments demonstrate that understanding the behavior of fibroblasts and leukocytes within microenvironments necessarily requires that we model the interactions of all the cellular populations concerned. An elegant example of this approach in vitro is the work of Lally and Smith\textsuperscript{94}, who developed a flow-based model of cellular recruitment to the rheumatoid synovium. Co-culturing fibroblasts from skin and RA synovial membrane with endothelial cells showed that IL-6 released from synovial (but not skin) fibroblasts was able to induce production of chemokines and adhesion molecules, resulting in greater neutrophil recruitment by synovial fibroblasts. Subsequent work interrogating the system using low-density gene arrays demonstrated that the effect of neutrophil-attracting chemokines such as CXCL5 released from synovial fibroblasts was dependent on the function of the chemokine transporter molecule DARC (Duffy antigen receptor for chemokines), which was also induced by fibroblast-to-endothelial cell co-culture.\textsuperscript{92}

**Constitutive Chemokines and Lymphoid Neogenesis**

RA is one of a number of inflammatory diseases in which the organization of the inflammatory infiltrate shares characteristics of lymphoid tissue. Follicular hyperplasia with germinal center formation can occur in autoimmune thyroid disease, myasthenia gravis, Sjögren’s disease, and RA and may occur during infection with Helicobacter pylori and *Borrelia burgdorferi*. The lymphoid infiltrates in the rheumatoid synovium can be divided into at least three distinct histologic groupings, varying from diffuse lymphocyte infiltrates through organized lymphoid aggregates to clear germinal center reactions. Moreover, there is conflicting evidence that such distinct histologic types correlate with other serum indicators of disease activity. This form of inflammatory lymphoid neogenesis relies on inappropriate but highly organized temporal and spatial expression by fibroblasts of the constitutive chemokines, particularly CXCL13 and CCL21, which are required for physiologic lymphoid
organogenesis. The elegant choreography of lymphocyte-stromal interactions within lymph nodes is organized by expression of adhesive and chemotactic cues in overlapping and combinatorial fashions. Once they have encountered new antigen, dendritic cells (DCs) specialized in the presentation of antigen to lymphocytes undergo a process of maturation under the local influence of inflammatory cytokines and bacterial and viral products. As a result, inflammatory chemokine receptors are downregulated and upregulation of the constitutive receptors CCR4, CCR7, and CXCR4 occurs, causing DCs to migrate into local draining lymphatics and thereby into peripheral lymph nodes. Trafficking of B and T cells is regulated by CXCL13 (BCA-1, B cell–attracting chemokine 1), its receptor CXCR5, and CCL21 and CCL19 (EBL-1-ligand chemokine, ELC), which are both CCR7 agonists. Within the lymph node CXCR5-bearing B cells are attracted to follicular areas, whereas T cells and DCs are maintained within parafollicular zones by local expression of CCL21 and CCL19. Some T cells that have been successfully presented with their cognate antigen by DCs then upregulate CXCR5, allowing them to migrate toward and interact with B cells.93–95

The genesis of lymphoid follicular structures in diseases such as diabetes and RA appears to rely on expression of such constitutive chemokines, in association with the lymphotoxins alpha and beta (LT-α and LT-β) and TNF.96 In this context it is important to note that transgenic animals overexpressing the TNF gene display increased formation of focal lymphoid aggregates and develop a chronic arthritis similar to RA.97 Clearly, one of the many mechanisms of action of anti-TNF therapy may be the dissolution of such aggregates. In transgenic mouse models, expression of CXCL13 in the pancreatic islets was sufficient for the development of T- and B-cell clusters, but because they lacked follicular dendritic cells, it was not sufficient for true germinal center formation.98 CCL21 does appear to be sufficient in some cases for lymph node formation; murine pancreatic islet models have demonstrated formation of lymph node–like structures in the presence of CCL21, and lymphoid infiltrates in response to CCL19 expression. The degree of lymphoid organization seen in the rheumatoid synovium has been shown to correlate with expression of the chemokines CCL21 and CXCL13, although these chemokines are also associated with less organized lymphoid aggregates.99

Expression of CCL21 is restricted to a population of perivascular fibroblastic reticular cells with common phenotype and function in secondary lymphoid and inflammatory aggregate tissues.100 CXCR5 is overexpressed in the rheumatoid synovium, consistent with a role in recruitment and positioning of B and T lymphocytes within lymphoid aggregates of the RA synovium. It therefore seems likely that expression of lymphoid-constitutive chemokines contributes significantly to the entry, local organization, and egress of lymphocytes in the RA synovium. It also seems that the ectopic expression of chemokines is a general characteristic of a number of chronic rheumatic conditions because another B cell–attracting chemokine CXCL13 (BCA-1) is inappropriately expressed by fibroblasts in the salivary glands of patients with Sjogren’s syndrome.101 Interestingly, the ectopic lymphoid structures seen in RA are capable of appropriate secondary lymphoid tissue structures including the production of class-switched high-affinity antibody production, as evidenced by the expression of activation-induced cytidine deaminase (AID), the enzyme required for somatic hypermutation and class-switch recombination (CSR) of Ig genes.102 CXCR3 expressing plasma cells are also present in the rheumatoid synovium, and their recruitment is once more supported by ectopic production of the CXCR3 ligand CXCL9 by fibroblasts, particularly in the sublining region where aggregates are located.103

Role of Fibroblast Subsets in Disease

It is possible that the various potential sources of expanded populations of synovial fibroblasts discussed earlier may correspond with functionally diverse fibroblast lineages and subpopulations. Kasperkovitz and collaborators9 showed using microarray analysis that the transcriptional profile of RA synovial fibroblasts clustered into two broad groups representing “high” (myofibroblastic) and “low” (growth factor producing) populations. These clusters were shown to be representative of the heterogeneity and of the degree of inflammation in the tissue of origin, suggesting that transcriptionally and functionally distinct populations of fibroblasts exist in joints. For example, some CD248+ cells may correspond with a pluripotential, stem cell–like population of pericytic cells lying in close apposition to endothelial cells, which provide a supply of new stromal cells during inflammation.104 Interestingly, deletion or removal of the intracellular portion of CD248 can reduce stromal cell accumulation and ameliorate models of arthritis such as murine collagen antibody-induced arthritis (CAIA).105 Furthermore, the prevailing conditions of hypoxia within the rheumatoid synovium may enhance expression of CD248, which is regulated by hypoxia-inducible factor-2 (HIF-2) binding to a hypoxia response element, and which in turn participates in angiogenesis.106 Whether such markers remain associated with functionally distinct subpopulations or simply contribute to a larger local pool of multipotential mesenchymal precursors is as yet unknown, but the discovery of markers apparently linked to function has provided the tools with which such questions can be answered. The results of studies demonstrating the association of stromal subpopulations with disease outcomes and response to therapy are eagerly awaited, as are the first attempts to target stromal markers therapeutically. The other group of markers that has been associated with synovial fibroblasts has come from the field of oncology.

Lessons Learned from Cancer

Alongside the field of inflammation, oncology has also been experiencing a surge in interest in the biology of fibroblasts and stromal cells, as well as the mechanisms by which they interact with primary transformed tumor cells.107 A number of important cytokines that contribute to cancerous transformation of healthy cells by so-called cancer-associated fibroblasts have been described. These include hepatocyte growth factor (HGF) and TGF-β. Crucially, tumor-associated fibroblasts appear able to transform normal, in addition to premalignant, cells.107 The importance of tumor-associated fibroblasts, termed cancer-associated fibroblasts (CAFs), has been demonstrated in breast cancer
where human breast cancer implants were unable to grow successfully when implanted into mice without their co-administration with human tumor-derived fibroblasts. Intriguingly, similar molecular signals have been implicated in the predilection for cancer cells to metastasize to certain sites. In particular, the ectopic expression and function of the CXCL12/CXCR4 ligand-receptor pair, in a manner reminiscent of RA, has been implicated in the persistence and tissue tropism of metastatic cells in breast cancer. Tumor-associated fibroblasts secrete CXCL12, resulting in increased promotion of carcinoma cell proliferation, migration, and invasion compared with control fibroblasts but also leading to recruitment of endothelial cell precursors. Furthermore, molecules that mark fibroblast subpopulations in the joint are associated with active, invasive cancer. These include the expression of tumor-associated stromal markers such as fibroblast activation protein (FAP), galectin-3, and S100A4. Interestingly, galectin-3 has been shown to be subject to epigenetic regulation. Both gp38 and CD248 are also heavily implicated in tumor progression.

These similarities between RA synovial fibroblasts and cancer-associated fibroblasts overlap with observations that the persistent phenotype of RA synovial fibroblasts itself includes elements normally associated with “transformed” cells. These include loss of density and anchorage limitation for growth, which usually curtails in vitro fibroblast culture, firm adherence to ECM components of cartilage, and the invasiveness that is most aptly demonstrated in the chimeric SCID mouse model. Another defining characteristic of RA synovial fibroblasts that helps to explain their phenotype is dysregulation of proto-oncogenes and tumor suppressor genes. Once again, epigenetic regulation is likely to underlie this phenotype. However, the precise mechanisms maintaining the persistent phenotype at a whole genome level are yet to be elucidated. Furthermore, RA is a systemic disease involving multiple joints; therefore whether the fibroblast phenotype results from a global change in fibroblast gene expression or whether the phenotype is locally imprinted by exposure to a characteristic cytokine, matrix, and cellular milieu is yet to be established. Recent data now appear to confirm that human RA synovial fibroblasts within the SCID mouse model may travel systemically through lymphatics and the bloodstream to unpopulated samples of cartilage and then invade. Therefore at least one possibility is that locally imprinted, “activated” fibroblasts may export destructive arthritis to joints where mild injury or immune response has occurred over time. In the cancer field, the concept of tumor stroma “normalization” has now become an accepted aspect of new oncology therapies. Clinical studies of angiogenesis inhibitors and antibodies against ECM components such as tenasin have been favorable, while inhibitors of MMPs, overexpression of TIMPs, and blockade of integrin signaling have all shown promise in preclinical trials. Results of studies examining the interactions between endothelial cells and their associated pericytes underlie the importance of targeting the stroma as a whole. Bergers and colleagues have shown that endothelial cells release PDGF, which induces VEGF production from pericytes leading to bidirectional conversations between the two cell types. Interrupting these conversations by using PDGF inhibitors proved to be more effective therapy than using VEGF inhibitors alone. Interestingly, although VEGF inhibitors lost their inhibitory effect in later-stage tumors, targeting of the pericytes helped even late-stage tumors to regress. The authors have subsequently shown that pericyte precursors are partly recruited from the bone marrow to tumor perivascular sites.

**SUMMARY**

Fibroblasts are structural mesenchymal cells that form the cellular infrastructure for most internal organs, as well as for bordering membranes such as the synovial membrane. They are prominently involved in the deposition and resorption of the ECM and thus are responsible for maintaining tissue homeostasis. However, fibroblasts are far more than structural, passively responding cells that build the “backbone” for organ-specific function. Rather, they are sensitive to environmental changes. They react in a specific manner to a variety of stimuli and are capable of actively influencing not only the composition of the ECM but also the cellular composition of tissues and barrier membranes. Under inflammatory disease conditions, fibroblasts act as organ-specific, innate immune system sentinel cells and are involved in the progression of organ damage, as well as in the switch from acute resolving to chronic persisting inflammation. We now know that functionally distinct fibroblast subsets exist and can be identified with new markers in order to understand better mechanisms of developmental patterning, wound healing, and persistent inflammatory responses, which appear to depend in large part on epigenetic modifications. This notion is particularly true for fibroblast-like synoviocytes, which play a critical role in the pathogenesis of RA and possess a characteristic, invasive, and activated phenotype. In addition to contributing to the recruitment of inflammatory cells to the joint, they modulate the survival and behavior of these cells and are, in turn, regulated by the newly recruited cells. More importantly, fibroblast-like synoviocytes are crucial components in the hyperplastic lining layer and in cartilage destruction. New data raise the possibility of epigenetically programmed aggressive cells exporting arthritis from inflamed to uninflamed joints in the early stages of arthritis, but at the same time offering the possibility of specifically targeting stromal subpopulations of choice.

**References**

60. Bechtel W, McCrohan S, Zeisberg EM, et al: Methylation deter-
Mast Cells

PETER A. NIGROVIC • DAVID M. LEE

KEY POINTS

Mast cells arise in the bone marrow, circulate as immature precursors, and develop into functional mast cells upon entering peripheral tissues.

The phenotype of mast cells is diverse, plastic, and governed by signals from lymphocytes, fibroblasts, and other elements of the microenvironment.

In healthy tissues, mast cells serve as immunologic sentinels and participate in both innate and adaptive immune responses to bacteria and parasites.

Mast cells accumulate in injured and inflamed tissue, where they may amplify or suppress inflammation.

Mast cells have been implicated in multiple autoimmune diseases, including inflammatory arthritis.

BASIC BIOLOGY OF MAST CELLS

Development and Tissue Distribution

Mast cells are distinctive in appearance. Ranging in size from 10 to 60 µM and with a centrally located round or oval nucleus, their abundant cytoplasm is filled with multiple small granules. They were named Mastzellen in 1878 by the German pathologist Paul Ehrlich, who believed incorrectly that they were overfed connective tissue cells (mästen, German, “to feed or fatten an animal”).

Electron microscopy reveals that the plasma membrane of mast cells exhibits multiple thin cytoplasmic extensions, providing a broad interface with surrounding tissue (Figure 16-1A). The tissue distribution of mast cells is extensive; within tissue, mast cells tend to cluster around blood vessels and nerves, and near epithelial and mucosal surfaces. They are also found in the lining of vulnerable body cavities such as the peritoneum and the diarthrodial joint. Given this localization, mast cells are among the first immune cells to encounter pathogens invading into tissue from the external world or via the bloodstream, consistent with their role as immune sentinel cells.

Mast cells are of hematopoietic origin, arising in the bone marrow and depositing in tissues after migrating through the bloodstream (Figure 16-2). Unlike most other myeloid cells, such as monocytes and neutrophils, mast cells do not terminally differentiate in the bone marrow but rather circulate as committed progenitors, bearing the surface signature CD34+/c-kit+/CD137. Further developmental details have been worked out most extensively in the mouse. Upon entering the tissues, murine mast cells may mature into classic granulated cells or may remain as ungranulated progenitors, awaiting local signals to differentiate fully. Comparison of murine lung and intestine has demonstrated that these tissues use distinct pathways to regulate the constitutive and inducible recruitment of mast cell progenitors, illustrating that mast cell homing is a precisely controlled process. Tissue homing is modulated prominently by lymphocytes, including regulatory T cells (Tregs).

Once resident in tissues, mast cells may live for many months. Unlike other myeloid lineage cells such as macrophages and neutrophils, mature mast cells remain capable of mitotic division, although recruitment of circulating progenitors appears to greatly exceed local replication as a pathway to expand the number of mast cells in a tissue. Mechanisms of reducing mast cell numbers include apoptosis, demonstrated in tissue mast cells deprived of the cytokine stem cell factor, a critical survival signal for mast cells. Under certain conditions, mast cells may emigrate...

Mast cell origin and differentiation. Mast cells arise in the bone marrow, circulate as committed progenitors, and differentiate into mature mast cells upon entering tissue. Human mast cells may be classified on the basis of granule proteases into tryptase+ mast cells (MC\textsubscript{T}) and tryptase+/chymase+ mast cells (MC\textsubscript{TC}), with characteristic tissue localization and mediator production. SCF, stem cell factor. (Adapted from Gurish MF, Austen KF: The diverse roles of mast cells, J Exp Med 194:F1–F5, 2001. Illustration by Steven Moskowitz.)

Mast Cell Heterogeneity: Common Progenitor, Multiple Subsets, and Phenotypic Plasticity

Although all types of mast cells derive from a common progenitor lineage, the phenotype of fully differentiated tissue mast cells is heterogeneous. Human mast cells are conventionally divided into two broad classes based on the protease content of their granules (see Figure 16-2). MC\textsubscript{TC} display rounded granules containing the enzymes tryptase and chymase; the smaller and more irregularly shaped granules of MC\textsubscript{T} contain tryptase but not chymase. MC\textsubscript{TC} also express other proteases, including carboxypeptidase and cathepsin G. MC\textsubscript{T} cells bearing only chymase have been reported but are controversial. These subtypes differ in tissue distribution. MC\textsubscript{TC} tend to be found in connective tissue, such as normal skin, muscle, intestinal submucosa, and synovium; MC\textsubscript{T} predominate in mucosal sites, including the lining of the gut and respiratory tract, although in fact both are present in many locations. Beyond protease signature, other differences between these subsets include their profiles of cytokine elaboration and cell surface receptor expression; however, tissue-specific phenotypic differences are noted within each type.

The relationship between MC\textsubscript{TC} and MC\textsubscript{T} mast cells is controversial. Are they committed subsets, akin to CD4 and CD8 lymphocytes, or functional states that mast cells assume under the influence of the microenvironment? In
the mouse, where an analogous distinction exists between connective tissue mast cells (CTMCs) and mucosal mast cells (MMCs), evidence for phenotypic plasticity is strong. Both in culture and in vivo, single CTMCs may differentiate into (or give rise to) MMCs and vice versa.\textsuperscript{17,18} Mast cells with intermediate protease expression are found, and serial observations suggest that exposure to an inflammatory stimulus can induce progressive change from one class to another, although whether this occurs at a single-cell level has not been definitively established.\textsuperscript{17} Similarly, in murine and human mastocytosis, clonally expanded mast cells display divergent phenotypes depending on the tissue of residence.\textsuperscript{20,21} In aggregate, these data favor the hypothesis that mast cells assume a particular phenotype under the control of local signals but can change radically if conditions change.

**Stem Cell Factor**

One of the most important signals from tissue to local mast cells is stem cell factor (SCF).\textsuperscript{10} The receptor for SCF, c-kit, is expressed widely on hematopoietic lineages early in differentiation, but among mature lineages, mast cells are one of the few cell types that express c-kit at a high level. Stimulation of mast cells by SCF promotes maturation and phenotypic differentiation, blocks apoptosis, and induces chemotaxis. It may also activate mast cells directly to release mediators. In both mouse and humans, SCF remains an irreplaceable survival signal for tissue mast cells. Accordingly, mice with defects in SCF or c-kit are strikingly deficient in mature tissue mast cells (examples include W/W\textsuperscript{v}, Sl/Sl\textsuperscript{I}, and W\textsuperscript{wsh} strains). Similarly, clonal mast cells obtained from patients with systemic mastocytosis commonly exhibit activating mutations in c-kit.\textsuperscript{22}

SCF occurs in two alternate forms resulting from differential mRNA splicing: soluble and membrane bound.\textsuperscript{10} The importance of this latter form is clear from Sl/Sl\textsuperscript{I} mice, which lack only the membrane-bound isoform yet exhibit very few tissue mast cells.\textsuperscript{23} SCF is synthesized by multiple lineages, including mast cells themselves. Expression by fibroblasts is likely especially important, given the intimate physical contacts observed between fibroblasts and mast cells in situ. Rodent mast cells co-cultured with fibroblasts demonstrate enhanced survival, connective tissue phenotypic differentiation, and heightened capacity to elaborate proinflammatory eicosanoids—effects mediated at least in part by direct contact, including interactions between SCF and c-kit.\textsuperscript{24,25} The extent of similar regulation in human mast cells is uncertain.\textsuperscript{26} Expression of SCF has also been documented on other lineages, including macrophages, vascular endothelium, and airway epithelium, and is likely a critical pathway by which tissues modulate the local mast cell population.

**T Lymphocytes and Other Cells**

It is interesting to note that T lymphocytes exert a profound effect on mast cell phenotype. SCID mice lacking T cells fail to develop mucosal mast cells, a defect that may be corrected by T cell engraftment.\textsuperscript{27} An analogous observation has been made in humans deficient in T cells as the result of congenital immunodeficiency or acquired immunodeficiency syndrome (AIDS). Intestinal biopsy in these patients shows that mucosal mast cells (MC\textsubscript{T}) are strikingly reduced, but connective tissue (MC\textsubscript{CT}) mast cells are present in normal numbers.\textsuperscript{28} The pathways by which T cells exert this striking effect are not defined, although it is clear that T cell cytokines such as interleukin (IL)-3, IL-4, IL-6, IL-9, and transforming growth factor (TGF)-β may have profound effects on the phenotype of mast cells matured in culture.\textsuperscript{29,31} By contrast, interferon (IFN)-γ inhibits mast cell proliferation and may induce apoptosis. These observations imply that cells recruited to an inflamed tissue may profoundly impact the phenotype of local mast cells. The rheumatoid synovium may well exemplify this phenomenon: Normally populated by MC\textsubscript{CT} mast cells, large numbers of MC\textsubscript{T} are identified in the inflamed synovium, typically in regions rich in infiltrating leukocytes, while MC\textsubscript{CT} reside in deeper, more fibrotic areas of the joint.\textsuperscript{32} It is interesting to note that T regulatory cells can also directly impact mast cell function, including receptor expression and degranulation.\textsuperscript{33,34}

Other cells beyond T cells may potentially interact with mast cells in the tissues. In particular, fibroblasts and mast cells commonly demonstrate close physical interactions.\textsuperscript{35} Beyond SCF, fibroblasts elaborate cytokines such as the IL-1 family member IL-33, which can exert determinative effects on mast cell protease expression and effector phenotype.\textsuperscript{36,37}

**Different Functions for MC\textsubscript{T} and MC\textsubscript{CT} Mast Cells?**

The preservation of distinct types of mast cells in multiple species implies distinct and nonoverlapping roles for these subtypes. However, our understanding of functional differences between MC\textsubscript{T} and MC\textsubscript{CT} remains limited. One hypothesis is that MC\textsubscript{T} play a proinflammatory role and MC\textsubscript{CT} specialize in matrix remodeling.\textsuperscript{38} This hypothesis makes sense of (1) the promotion of MC\textsubscript{T} development by T cells patrolling the tissues; (2) the partitioning of MC\textsubscript{T} and MC\textsubscript{CT} mast cells to inflamed and fibrotic areas respectively; and (3) the preferential expression of the proinflammatory mediators IL-5 and IL-6 by MC\textsubscript{T} and the profibrotic IL-4 by MC\textsubscript{CT}.\textsuperscript{39} Not all observations fit comfortably into this dichotomy, however. For example, the potent proinflammatory anaphylatoxin receptor C5aR (CD88) is expressed on MC\textsubscript{CT} but not on MC\textsubscript{T}.\textsuperscript{40} Ultimately, too little is known about the actual functional importance of these subsets to permit firm conclusions.

**Mast Cell Activation**

**IgE**

The canonical pathway to mast cell activation is via IgE and its receptor FceRI. With a \(K_d\) of \(10^{-7}\)M, this receptor is essentially constantly saturated with IgE at typical serum concentrations.\textsuperscript{41} Such binding not only sensitizes mast cells to the target antigen but also helps to promote mast cell survival and, in some cases, cytokine production.\textsuperscript{42,43} Crosslinking of FceRI-bound IgE by multivalent antigen induces a brisk and vigorous response. Within minutes, granules within the mast cell fuse together and with the surface membrane create a set of labyrinthine channels that allow...
rapid release of granule contents (see Figure 16-1B). This compound exocytosis event, termed anaphylactic degranulation, is followed within minutes by the elaboration of eicosanoids newly synthesized from arachidonic acid cleaved from internal membrane lipids. Finally, signals transduced via FcεRI induce the transcription of new genes and the elaboration of a wide range of chemokines and cytokines (Figure 16-3). Upon termination of the stimulation event, the surface membrane closes over the granule-formed channels; these subsequently bud off within the cytoplasm, re-creating discrete granules using the original membranes. These granules become recharged with mediators through a process that occurs gradually over days to weeks.

**IgG and Immune Complexes**

IgE is only one among many pathways of mast cell activation. One key trigger for mast cell activation in both human and mouse is IgG, acting via receptors for the Fc portion of IgG (FcγR). The importance of this pathway was demonstrated first in mice rendered genetically deficient in IgE. Contrary to expectations, these animals remained susceptible to anaphylaxis mediated through IgG and the low-affinity IgG receptor FcεRII. The human counterpart of this receptor, FcγRIIa, is equally capable of inducing activation of human mast cells. Human mast cells exposed to IFN-γ may also be induced to express the high-affinity IgG receptor FcεRI, rendering them susceptible to IgG-mediated activation, although expression of this receptor in vivo has not been shown.

These IgG receptors contribute to involvement of mast cells in IgG-driven diseases. Thus, in the mouse, mast cells participate in IgG-mediated immune complex peritonitis, the cutaneous Arthus reaction, and experimental murine bullous pemphigoid. Activation via Fc receptors also mediates mast cell participation in antibody-mediated murine arthritis.

**Soluble Mediators and Cell-Cell Contact**

Mast cells may coordinate with immune and nonimmune lineages via mechanisms beyond antibody response, including soluble mediators and surface receptors. Examples of such signals include the cytokine tumor necrosis factor (TNF) and the neurogenic peptide substance P, which can induce mast cell degranulation. Physical contact with other cells can also induce mast cell activation. For example, CD30 on lymphocytes can interact with CD30L on mast cells to induce the production of a range of chemokines.

It is interesting to note that ligation of CD30L does not induce the release of granule contents or lipid mediators, illustrating the selectivity of response of which mast cells are capable.

**Danger and Injury**

Mast cells are equipped to recognize danger in the absence of guidance from other lineages via a range of pathogen receptors, including multiple Toll-like receptors (TLRs) and CD48, a surface protein recognizing the fimbrial antigen FimH. These receptors are implicated in the response of mast cells to pathogens. Mast cells may also be activated through complement, including the anaphylatoxins C3a and C5a.

Finally, mast cells can respond directly to physical stimuli such as trauma, temperature, and osmotic stress. Together, these receptors enable mast cell involvement in a broad range of immune and nonimmune processes.

**Inhibitory Signals for Mast Cells**

As with other immune lineages, mast cells are subject to both negative and positive regulation. Examples of inhibitory receptors on the surface of mast cells include the IgG receptor FcγRIIB and the integrin-binding immunoglobulin superfamily member gp49b1. The importance of these receptors is demonstrated in genetically deficient animals. Mice lacking FcγRIIB demonstrate a striking propensity to activation via both IgG and IgE (which bind with low affinity to FcγRIIB as well as to FcεRI), but gp49b1-null mice are unusually susceptible to IgE-mediated anaphylaxis. Of note, no human orthologue of gp49b1 is known, thus the relevance of this pathway in modulating MC activity in humans remains unclear. Nevertheless, modulating the surface expression of inhibitory receptors serves as an important mechanism for regulation of the activation threshold of mast cells in tissues.
Mast Cell Mediators

Granule Contents: Proteases, Amines, Proteoglycans, and Cytokines

Mature mast cells package a range of mediators in their granules, ready for immediate release through fusion with the surface membrane. The most abundant of these are the neutral proteases, named for their enzymatic activity at neutral pH, but vasoactive amines, proteoglycans such as heparin, and pre-formed cytokines play distinct roles in the biologic consequences of mast cell degranulation. The release of these mediators is not all or none. In addition to anaphylactic degranulation, mast cells may release only a few granules at a time in a process termed piecemeal degranulation. Further, mast cells can release one type of granule but not another. Alternately, mast cells may be induced to elaborate cytokines and chemokines with no release of granule contents, as illustrated by activation via CD30L. Thus, although the mast cell is well equipped to release large volumes of pre-formed mediators, it is equally capable of responses tailored to the activating stimulus.

Tryptase. Named for its enzymatic similarity to pancreatic trypsin, tryptase is the most abundant granule protein in human mast cells. It is an essentially specific marker for mast cells, synthesized in scant amounts by basophils but by no other lineage. The enzyme found in granules is the β-isomer, which is enzymatically active upon formation of a homotetramer that relies on the scaffolding function of the proteoglycan heparin. Mast cells also synthesize α-trypsin, a protein incapable of forming homotetramers and so enzymatically inactive. Unlike β-trypsin, the α-isomer is not stored in granules but is constitutively released into the circulation, where its function is unknown. The distinction between tryptase isomers is important for diagnostic reasons: As a marker of degranulation, systemic levels of β-trypsin serve as a marker of recent anaphylaxis. By contrast, α-trypsin levels reflect total body mast cell load and serve as a useful biomarker in systemic mastocytosis.

Tryptase directly cleaves structural proteins such as fibronectin and type IV collagen and enzymatically activates stromelysin, an enzyme responsible for activating collagenase. Tryptase also promotes hyperplasia and activation of fibroblasts, airway smooth muscle cells, and epithelium. Cleavage of protease-activated receptors such as PAR-2 may contribute to some of these activities, although other studies have documented PAR2-independent tryptase activation of mesenchymal cells. In aggregate, these effects suggest an important role for tryptase in matrix remodeling.

A further contribution to the inflammatory milieu is suggested by the capacity of tryptase to promote neutrophil and eosinophil recruitment and to cleave C3, C4, and C5 to generate anaphylatoxins. It is interesting to note that tryptase can potentially downregulate inflammation by cleaving IgE and IL-6.

Chymase. This chymotrypsin-like neutral protease is found in the MC_Tc subset of human mast cells, packaged within the same granules as tryptase. Similar to tryptase, chymase can cleave matrix components and activate stromelysin, although it can also activate collagenase directly, suggesting a role in matrix remodeling. Chymase can influence cytokine function, with the capacity to cleave pro-IL-1β to generate active cytokine, as well as to inactivate proinflammatory cytokines such as IL-6 and TNF.

Vasoactive Amines. Human mast cells are capable of synthesizing and storing the biogenic amines histamine and serotonin, implicated in vascular leak. Histamine, by far the more abundant, is a vasoactive amine found in both MC_T and MC_Tc mast cells, although it is not unique to this lineage. Histamine is involved in the wheal-and-flare response to cutaneous allergen challenge via augmented vascular permeability, transendothelial vesicular transport, and neurogenic vasodilation. These effects are mediated principally via the H1 receptor. Three other histamine surface receptors, H2 through H4, are distributed widely on immune and nonimmune lineages, with effects as diverse as gastric acid secretion, Langerhans cell migration, and B cell proliferation.

Heparin and Chondroitin Sulfate E. These large proteoglycans enable the ordered packing of mediators within human mast cell granules. Negatively charged carbohydrate side chains complex tightly with positively charged proteins, allowing very high concentrations of β-trypase and other proteases. Heparin, produced exclusively by mast cells, facilitates the activity of tryptase by making possible proteolytic self-activation within the granule and stabilizing the active tetrameric form of this enzyme. Heparin also has a wide range of effects beyond the mast cell. Heparin is potently angiogenic. Heparin binding activates antithrombin III, providing the basis for use as an anticoagulant, while inhibiting chemokines and both classical and alternative pathways of complement activation, as well as the function of Treg cells. The physiologic role of these extracellular activities of mast cell–derived heparin is uncertain.

Pre-Formed Cytokines. Mast cells are able to store certain cytokines in their granules for rapid release. The first of these to be documented was TNF. In the mouse, this pool of TNF is implicated in the rapid recruitment of neutrophils to the peritoneum during peritonitis. Other cytokines that may be stored in granules include IL-4, IL-16, basic fibroblast growth factor (bFGF), and vascular endothelial growth factor (VEGF).

Newly Synthesized Mediators: Lipid Mediators, Cytokines, Chemokines, and Growth Factors

Beyond pre-formed mediators stored within granules, activated mast cells elaborate a range of mediators that are generated de novo. These mediators are released minutes to hours after stimulation, broadening and extending the impact of activated mast cells on surrounding tissues.

Lipid Mediators. Within minutes of activation, mast cells begin to release metabolites of membrane phospholipids. This process is rapid because the relevant enzymes, beginning with phospholipase A2, responsible for harvesting phospholipids from the outer leaflet of the nuclear membrane, are already present in the cytoplasm and need only to be activated through signals mediated by calcium flux and the phosphorylation of intracellular messengers. The hallmark prostaglandin of human mast cells is prostaglandin D2 (PGD₂), which is capable of inducing bronchoconstriction, vascular leak, and neutrophil recruitment. Smaller quantities of other prostaglandins as well as thromboxane
are also made. Mast cell–derived leukotrienes have similar but generally more potent activity. Leukotriene C_4 (LTC_4) is the major leukotriene species generated by human mast cells; together with its metabolites LTD_4 and LTE_4, it serves as a potent inducer of vascular leak. Smaller quantities of the chemotaxins LTβ and platelet-activating factor (PAF) are also generated. The particular profile of lipid mediators produced by mast cells can change with local environmental signals and the resulting state of differentiation. Thus, mast cells from skin generate PGD_2 in excess of LTC_4, and both species are elaborated in roughly equal proportions by mast cells isolated from lung and osteoarthritic synovium.65

Cytokines, Chemokines, and Growth Factors. Within hours of activation, mast cells begin to elaborate newly synthesized mediators as the end result of induced gene transcription and translation. The range of such mediators is broad (see Figure 16-3). They include the canonical proinflammatory mediators TNF, IL-1, and IL-6; the Th2 cytokines IL-4, IL-5, IL-10, and IL-13; chemotactic factors including IL-8, MIP-1α, and regulation upon activation normal T cell expressed and secreted (RANTES); and growth factors for fibroblasts, blood vessels, and other cells such as bFGF, VEGF, and platelet-derived growth factor (PDGF).66 As noted earlier, some of these may also be stored pre-formed in granules for rapid release. The panel of mediators generated depends on the state of differentiation as well as the activating signal, and may occur in the absence of degranulation.

ROLE OF MAST CELLS IN HEALTH AND DISEASE

Our understanding of the role of mast cells in health and disease has been aided greatly by the availability of mice lacking mast cells through defects in the SCF/c-kit axis. Although these mice exhibit multiple phenotypic abnormalities, they are viable, excluding an obligate basal role for mast cells in the structure and function of most tissues. Yet under physiologic stress, such as imposed by experimental models of disease, multiple differences from wild-type become evident. In many cases, these abnormalities may be corrected by engraftment with cultured mast cells,70 directly implicating mast cells in a remarkably broad range of disease processes (Table 16-1). Interpretation of such experiments is limited by incomplete physiologic restoration of the mast cell compartment and by residual effects of deficient c-kit signaling in other lineages. However, together with in vitro experiments and careful observation of normal subjects, animal experiments in mast cell–deficient mice have contributed greatly to recent progress in our understanding of mast cell physiology and pathophysiology.

### Table 16-1 Participation of Mast Cells in Murine Models of Disease

<table>
<thead>
<tr>
<th>Beneficial to Host</th>
<th>Reference</th>
<th>Harmful to Host</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiogenesis</td>
<td>147</td>
<td>Anaphylaxis*</td>
<td>98</td>
</tr>
<tr>
<td>Anxiety control</td>
<td>177</td>
<td>Arthritis*</td>
<td>163, 164</td>
</tr>
<tr>
<td>Bacterial cystitis</td>
<td>59</td>
<td>Aortic aneurysm*</td>
<td>178, 179</td>
</tr>
<tr>
<td>Bacterial peritonitis*</td>
<td>94, 103</td>
<td>Asthma*</td>
<td>100</td>
</tr>
<tr>
<td>Bone remodeling</td>
<td>142</td>
<td>Atherosclerosis*</td>
<td>180</td>
</tr>
<tr>
<td>Dermatitis</td>
<td>133</td>
<td>Atrial fibrillation</td>
<td>181</td>
</tr>
<tr>
<td>Envenomation*</td>
<td>182</td>
<td>Burn</td>
<td>183</td>
</tr>
<tr>
<td>Glomerulonephritis*</td>
<td>184</td>
<td>Bullous pemphigoid*</td>
<td>52</td>
</tr>
<tr>
<td>Graft tolerance*</td>
<td>132</td>
<td>Cardiomyopathy</td>
<td>185</td>
</tr>
<tr>
<td>Intestinal epithelial barrier*</td>
<td>186</td>
<td>Colon polyps</td>
<td>187</td>
</tr>
<tr>
<td>Parasites, intestine</td>
<td>106, 108</td>
<td>Dermatitis, irritant*</td>
<td>188</td>
</tr>
<tr>
<td>Parasites, muscle</td>
<td>109</td>
<td>Dermatitis, sunburn</td>
<td>189</td>
</tr>
<tr>
<td>Parasites, skin</td>
<td>190</td>
<td>Gastritis</td>
<td>191</td>
</tr>
<tr>
<td>Thromboembolism</td>
<td>192</td>
<td>Glomerulonephritis*</td>
<td>193</td>
</tr>
<tr>
<td>Tumor suppression*</td>
<td>194</td>
<td>Immune complex peritonitis*</td>
<td>50</td>
</tr>
<tr>
<td>Wound healing*</td>
<td>136</td>
<td>Ischemia-reperfusion injury</td>
<td>195, 196</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple sclerosis*</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neurogenic inflammation*</td>
<td>123, 124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Obesity*</td>
<td>197</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Peritonitis, irritant*</td>
<td>198</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Peritoneal adhesions</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pneumonitis</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scleroderma</td>
<td>139, 140</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tumor angiogenesis</td>
<td>201, 202</td>
</tr>
</tbody>
</table>

Mast cells are implicated in these processes by virtue of phenotypic abnormalities in mast cell–deficient mice, or in mice lacking mast cell–specific mediators. The asterisk (*) indicates that the phenotype has been shown to be reversible by engraftment with cultured mast cells, providing more direct evidence of a role for this lineage.
Mast Cells in Nonallergic Inflammation

Pathogen Defense: Mast Cells as Sentinels of Innate Immunity

The involvement of mast cells in atopic disease is clear but does not explain their remarkable evolutionary conservation. Rather, mast cells must somehow contribute to the survival of the organism. The most probable mechanism by which mast cells convey a survival advantage is defense against infection. This role is reflected in the localization of mast cells near epithelial surfaces, around blood vessels, and in other locations of potential invasion by pathogens.

Mast cells are competent defensive cells against bacteria. They express TLRs and other receptors against bacterial antigens, and upon activation are able to phagocytose bacteria and generate antimicrobial molecules such as cathelicidin. However, given their relatively small numbers, the most important function of mast cells in immune defense is to serve as sentinels, monitoring for early traces of infection and rapidly mobilizing neutrophils and other inflammatory cells when needed. Such a role has been clearly demonstrated in mouse models of bacterial peritonitis, in which mast cell–deficient animals exhibit high mortality. This susceptibility correlates with delayed recruitment of neutrophils via TNF and leukotrienes; both neutrophil influx and survival may be restored by correction of the mast cell deficit, although in severe infection, mast cell TNF may actually contribute to mortality. Clearance of bacteria from the lung is delayed in mast cell–deficient mice and can be similarly restored. Analogous observations have been made in other models of bacterial infection. Thus, mast cells may play an important role in defense of the host against bacterial infection.

Mast cells are also implicated in the defense against parasites. Mast cell–deficient animals exhibit abnormal clearance of multiple parasites from gut and skin, in a manner promoted by IgE. The mechanism of this defense remains uncertain but may include direct attack upon pathogens, recruitment of inflammatory lineages such as neutrophils and eosinophils, and lysis of tight junctions in the mucosal lining to facilitate the expulsion of helminths.

Mast Cells and the Adaptive Immune Response

In addition to recruiting innate effector cells, mast cells mobilize T and B lymphocytes, the adaptive arm of the immune system. Mast cells may express MHC II, as well as co-stimulatory molecules such as CD80 and CD86, rendering them effective antigen-presenting cells for CD4 T cells. Mast cells can also mobilize and potentiate CD8 T cell responses. They may migrate from peripheral tissues to lymph nodes carrying antigen and may contribute to the recruitment of T cells to lymph nodes via mediators such as MIP-1β and TNF, as well as suppression of Treg responses. Indeed, infection-induced lymph node hyperplasia is abrogated in the absence of mast cells. Further, mast cells can recruit CD4 and CD8 effector T cells to peripheral tissues via leukotriene B4, among other mediators. Finally, mast cells can contribute to the migration of cutaneous Langerhans cells and other dendritic cells to lymph nodes via mediators including histamine.

By means of the inducible expression of CD40L and cytokines, mast cells may stimulate B cells and induce class switching to IgA or IgE. The physiologic importance of these effects will vary with circumstances. For example, under some conditions delayed-type hypersensitivity responses in skin are mast cell dependent, but under others mast cells appear to play no role. The potential importance of the mast cell in adaptive immunity is highlighted by the recent demonstration that mast cell activators are effective vaccine adjuvants.

Neurogenic Inflammation

In addition to their perivascular localization, mast cells cluster near and even within peripheral nerves. A discrete function for them in these locations has not yet been identified; although the potential for bidirectional neuroimmune interaction is clear. Mast cell mediators such as histamine may directly activate neurons, and mast cells residing near stimulated neurons may be induced to degranulate. Indeed, vascular leak and neutrophil infiltration arising from infiltration of skin with the neurogenic mediator substance P are mediated by mast cells. Thus neurons may recruit mast cells as local effectors to initiate neurogenic inflammation.

Autoimmune Disease

Reconstitution experiments in mast cell–deficient mice have implicated mast cells in a variety of pathologic conditions (see Table 16-1). These include murine models of autoimmune diseases such as bullous pemphigoid, multiple sclerosis, scleroderma, and inflammatory arthritis. In pemphigoid, mast cells triggered via IgG antibodies against a hemidesmosomal antigen recruit neutrophils that are responsible for blister formation. The role of mast cells in murine experimental autoimmune encephalomyelitis (EAE) is more complex. Although the resistance of W/Wv mice to EAE corrects with mast cell engraftment, these cells fail to repopulate the brain and spinal cord, indicating that mast cells are not obligate local effector cells in this model. One mechanism for this activity appears to be promotion of the adaptive immune response, because mast cell engraftment into W/Wv animals improves T cell responses to immunization with the inciting myelin antigen. The contribution of mast cells to human scleroderma remains unknown. The participation of mast cells in arthritis is discussed in detail later.

MAST CELLS AS ANTI-INFLAMMATORY CELLS

Within the last few years, it has become evident that mast cells may also help moderate the immune response. One mechanism for this effect is degradation of proinflammatory mediators. Mast cell proteases may cleave and inactivate the cytokines IL-5, IL-6, IL-13, and TNF, as well as endothelin-1 and the anaphylatoxin C3a. The importance of this activity has been demonstrated in a murine sepsis model, in which mast cells reduced mortality...
by restraining excess inflammation in a protease-dependent manner. More broadly, mast cells are capable of producing mediators such as IL-10 that have immunosuppressive activity; even otherwise proinflammatory mediators such as TNF and granulocyte-macrophage colony-stimulating factor (GM-CSF) can be immunosuppressive under appropriate circumstances. Thus, mast cells promote immunologic tolerance to skin grafts and limit tissue inflammation related to ultraviolet-light injury.

**MAST CELLS AND CONNECTIVE TISSUE**

**Wound Healing and Tissue Fibrosis**

Mast cells have long been noted to accumulate at the borders of healing wounds. In normal human subjects undergoing experimental wounding and recurrent biopsies, mast cell numbers increase sixfold by day 10 after initial incision. These mast cells localize preferentially to fibrotic areas of the wound and strongly express IL-4, a cytokine capable of inducing fibroblast proliferation and collagen synthesis. In vitro studies confirm the stimulatory effects of mast cells on fibroblast growth; candidate fibroblast mitogens in addition to IL-4 include tryptase, histamine, LTC₄, and bFGF. Indeed, mast cell–deficient W/W° animals exhibit delayed contracture and healing of skin wounds in a manner repairable by local engraftment with cultured mast cells.

Mast cells also accumulate in sites of pathologic fibrosis, including the skin and lungs of patients with scleroderma. Because experimental skin fibrosis proceeds in mast cell–deficient mice with only relatively subtle differences in intensity or kinetics, it is unlikely that mast cells are an obligate effector lineage in human scleroderma, although they may contribute to disease progression.

**Bone**

Mast cells are also implicated in the remodeling of bone. Mast cells accumulate in sites of healing fracture; under normal circumstances, they may contribute productively to normal bone turnover. However, mast cells accumulate in osteoporotic bone, and systemic osteoporosis is a known complication of systemic mastocytosis. Heparin is a potentially important mediator of bone loss, in that it directly promotes differentiation and activation of osteoclasts. Mast cell products such as IL-1, TNF, and MIP-1α have similar activity.

**Angiogenesis**

A final and potentially quite important activity of mast cells on the stroma is the promotion of angiogenesis. Mast cells are not required for development of the normal vasculature, as is evident in the viability of mast cell–deficient mice. However, mast cells cluster at sites of early blood vessel growth in tumors and contribute appreciably to physiologic angiogenesis under certain experimental conditions. Heparin was the first proangiogenic mast cell mediator identified; bFGF and VEGF are other potent stimulators of endothelial migration and proliferation.

**MAST CELLS IN ARTHRITIS**

The normal synovium is populated by a limited number of mast cells. These cells are not found in the immediate lining layer but rather reside in the synovial sublining, near blood vessels and nerves, constituting almost 3% of cells within 70 µM of the synovial lumen. In both mice and humans, their phenotype is principally MC⁰⁰, similar to mast cells found in most other connective tissue sites. The several-fold increased density of mast cells in the immediate vicinity of the synovial lining, compared with more distant connective tissue, supports the hypothesis that mast cells contribute to surveillance of the articular cavity. Extrapolating from the activity of mast cells near other vulnerable body cavities, such as the peritoneum, it is likely that synovial mast cells help to monitor the joint for early evidence of infection.

Under conditions of arthritis, the population of synovial mast cells may expand remarkably (Figure 16-4). More than two-thirds of synovial specimens from patients with rheumatoid arthritis (RA) exhibit abnormal numbers of mast cells, averaging in excess of tenfold above normal. Consistent with these histologic findings, synovial fluid from rheumatoid joints contains appreciable quantities of histamine and tryptase. Unlike the normal joint, in the RA joint both subtypes of mast cells are present in roughly equal numbers; MC⁰ cells are located nearer to the pannus and infiltrating leukocytes, and MC⁰ cells cluster in deeper, more fibrotic areas of the synovium. Mast cells have been noted near the junction of pannus and cartilage. Rare mast cells are also identified in synovial fluid. The absence of mitotic figures and of staining for the proliferation antigen Ki-67 in this population suggests that they arise not from local replication but rather by recruitment of circulating progenitors. Although the signals driving this recruitment are unknown, inflammatory cytokines such as TNF enhance expression of the mast cell chemotactic and survival factor SCF on synovial fibroblasts, suggesting one

![Figure 16-4](image-url) Mast cells in the rheumatoid synovium. Stained red by an antibody against tryptase, mast cells are abundant in this synovial biopsy from a patient with chronic rheumatoid arthritis. Note the proliferation of mast cells in the synovial sublining. (Reproduced with permission from Nigrovic PA, Lee DM: Synovial mast cells: role in acute and chronic arthritis, Immunol Rev 217:19–37, 2007.)
mechanism for this dramatic expansion. Indeed, degree of inflammation is the best predictor of the number of mast cells within the joint. Incompletely identified factors in RA synovial fluid can potently promote mast cell differentiation and growth.

Hyperplasia of the mast cell population is not specific for rheumatoid arthritis but is observed in a wide range of inflammatory joint disorders (Table 16-2). Expansion is also noted in osteoarthritis (OA), often to densities observed in RA. The levels of histamine and tryptase in OA synovial fluid are also comparable. It is interesting to note that unlike in RA, expansion in OA results from an increase in the numbers of MC\(_T\) mast cells, the subtype generally associated with T cells and inflammation.

**Mast Cells in Acute Arthritis: Insights from Animal Models**

Recent experimental work in mice has begun to shed light on the role of mast cells in inflammatory arthritis. Several mast cell–deficient strains demonstrate striking resistance to arthritis induced by IgG autoantibodies—a defect that may be repaired by engraftment with cultured mast cells expressing receptors for IgG and C5a. A number of mechanisms contribute to this arthritogenic activity. First, mast cells induce vascular permeability, facilitating entry of autoantibody into the joint. Second, mast cells release proinflammatory mediators including IL-1 that help to establish inflammation, presumably via effects on endothelium and other local populations such as macrophages and fibroblasts. These actions appear to be most critical at the initiation of disease, constituting a “jump start” for acute inflammation within the joint. This function is in line with the activity of mast cells in other models of IgG-mediated disease, such as IgG-mediated immune complex peritonitis, murine bullous pemphigoid, and anaphylaxis. In each of these models, mast cells resident in tissue for the purpose of immune defense become co-opted by autoantibodies to initiate inflammatory pathology (Figure 16-5, top).

**Mast Cells in Chronic Arthritis**

In contrast to the acute phase of joint inflammation, the contribution of mast cells in the context of established arthritis is less well understood. The sheer numbers of these cells in arthritic synovium implies a substantial role. Taking into account the spectrum of mast cell activity elsewhere, it is likely that mast cells participate both in the inflammatory process and in the mesenchymal response (Figure 16-5, bottom).

An ongoing contribution of mast cells to inflammatory arthritis is suggested by several observations. First, as noted, prominent among infiltrating synovial mast cells are MC\(_T\) cells, typically associated elsewhere with the elaboration of cytokines such as IL-6 with documented pathogenic activity in rheumatoid arthritis. Immunofluorescence staining has identified TNF and IL-17 in RA synovial mast cells and elaboration of other proinflammatory mediators is probable. Second, mast cells from RA but not OA synovium express the receptor for the anaphylatoxin C5a, a mediator readily documented in synovial fluid. Immune complexes within RA joints, and potentially IgE antibodies against citrullinated peptides, provide other candidate pathways to activation of synovial mast cells. Indeed, ultrastructural data support ongoing degranulation of mast cells in the RA synovium. Finally, studies of c-kit inhibition in murine and human arthritis suggest efficacy, although it remains unclear whether these agents functionally antagonize tissue mast cells, and whether such antagonism explains their efficacy.

The effects of mast cells on the established inflammatory synovial infiltrate are difficult to predict. As in acute arthritis, activated mast cells may promote the recruitment and activation of leukocytes. Alternatively, protease cleavage of inflammatory mediators and elaboration of cytokines such as IL-10 and TGF-β could downmodulate inflammation, potentially in concert with regulatory T cells, as has been observed in tolerance of skin grafts in mice.

Mast cells likely modulate the stromal response to inflammation as well. Expansion and activation of synovial fibroblasts is a key pathogenic process within RA, and the capacity of mast cells to promote such changes is well established. Through interaction with osteoclasts, mast cells could promote both focal erosions and periarticular osteopenia. Mast cell tryptase not only promotes inflammation but can contribute directly to joint injury in an amplifying manner by activating synovial fibroblasts to produce chemokines, or by acting directly on susceptible substrates such as cartilage aggregate. Finally, by producing proangiogenic mediators, mast cells may enable growth of the vascular supply required for profound expansion of the thin synovial layer into thick pannus. Confirmation of these roles awaits further experimental data.

**CONCLUSIONS**

Mast cells are potent immune cells characterized by phenotypic diversity and an extremely broad range of functions in health and disease. In addition to mediating atopic disease, mast cells serve as important sentinels against pathogen invasion. Under certain conditions, it is likely that they also participate in control of the immune response and remodeling of tissue matrix. Abrupt activation of mast cells by autoantibodies and potentially other signals has been identified in a range of inflammatory diseases, including arthritis. Such activation may represent a key pathologic step in the development of tissue inflammation and injury, and could present an interesting target for the development of anti-inflammatory therapies.
Figure 16-5  Potential roles of mast cells in acute and chronic arthritis. In the acute phase of joint inflammation, mast cells may contribute to initiation of arthritis by inducing vascular permeability, recruiting and activating circulating leukocytes, and stimulating local fibroblasts and macrophages. In established arthritis, these activities may be joined by effects on the stroma, including promotion of pannus formation, angiogenesis, fibrosis, and injury to cartilage and bone. Potential anti-inflammatory effects of mast cells are not depicted. The mediators listed are representative and do not constitute a complete list. bFGF, basic fibroblast growth factor; IL, interleukin; LTC4, leukotriene C4; MIP, macrophage inflammatory protein; MMP, matrix metalloproteinase; PDGF, platelet-derived growth factor; PMN, polymorphonuclear neutrophil; TNF, tumor necrosis factor; VEGF, vascular endothelial growth factor. (Reproduced with permission from Nigrovic PA, Lee DM: Synovial mast cells: role in acute and chronic arthritis, Immunol Rev 217:19–37, 2007. Illustration by Steven Moskowitz.)

Selected References

50. Zhang Y, Ramos BF, Jakschik BA: Neutrophil recruitment by tumor

46. Oettgen HC, Martin TR, Wynshaw-Boris A, et al: Active anaphy-

44. Dvorak AM, Schleimer RP, Schulman ES, Lichtenstein LM: Human


37. Deleted in proofs.


35. Levi-Schaffer F, Austen KF, Gravallese PM, Stevens RL: Coculture


32. Gotis-Graham I, McNeil HP: Mast cell responses in rheumatoid


mucosa of patients with defective T lymphocyte function. J Immunol


counterbalance one another while regulating mast cell homeostasis,


32. Cotis-Graham I, McNeil HP: Mast cell responses in rheumatoid

synovium: association of the MCTC subset with matrix turnover and


suppress mast cell degranulation and allergic responses through


cell-mast cell interactions alter IgE receptor expression and signaling,


35. Levi-Schaffer F, Austen KE, Gravallese PM, Stevens RL: Coculture

of interleukin 3-dependent mouse mast cells with fibroblasts results

in a phenotypic change of the mast cells, Proc Natl Acad Sci U S A


the expression of granule accumulation of tryptase via interleukin-3


37. Deleted in proofs.

38. Mahood LF, Ochs HD: Mast cells in transplantation, Transplantation


signaling pathways in mast cells that lead to cytokine production and


42. Dvorak AM, Schleimer RP, Schulman ES, Lichtenstein LM: Human

mast cells use conservation and condensation mechanisms during

recovery from degranulation: in vitro studies with mast cells purified


43. Oettgen HC, Martin TR, Wynshaw-Boris A, et al: Active anaphyl-


44. Miyajima I, Dombrowicz D, Martin TR, et al: Systemic anaphylaxis

in the mouse can be mediated largely through IgG1 and Fc gammaRIII:

assessment of the cardiopulmonary changes, mast cell degranulation,

and death associated with active or IgE- or IgG1-dependent

definition of mast cells: endogenous proteases are responsible for


capable of serotonin synthesis and release, J Allergy Clin Immunol


46. Humphries DE, Wong GW, Friend DS, et al: Heparin is essential for the

storage of specific granule proteases in mast cells, Nature 400:


stimulates migration of capillary endothelial cells in vitro, J Exp Med


49. Gordon JR, Galli SJ: Mast cells as a source of both preformed and


influx and bacterial clearance at sites of infection through TNF-


I. Ultrastructural in situ and in vitro immunologic characterization,


52. Schwartz LB, Bradford TR: Regulation of tryptase from human lung

mast cells by heparin: stabilization of the active tetramer, J Biol Chem


Full references for this chapter can be found on www.expertconsult.com.
References

1. Ehrlich P: Beiträge zur Theorie und Praxis der Histologischen Färbung (thesis), Saxony, Germany, 1878, Leipzig University.
34. Deleted in proofs.
44. Miyajima I, Dombrowicz D, Martin TR, et al: Systemic anaphylaxis in the mouse can be mediated largely through IgG1 and Fc gammaRII: assessment of the cardiopulmonary changes, mast cell
degradation, and death associated with active or IgE- or IgG1-dependent passive anaphylaxis, J Clin Immunol 99:901–914, 1997.


128. Heissig B, Ral1 S, Akiyama H, et al: Low-dose irradiation promotes tissue revascularization through VEGF release from mast cells and...


Platelets are small circulating cytoplasmic fragments that play a crucial role in hemostasis. They are produced in the bone marrow by megakaryocytes. Single platelets circulate freely in the bloodstream; after vascular injury, platelets adhere to the subendothelium, resulting in responses that contribute to formation of the hemostatic plug. These responses include aggregation, secretion of bioactive compounds, and production of procoagulant activity. Platelets also secrete soluble factors that contribute to wound repair by altering vascular tone and permeability, promoting cell growth, and stimulating scavenger cells such as monocytes. During the inflammatory response, many of the activities that lead to hemostasis contribute to inflammation. Inflammation initiates clotting, decreasing the activity of natural anticoagulant mechanisms and impairing the fibrinolytic system. Conversely, activated platelets release chemotactic factors that promote leukocyte adhesion, which facilitates their extravasation into inflammatory foci. Platelets secrete a variety of factors that can alter vascular tone and permeability. Last, platelets are a major source of transforming growth factor (TGF)-β, a potent stimulus of fibrosis. Taken together, these activities make platelets contributors to the inflammatory response and to the pathogenesis of systemic rheumatic diseases.

Platelets have been implicated in the pathogenesis of several rheumatic diseases, including rheumatoid arthritis and systemic lupus erythematosus; in the latter case, they have been particularly implicated in atherothrombotic complications. Platelet surface proteins also participate in inflammation by serving as sites for leukocyte adhesion (e.g., P-selectin, glycoprotein Ibα (GPIbα)) or as agonists for counterreceptors on leukocytes (e.g., CD40 ligand, platelet-activating factor). Platelet surface proteins also participate in inflammation by serving as sites for leukocyte adhesion (e.g., P-selectin, glycoprotein Ibα (GPIbα)) or as agonists for counterreceptors on leukocytes (e.g., CD40 ligand, platelet-activating factor). Platelet surface proteins also participate in inflammation by serving as sites for leukocyte adhesion (e.g., P-selectin, glycoprotein Ibα (GPIbα)) or as agonists for counterreceptors on leukocytes (e.g., CD40 ligand, platelet-activating factor).

The development of antiplatelet agents offers the promise of new therapeutic modalities. 

Platelets are the smallest blood cells; they are cytoplasmic fragments derived from their bone marrow precursor, the megakaryocyte. Resting platelets have a smooth disk shape and are 3.6 ± 0.7 µm in diameter. On activation, platelets undergo a shape change, becoming a compact sphere with numerous long dendritic extensions, markedly increasing their surface area. In humans, normal platelet counts range from 150,000/µL to 450,000/µL. The main function of platelets is to maintain vascular integrity, thereby playing a crucial role in hemostasis.

The plasma membrane of platelets is a typical lipid bilayer, having an extensive series of complex invaginations termed the canicular system. The role of this surface-connected tubular system seems to be to facilitate the quick release of secreted substances to the extracellular environment. The platelet membrane bears numerous glycoprotein (GP) receptors. Platelet surface phospholipids play an important role in coagulation and are a source of arachidonic acid, a precursor of important vasoactive substances such as thromboxane A₂, a potent vasoconstrictor and platelet-aggregating agent, and of leukotrienes, which can amplify the inflammatory response. Platelet surface GPs are receptors that mediate adhesion to subendothelial tissue and subsequent aggregation to form the hemostatic plug. The largest GP is termed I and the smallest IX. The labels α and β distinguish between two separate electrophoretic bands that initially were considered one (e.g., GPI became GPIa and GPIb). The platelet GPIb-IX-V is an important receptor that binds to von Willebrand factor (vWF) exposed in the subendothelial matrix, causing the attachment of platelets. Deficiency of any component of the GPIb-IX-V complex or of vWF leads to the congenital bleeding disorder Bernard-Soulier disease (GPIb-IX-V complex) or von Willebrand disease (vWD), respectively. vWF, in addition to its important role in hemostasis, has been suggested to promote inflammation, facilitating neutrophil diapedesis by destabilization of the endothelial barrier.

ADAMTS-13 (a disintegrin-like metalloprotease with thrombospondin type I repeats 13) is a plasma protease that cleaves vWF into smaller multimers, reducing its hemostatic potency. Mutations in the ADAMTS-13 gene and autoantibodies against ADAMTS-13 have been shown to cause familial and acquired thrombotic thrombocytopenic purpura, respectively. In mice, ADAMTS-13 has powerful natural antithrombotic activity, and recombinant ADAMTS-13 has proved useful in preventing ischemic brain injury in experimental stroke. It has been suggested that ADAMTS-13 might act as a link between thrombosis and inflammation. In inflammatory models, ADAMTS-13...
has played an important role in preventing vWF-induced secretion of Weibel-Palade bodies by endothelial cells and, consequently, reducing the adhesion and extravasation of leukocytes.\(^4\) Other interactions that contribute to initial platelet adhesion are mediated by collagen receptors GPIIb-IIIa (integrin \(\alpha_{IIb}\beta_3\)) and GPVI, which bind to collagen in the subendothelial matrix.\(^14\) The most abundant platelet surface receptor, GPIIb-IIIa (integrin \(\alpha_{IIb}\beta_3\)), is activated by adhesion to collagen or vWF or by soluble agonists, such as thrombin. After activation, GPIIb-IIIa binds fibrinogen, leading to platelet aggregation.\(^4\) Deficiency of this GP results in Glanzmann's thrombasthenia, a disorder characterized by petechial bleeding and the absence of platelet aggregation and clot retraction.\(^15\)

The cytoplasm of platelets is rich in actin and myosin, which provide platelets the ability to change shape and to retract clots. Platelet cytoplasm consists of mitochondria, lysosomes, glycogen stores, and three types of granules that contain numerous biologically active molecules (Table 17-1). These granules are classified according to their ultrastructure, density, and contents as alpha granules, lysosomes, and dense granules. Although most of the contents of these granules are made in megakaryocytes, some are taken up from the plasma by megakaryocytes and platelets.

Alpha granules contain numerous proteins and growth factors, such as platelet-derived growth factor (PDGF), TGF-\(\beta\), platelet factor-4 (also referred to as CXCL4), and vWF, which are synthesized in the megakaryocyte.\(^16\) Other proteins, such as fibrinogen, enter the alpha granules from the plasma via GPIIb-IIIa receptor-mediated endocytosis.\(^17\) P-selectin (CD62P), an adhesion molecule, also is localized in the membrane of alpha granules\(^18\) and redistributes to the cell surface during platelet activation. Platelet P-selectin has been implicated in stabilizing platelet aggregates.\(^19\)

The best documented high-affinity counterreceptor for P-selectin is P-selectin glycoprotein ligand-1 (PSGL-1), a transmembrane sialomucin found on leukocytes and lymphoid cells,\(^20\) through whose interaction platelets participate in the inflammatory response.\(^21\)

Dense granules contain serotonin, adenosine diphosphate (ADP), adenosine triphosphate, and calcium. The dense granule membrane bodies are made in megakaryocytes, but they do not acquire their content of serotonin and calcium until platelets are released into the circulation.\(^22\) Another series of intracellular membrane vesicles serves as a reserve to increase membrane surface area on platelet activation.

As stated previously, platelets are small cytoplasmic fragments derived from megakaryocytes. Although megakaryocytes are rare in the bone marrow (approximately 0.1% of all nucleated cells), they are easily recognized by their giant size (50 to 100 \(\mu\)m diameter) and large, multilobed nucleus. Megakaryocytes have two unique characteristics: (1) They undergo a process known as endomitosis, in which the nucleus accumulates many times the normal number of chromosomes, and (2) they have specialized structures in the cytoplasm that permit fragments to be shed, as platelets, into the bloodstream.\(^23\)

With a life span of just about 10 days, every day, about \(2 \times 10^{11}\) platelets are released into the bloodstream of healthy adults by mature megakaryocytes. This quantity can be increased 10-fold under specific conditions. In humans, as in other species, there is an inverse relationship between platelet count and mean platelet volume.\(^24\) This suggests that platelet production by bone marrow megakaryocytes is regulated to maintain a constant total platelet mass. The tendency toward a stable platelet mass explains the wide variation in platelet count in healthy donors (150,000/\(\mu\)L to 450,000/\(\mu\)L). Megakaryocytes normally replace about 10% of the platelet mass daily. In response to the increased need for platelets, megakaryocytes modify their number, size, and ploidy. Changes in free levels of thrombopoietin, the main physiologic regulator of platelet production, are responsible for these morphologic and functional adaptations in megakaryocytes.

Thrombopoietin is an 80- to 90-kD GP produced mainly by the liver and released at a constant rate into the circulation. Thrombopoietin acts through its receptor, also known as e-Mpl, which is present in platelets, megakaryocytes, and, to a lesser extent, most other hematopoietic precursor cells. Thrombopoietin prevents apoptosis of megakaryocytes, while increasing their number, size, and maturation,\(^25\) but it does not seem to increase the rate of shedding of platelets into the circulation.\(^26\) On circulating platelets, thrombopoietin is not a sufficiently strong stimulus to trigger platelet function, but it reduces the threshold for activation by other agonists, such as ADP.\(^27\) Binding to the platelet thrombopoietin receptor is the major route of catabolism, however, of circulating thrombopoietin. When the platelet production rate decreases, the platelet mass and the quantity of thrombopoietin receptor decrease; consequently, thrombopoietin concentrations increase and megakaryocyte growth is stimulated. In conditions of high platelet mass (e.g., hypertransfusion of platelets), the number of

---

Table 17-1 Platelet Granule Compounds and Granule Membrane Components with a Role in the Hemostatic/Inflammatory Response

<table>
<thead>
<tr>
<th>Platelet Granules</th>
<th>Actions</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dense granule</td>
<td>Proaggregating factors</td>
<td>Serotonin, histamine, ADP, ATP, Ca(^{2+}), Mg(^{2+})</td>
</tr>
<tr>
<td>Alpha granule</td>
<td>Adhesive glycoproteins, Growth factors, Platelet aggregation and chemotaxis, Hemostasis factors, Tissue destruction</td>
<td>P-selectin, CD31, GPIIb-IIIa, fibronectin, vitronectin, thrombospondin TGF-(\beta), PDGF, EGF, VEGF, (\beta)-Thromboglobulin, PF4 (CXCL4), CC and CXC chemokines, Fibrinogen, vWF</td>
</tr>
<tr>
<td>Lysosome</td>
<td>Hydrolases, collagenase, cathepsins D and E</td>
<td></td>
</tr>
</tbody>
</table>

ADP, adenosine diphosphate; ATP, adenosine triphosphate; EGF, epidermal growth factor; GPIIb-IIIa, glycoprotein IIb-IIIa; PDGF, platelet-derived growth factor; PF4, platelet factor-4; TGF, transforming growth factor; VEGF, vascular endothelial growth factor; vWF, von Willebrand factor.

thrombopoietin receptors increases, thrombopoietin concentrations decrease, and megakaryocyte growth decreases. In addition to thrombopoietin, other soluble factors, such as interleukin (IL)-3, IL-6, IL-11, stem cell factor, or granulocyte-macrophage colony-stimulating factor, seem to promote megakaryocyte growth and maturation. Some of these soluble proteins may play a relevant role in thrombocytosis conditions.28

Under normal conditions, the spleen stores about one-third of circulating platelets. Circumstances that increase splenic volume, such as hepatic cirrhosis or portal hypertension, cause a reduction in the circulating platelet count by a sequestration within the splenic sinuses. Hypersplenism does not reduce platelet life span, however; rather, it reduces the circulating platelets available for effective hemostasis. After senescence, platelets are removed from the circulation by the reticuloendothelial system. Only a small fraction of the circulating platelets available for effective hemostasis. Does not reduce platelet life span, however; rather, it reduces the number of circulating platelets available for effective hemostasis. After senescence, platelets are removed from the circulation by the reticuloendothelial system. Only a small fraction of the circulating platelets available for effective hemostasis.

FUNCTION OF PLATELETS

In response to vascular injury, platelets adhere to subendothelium, secreting a variety of potent agonists and aggregating to form a hemostatic plug. During the inflammatory response, these physiologic responses of platelets can promote and exacerbate inflammation. In this sense, platelets are authentic inflammatory cells.

HEMOSTASIS

When a blood vessel is injured, a complex process involving biochemical reactions and cell-cell and cell-matrix interactions, termed hemostasis, occurs. The initial hemostatic response is mediated by platelets that form the platelet plug (Figure 17-1).

Under physiologic conditions, the undamaged endothelium prevents the adherence of platelets by several mechanisms. These mechanisms include a cell-associated ecto-ADPase (CD39) and the production of nitric oxide and prostacyclin.29 When blood vessel integrity is disrupted, the first reaction is vasoconstriction, which reduces blood loss. Simultaneously, subendothelial matrix elements are exposed, and platelets are rapidly transformed into sticky cellular elements capable of adhering to the underlying surface. Platelet adhesion is initially mediated by the interaction of the GPIb-IX-V receptor complex with vWF in the subendothelial matrix.30 This interaction transduces signals through the GPIb-IX-V complex that activate platelet integrins.30 The activation of GPIa-IIa and GPIb-IIIa integrins allows the binding to collagen (GPIa-IIa) and vWF (GPIb-IIIa), mediating the stable adhesion of platelets to the subendothelial surface. In addition to vWF, the active form of GPIb-IIIa binds fibrinogen.31 The association of soluble fibrinogen with GPIb-IIIa creates bridges between platelets that result in platelet aggregation and thrombus growth. In concert with aggregation, platelets release their intracellular granules, amplifying the hemostatic response (see Table 17-1).32,33 The outcome is the formation of a platelet plug and triggering of the coagulation cascade, which leads to thrombin generation and resulting fibrin clot formation (Figure 17-2).

One response of platelets to activation by stimuli such as shear stress or collagen is the release of vesicles called platelet microparticles, fragments 0.1 to 0.2 µm in diameter that carry antigens present in intact platelets. These platelet-derived microparticles may play a role in normal hemostasis.34-35 The number of clinical disorders associated with elevated platelet microparticles is increasing,36-37 including several rheumatic diseases in which the number of circulating platelet microparticles seems to be associated with disease activity.38-41 The relevance of platelet-derived microparticles in the physiopathology of those disorders needs to be fully clarified; however, it has been demonstrated that platelets intensify the inflammatory response in joint42 (Figure 17-3).

Under physiologic conditions, the undamaged endothelium prevents the adherence of platelets by several mechanisms. These mechanisms include a cell-associated ecto-ADPase (CD39) and the production of nitric oxide and prostacyclin. When blood vessel integrity is disrupted, the first reaction is vasoconstriction, which reduces blood loss. Simultaneously, subendothelial matrix elements are exposed, and platelets are rapidly transformed into sticky cellular elements capable of adhering to the underlying surface. Platelet adhesion is initially mediated by the interaction of the GPIb-IX-V receptor complex with vWF in the subendothelial matrix. This interaction transduces signals through the GPIb-IX-V complex that activate platelet integrins. The activation of GPIa-IIa and GPIb-IIIa integrins allows the binding to collagen (GPIa-IIa) and vWF (GPIb-IIIa), mediating the stable adhesion of platelets to the subendothelial surface. In addition to vWF, the active form of GPIb-IIIa binds fibrinogen. The association of soluble fibrinogen with GPIb-IIIa creates bridges between platelets that result in platelet aggregation and thrombus growth. In concert with aggregation, platelets release their intracellular granules, amplifying the hemostatic response (see Table 17-1).

The outcome is the formation of a platelet plug and triggering of the coagulation cascade, which leads to thrombin generation and resulting fibrin clot formation (Figure 17-2).

Figure 17-2 Anatomy of a platelet plug. Electron micrograph of a group of platelets (P) attached to an endothelial cell (EC) in the initial platelet plug formation. Several dense granules (d) and alpha granules (α) are visible. The central platelet shows long dendritic extensions or filopodia (F). (Courtesy Dr. Lucio Díaz-Flores.)

Figure 17-1 Platelet plug formation. Platelet activation can be initiated by several mechanical (vessel wall injury, disruption of atherosclerotic plaques) or chemical (adenosine diphosphate, thromboxane A2, and thrombin) stimuli. In response to vessel wall injury, platelets attach to subendothelial matrix (adhesion); this is followed by fibrinogen-mediated platelet-platelet interaction (aggregation). Simultaneously, platelets release their intracellular granule contents (secretion), leading to recruitment of additional circulating platelets (amplification).
**Table 17-2** Platelet Components Implicated in the Inflammatory Response

<table>
<thead>
<tr>
<th>Platelet Component</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface molecules</td>
<td>P-selectin (CD62P), PECAM (CD31), GPIbα</td>
</tr>
<tr>
<td>Soluble factors</td>
<td>PAF, ROS, CD154 (CD40 ligand), Serotonin, histamine, β-Thromboglobulin, PF4, Acid hydrolases, ROS, PDGF, TGF-β, Thrombin, fibrin</td>
</tr>
<tr>
<td>End products of platelet procoagulant activity</td>
<td>Tissue destruction</td>
</tr>
</tbody>
</table>

GPIb-IIIa is a member of a family of cell-adhesion receptors termed integrins. It also is referred to as integrin αIIbβ3 or CD41/CD61. Although integrins are expressed on virtually all nucleated cells, GPIb-IIIa is restricted to megakaryocytes and platelets. It is the most abundant receptor on the platelet surface, averaging 80,000 copies per platelet. GPIb-IIIa recognizes at least five different adhesive ligands:

- **fibronectin**, fibrinogen, vWF, thrombospondin, and vitronectin.

Cells can modify integrin functions through dynamic modulation of receptor affinity. On resting platelets, GPIb-IIIa does not bind soluble fibrinogen. After platelet stimulation (e.g., by thrombin, collagen, or ADP), GPIb-IIIa undergoes a conformational change, however, and is converted from a low-affinity to a high-affinity fibrinogen receptor, a process known as **inside-out signaling**. In this situation, fibrinogen bridges the activated platelets, and platelet aggregation occurs. Simultaneously, the cytosolic portion of the activated GPIb-IIIa binds to platelet cytoskeleton proteins and mediates platelet spreading and clot retraction in what is referred to as **outside-in integrin signaling**. GPIb-IIIa integrates receptor-ligand interactions on the external face of the membrane with cytosolic events in a bidirectional fashion. This is the final common pathway for platelet aggregation, regardless of the mode of platelet stimulation. The importance of GPIb-IIIa integrin is illustrated by Glanzmann’s thrombasthenia, a bleeding disorder caused by mutations in the gene for the αIIb- or the β3-subunit, and by the clinical utility of GPIb-IIIa antagonists as antithrombotic agents in the treatment of thrombotic diseases. Glycoprotein IIb-IIIa inhibitors are now recommended by international guidelines in patients with acute coronary syndromes undergoing percutaneous coronary intervention.

**ROLE OF PLATELETS IN THE INFLAMMATORY RESPONSE**

The accumulation of leukocytes in tissue is an essential event for the inflammatory response. The current paradigm of leukocyte extravasation requires a multistep cascade of sequential leukocyte–endothelial cell interactions, in which members of three different families of adhesion receptors participate: selectins, integrins, and the immunoglobulin superfamily. Platelets contribute in many ways to leukocyte accumulation in the inflammatory foci (Table 17-2).

In flowing blood, leukocytes roll on adherent activated platelets, mainly through the interaction of platelet P-selectin with its major leukocyte ligand, PSGL-1. This initial rolling of leukocytes on platelet P-selectin is followed by their firm adhesion and subsequent migration—processes that depend on the leukocyte integrin Mac-1 (αMβ2, CD11b/CD18). Mac-1 adheres firmly to platelets through direct binding to glycoprotein Ibα (GPIbα, CD42b). These interactions provide molecular mechanisms for leukocyte recruitment to hemostatic plugs where platelets have been previously deposited in response to vascular injury. Parallel lines of investigation have shown that resting platelets are able to roll on activated endothelial cells, apparently through an interaction between PSGL-1 expressed in platelets and the endothelial P-selectin.
physiologic function of platelet rolling on stimulated endothelial cells needs to be clarified. If this contact results in activation of platelets, however, those platelets may release proinflammatory mediators, such as cytokines, chemokines, and eicosanoid precursors, or growth factors that stimulate tissue healing. Activated platelets in circulation stimulate secretion of Weibel-Palade bodies from endothelial cells in vivo; this leads to P-selectin–mediated leukocyte rolling. Given the important role of platelet P-selectin in chronic inflammatory processes, this effect of activated platelets might represent an important pathway of platelet-induced inflammation.

In addition to the adhesion molecules, activated platelets express on their surface two major proinflammatory mediators: platelet-activating factor (PAF) and CD40 ligand (CD154). PAF is a potent platelet-aggregating phospholipid produced by macrophages, mast cells, platelets, endothelial cells, neutrophils, and monocytes. Upon cell activation, PAF is rapidly synthesized and translocated to the plasma membrane of endothelial cells, where it recognizes its receptor in neutrophils, resulting in β2 integrin–mediated adhesion of leukocytes to the endothelial surface.

In the same way, PAF can signal neutrophils when it is displayed on the surface of adherent activated platelets acting in cooperation with P-selectin to tether neutrophils. The biologic action of PAF is physiologically inactive by plasma and cellular acetylhydrolase. A role of PAF in the pathogenesis of chronic inflammatory arthritis has been proposed; however, a well-controlled clinical trial failed to show any beneficial effect of a PAF antagonist in patients with active rheumatoid arthritis (RA).

CD40 is a transmembrane protein member of the tumor necrosis factor (TNF) receptor family. CD40 is present on many cells, including B cells, monocytes, macrophages, dendritic cells, and vascular endothelial cells. Platelets are the major peripheral blood source of CD154, the ligand of CD40, and they express it on their surface within seconds of exposure to an agonist. The interaction of CD154 on activated platelets with CD40 on endothelial cells causes a proinflammatory reaction of the endothelium characterized by expression of inflammatory adhesion molecules, such as E-selectin, vascular cell adhesion molecule-1 (CD106), and intercellular adhesion molecule-1 (CD54), and secretion of the chemokines IL-8 (CXCL8) and monocyte chemotactic protein-1 (CCL2). CD154 expressed on activated platelets can provide a potent stimulus to the inflammatory response. Clinical data from an open-label study suggested that the blockade of CD154 with a biologic may induce a prothrombotic state in patients with lupus nephritis through a mechanism not clarified. A phase II, double-blind, placebo-controlled study evaluating the safety and efficacy of a humanized monoclonal antibody against CD154 in patients with active systemic lupus erythematosus failed to show clinical efficacy. In this study, the type and frequency of adverse events were similar between the intervention and placebo groups.

When platelets adhere, they release numerous growth factors, such as PDGF, TGF-β, and other factors that are chemotactic for monocytes, macrophages, and fibroblasts. These growth factors may play an important role in the chronic inflammatory response by mediating a fibroproliferative response. PDGF is a homodimer or heterodimer molecule of A and B chains produced by platelets, monocytes or macrophages, endothelial cells, and vascular smooth muscle cells (under some conditions). This molecule plays an essential role in tissue repair and wound healing. PDGF is a potent mitogen and chemoattractant for smooth muscle cells, connective tissue cells, and macrophages; it contributes to the formation of lesions of atherosclerosis, a disorder strongly related to the inflammatory response. It has been shown that PDGF is a potent mitogen for synovial fibroblasts isolated from patients with RA.

TGF-β has three isoforms (TGF-β1, TGF-β2, and TGF-β3) secreted by virtually all cell types as latent complexes that need to be processed to exhibit biologic activity. Several effects have been associated with TGF-β: (1) It is chemotactic for various cell types, including leukocytes; (2) it inhibits proliferation of most cells; (3) it induces the synthesis and deposition of extracellular matrix; and (4) it stimulates the formation of granulation tissue. The net result is that TGF-β is mainly an inhibitor of the inflammatory response. Carefully regulated expression of active TGF-β is essential for resolution of inflammation and repair. Systemic administration of TGF-β1 has antagonized the development of polyarthritis in susceptible rats. Overproduction of this cytokine has been associated with several fibrotic processes. TGF-β is a major cytokine involved in the pathogenesis of fibrosis in systemic sclerosis. Blockade of cell surface molecules capable of activating latent TGF-β and blockade of ligand by antibody, soluble TGF-β receptors, and a recombinant latency-associated peptide, as well as inhibitors for ALK5 and Smad3, are potential strategies for abolishing the pathologic activation of TGF-β in systemic sclerosis.

Several reactive oxygen species are released from unstimulated platelets and after platelet stimulation with agonists such as collagen or thrombin. Because reactive oxygen species have been implicated in direct tissue injury and in inflammatory reactions through promotion of adhesive interactions between inflammatory and endothelial cells, reactive oxygen species originating from platelets may act as an autocrine or paracrine mediator that participates in the amplification of the inflammatory response in disorders such as rheumatic diseases.

**PLATELETS AND RHEUMATIC DISEASES**

**Alterations in Platelet Numbers in Rheumatic Diseases**

Increases in platelet counts have three major causes: (1) reactive or secondary thrombocytosis; (2) familial thrombocytosis; and (3) clonal thrombocytosis, including essential thrombocythemia and related myeloproliferative disorders. The platelet count is frequently elevated in patients with active RA and juvenile chronic arthritis, owing to reactive thrombocytosis. The level of thrombocytosis correlates with clinical and laboratory parameters of disease activity. Relapses of RA are often accompanied by increases in platelet count, whereas remissions are associated with their reduction, to normal limits. This activity indicates that
the thrombocytosis observed in patients with rheumatic disease is reactive or occurs secondary to the chronic inflammatory process. Although the mechanism responsible for thrombocytosis is uncertain, increased intravascular coagulation with a compensatory increase in platelet production has been suggested as a possible cause.\(^\text{63}\) More recently, several studies suggested that inflammatory cytokines with a minor role in the physiologic production of platelets, such as IL-6, IL-1, or TNF, among others,\(^\text{84,85}\) may be active mediators in the regulation of thrombopoiesis during the reactive thrombocytosis that occurs in the inflammatory process.

Reduced platelet count, or thrombocytopenia, is common in rheumatic diseases. The mechanisms involved in thrombocytopenic states include reduction in platelet production, sequestration, and rapid platelet destruction. Several drugs used in rheumatic diseases are able to suppress the bone marrow. Among drugs that can produce thrombocytopenia because of megakaryocytic hypoplasia are gold, cyclophosphamide, methotrexate, penicillamine, and azathioprine. The effect these compounds have on suppressing megakaryocyte replication depends on the time and dose of exposure; reduced elimination of these drugs places patients at increased risk for this complication.\(^\text{86}\)

The normal spleen contains about 30% of the platelet mass, and splenomegaly can result in a low circulating count without reduction in the platelet life span.\(^\text{97}\) Several rheumatic diseases may lead to this type of thrombocytopenia. The best known is Felty's syndrome, an uncommon but severe subset of seropositive RA complicated by granulocytopenia and splenomegaly. In this disorder, thrombocytopenia usually is not life threatening.

Another related disease is immune-mediated platelet destruction,\(^\text{88}\) a disorder termed idiopathic thrombocytopenic purpura. Autoantibodies cause idiopathic thrombocytopenic purpura, and platelet surface proteins, including GPIb-IIIa, GP Ib-IX, GPIa-IIa, GPV, and GPVI, can be antigenic targets of such autoantibodies.\(^\text{89,90}\) Circulating platelets coated with immunoglobulin (Ig)G autoantibodies undergo accelerated clearance through Fcγ receptors expressed by macrophages in the spleen and liver. In some cases of idiopathic thrombocytopenic purpura, platelet production seems to be reduced, either by intramedullary destruction of antibody-coated platelets or by inhibition of megakaryocyte proliferation.\(^\text{91}\) The level of thrombopoietin is not increased,\(^\text{92}\) suggesting a normal megakaryocyte mass. Idiopathic thrombocytopenic purpura is present in 15% to 25% of patients with systemic lupus erythematosus\(^\text{93}\) and in about 25% of patients with antiphospholipid syndrome.\(^\text{94}\)

In contrast, the thrombocytopenia that occurs during episodes of systemic vasculitis has a more complex pathogenesis, a worse clinical course, and a poorer outcome.\(^\text{95,96}\) Immune thrombocytopenia is rare in RA except when related to therapy. Among the drugs that can produce thrombocytopenia in RA, intramuscular gold salts are the most clearly associated with drug-induced immune thrombocytopenia. About 1% to 3% of patients receiving intramuscular gold salts for the treatment of RA develop a thrombocytopenia, which may be life-threatening. Although, as stated previously, bone marrow suppression can occur in patients undergoing gold treatment, thrombocytopenia is usually due to immune destruction of platelets associated with an active marrow.\(^\text{97}\)

### Role of Platelets in the Pathogenesis of Rheumatic Diseases

The role that platelets play in amplification of the inflammatory response provides a basis for their involvement in rheumatic diseases. Most of the available evidence implicating platelets in the pathogenesis of rheumatic disorders is indirect and circumstantial; however, current findings based on pharmacologic and genetic experimental procedures demonstrate a previously unappreciated role for platelet microparticles in the pathogenesis of rheumatic diseases.\(^\text{42}\)

Platelets have been implicated in the pathogenesis of RA,\(^\text{288}\) on the basis of several studies that have documented the presence of platelet proteins in the synovial fluid of RA patients\(^\text{99}\) and on the observation that labeled platelets localize only to joints with clinically active inflammation.\(^\text{100}\) Levels of plasma-soluble P-selectin are increased in RA patients compared with controls,\(^\text{101}\) indicating platelet activation in this disease. A direct correlation has been observed between platelet-derived microparticle levels and disease activity in RA patients; this suggests that generation of platelet microparticles\(^\text{102,103}\) contributes to the pathogenesis of RA. Microparticles are abundant in RA, psoriatic arthritis, and juvenile idiopathic arthritis synovial fluid both in suspension and adhered to the surface of leukocytes. Solid evidence supports that platelet microparticles are generated by platelet activation via the collagen receptor GPVI, locally in the synovial tissue. Once in the joint, microparticles seem to play a proinflammatory role, exerting potent IL-1–mediated activation of resident synoviocytes\(^\text{42}\) (Figure 17-4). These findings support platelet GPVI as a potential target for arthritis treatment.

Several studies have focused on the presence of activated platelets in patients with systemic lupus erythematosus.\(^\text{102-104}\) The risk for thrombosis is increased significantly in these patients, and platelets have been implicated in the prothrombotic state of systemic lupus erythematosus through the release of microparticles\(^\text{105}\) and by the increased deposition of complement activation product C4d on the platelet surface.\(^\text{105,106}\) Patients with essential thrombocytopenia have an increased prevalence of antiphospholipid antibodies, which may be associated with a higher risk of thrombosis.\(^\text{107}\) The presence of activated platelets and enhanced aggregation of platelets have been described in patients with antiphospholipid syndrome,\(^\text{103}\) systemic sclerosis,\(^\text{108,109}\) primary Raynaud’s phenomenon\(^\text{108}\) and ankylosing spondylitis.\(^\text{110}\)

### INHIBITION OF PLATELET FUNCTION BY PHARMACOLOGIC AGENTS

Nonsteroidal anti-inflammatory drugs (NSAIDs) serve as the foundation of therapy in many rheumatic diseases. NSAIDs inhibit prostaglandin synthesis\(^\text{111}\) through blockade of cyclooxygenase (COX). These agents can interfere with platelet aggregation and secretion\(^\text{112,113}\) through the inactivation of platelet COX-1. In platelets, this enzyme is a rate-limiting step in the transformation of arachidonic acid into thromboxane A₂, a potent platelet-aggregating agent. In addition, some NSAIDs reduce platelet
aggregation by interfering with the activation of GPIIb-IIIa through a COX-independent mechanism. NSAIDs inhibit platelet function and can lead to bleeding complications in patients with rheumatic diseases. Newly developed potent antithrombotic agents also might provide new weapons in the treatment of rheumatic diseases. Among these are ticlopidine and its analogue, clopidogrel—two inhibitors of the P2Y₁₂ ADP receptor. Nowadays, many other members of this family of antiplatelet agents are being tested for control of procoagulant states. These agents have greater efficacy than aspirin for the prevention of recurrent stroke and may find a place in the antirheumatic armamentarium as a result of their potential anti-inflammatory properties. However, no information is available about clinical trials specifically designed to test the effect of P2Y₁₂ inhibitors in the control of rheumatic diseases. Agents that interfere directly with the adhesive function of integrin GPIIb-IIIa have come into therapeutic use. This new group of agents includes monoclonal antibodies (abciximab), cyclic peptides (eptifibatide), and other small molecules that have been approved for intravenous coronary angioplasty and stent procedures. Orally active GPIIb-IIIa blockers have been developed for long-term therapy, including secondary and even primary prevention of thrombotic diseases. However, available data from clinical trials of these oral agents have failed to show clinical benefit, whereas they have shown unexplained increased mortality.

Acknowledgments
This work was supported by grants from the National Institutes of Health (NIH) and from the Instituto de Salud Carlos III of Spain (FIS09/02209). We are indebted to Dra. Esmeralda Delgado-Frias for the artwork.

References

Figure 17-4 Model of how platelet microparticle (MP) generation might contribute to joint inflammation. Circulating platelets make contact with extracellular matrix collagen type IV of fenestrated subsynovial capillaries. This contact generates platelet MPs, which once in the synovial membrane activate resident synoviocytes through potent interleukin (IL)-1 activity.
79. Caccese D, Pratico D, Ghiselli A, et al: Superoxide anion and transforming growth factor-β levels are high when thrombocytopenia is due to megakaryocyte deficiency and low when due to increased platelet destruction, Blood 87:4068–4071, 1996.


The references for this chapter can also be found on [www.expertconsult.com](http://www.expertconsult.com).
Innate Immunity

STEVEN A. PORCELLI

KEY POINTS

- Innate immunity depends on recognition of conserved molecular patterns found in many microorganisms.
- Several families of pattern recognition receptors are responsible for triggering innate immune responses.
- Toll-like receptors and other pattern recognition receptors with leucine-rich repeat domains play a key role in innate immune recognition.
- Antimicrobial peptides are important effectors of innate immunity.
- Phagocytic cells and several types of innate-like lymphocytes are key cell types in mediating innate immunity.
- Innate immune responses have a strong impact on the development of adaptive immunity.
- Some defects in the innate immune system are associated with a predisposition to infection or to autoimmune disease.

It has become common practice in immunology to divide the mechanisms involved in host defense into adaptive and innate components; this provides a useful framework for classifying the numerous cells, receptors, and effector molecules that combine to make up the vertebrate immune system (Table 18-1). A specific immune response, such as the production of antibodies or T cells against a particular pathogen, is referred to as adaptive immunity because it represents an adaptation that occurs during the lifetime of an individual as a result of exposure to that pathogen. Adaptive immune responses involve the clonal expansion of T and B lymphocytes bearing a large repertoire of somatically generated receptors that can be selected to recognize virtually any pathogen. The adaptive immune system of any given individual is profoundly molded by the immunologic challenges encountered by that individual during the course of a lifetime. A hallmark of adaptive immune responses is that they are highly specific for the triggering agent, and they provide the basis for immunologic memory. This property of memory endows the adaptive immune response with its “anticipatory” property, which provides increased resistance against future infection with the same pathogen and also allows vaccination against future infectious threats.

Adaptive immunity is essential for the survival of all mammals and most other vertebrates, but a wide variety of other mechanisms that do not involve antigen-specific lymphocyte responses are also involved in successful immune protection. These diverse mechanisms are collectively known as innate immunity because they are not dependent on prior exposure to specific pathogens for their amplification. Such responses are controlled by the products of germ-line genes that are inherited and similarly expressed by all normal individuals. Innate immune mechanisms involve both constitutive and inducible components and use a wide variety of recognition and effector mechanisms. It has become clear in recent years that innate immune responses have a profound influence on the generation and outcome of adaptive immune responses. This ability of the innate immune system to instruct the responses of the adaptive immune system suggests many ways in which innate immunity can influence the development of both long-term specific immunity and autoimmune disease.

EVOLUTIONARY ORIGINS OF INNATE IMMUNITY

In spite of its obvious importance for most vertebrate organisms, the adaptive immune system is a relatively recent evolutionary development (Figure 18-1). In the great majority of present-day vertebrate species, the adaptive immune system is based on the ability to generate large families of variable lymphocyte receptors with immunoglobulin-like structures. This ability has been conserved owing to the acquisition of a specialized recombination system that mediates the assembly of gene segments in the T cell and B cell receptor families, which most likely occurred through invasion of the genome of a primitive vertebrate by a transposable element or virus carrying this machinery. This critical step in the evolution of the immune system can be traced back to the emergence of the ancestors of present-day jawed fish, which represent the most primitive extant species known to have adaptive immune systems based on the generation of large families of specific immunoglobulin-type receptors. Recently, other systems of variable lymphocyte
**Table 18-1 Contrasting Features of Innate and Adaptive Immune Systems**

<table>
<thead>
<tr>
<th>Property</th>
<th>Innate Immune System</th>
<th>Adaptive Immune System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receptors</td>
<td>Relatively few (several hundred?)</td>
<td>Many (potentially $10^{14}$ or more)</td>
</tr>
<tr>
<td></td>
<td>Fixed in genome</td>
<td>Encoded in gene segments</td>
</tr>
<tr>
<td></td>
<td>Gene rearrangement not required</td>
<td>Gene rearrangement required</td>
</tr>
<tr>
<td></td>
<td>Nonclonal</td>
<td>Clonal</td>
</tr>
<tr>
<td></td>
<td>All cells of a class identical</td>
<td>All cells of a class distinct</td>
</tr>
<tr>
<td>Distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Targets</td>
<td>Conserved molecular patterns</td>
<td>Details of molecular structure</td>
</tr>
<tr>
<td></td>
<td>Lipopolysaccharides</td>
<td>Proteins</td>
</tr>
<tr>
<td></td>
<td>Lipoteichoic acids</td>
<td>Peptides</td>
</tr>
<tr>
<td></td>
<td>Glycans and peptidoglycans</td>
<td>Carbohydrates</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td></td>
</tr>
<tr>
<td>Distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Targets</td>
<td>Conserved molecular patterns</td>
<td>Details of molecular structure</td>
</tr>
<tr>
<td></td>
<td>Lipopolysaccharides</td>
<td>Proteins</td>
</tr>
<tr>
<td></td>
<td>Lipoteichoic acids</td>
<td>Peptides</td>
</tr>
<tr>
<td></td>
<td>Glycans and peptidoglycans</td>
<td>Carbohydrates</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td></td>
</tr>
<tr>
<td>Self–non–self–discrimination</td>
<td>Perfect: selected over evolutionary time</td>
<td>Imperfect: selected in individual somatic cells</td>
</tr>
<tr>
<td>Action time</td>
<td>Immediate or rapid (seconds to hours)</td>
<td>Delayed (days to weeks)</td>
</tr>
<tr>
<td>Response</td>
<td>Microbicidal effector molecules</td>
<td>Clonal expansion or energy of specific T and B lymphocytes</td>
</tr>
<tr>
<td></td>
<td>Antimicrobial peptides</td>
<td>Cytokines (IL-2, IL-4, IFN-γ, others)</td>
</tr>
<tr>
<td></td>
<td>Superoxide</td>
<td>Specific antibody production</td>
</tr>
<tr>
<td></td>
<td>Nitric oxide</td>
<td>Specific cytolytic T cell generation</td>
</tr>
<tr>
<td></td>
<td>Cytokines (IL-1, IL-6, others)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chemokines (IL-8, others)</td>
<td></td>
</tr>
</tbody>
</table>

IFN, interferon; IL, interleukin.

**Figure 18-1** Ancient evolutionary origin of the innate immune system. Studies of the immune systems of a wide range of vertebrates and invertebrates have revealed that even the most primitive invertebrates possess many components of innate immunity (e.g., pattern recognition receptors of the lectin and Toll-like families, antimicrobial peptides, complement proteins). The innate immune system is thus extremely ancient, having arisen early in the evolution of multicellular life. In contrast, the adaptive immune system is a much more recent development that did not appear until emergence of the ancestors of present-day sharks and rays, approximately 400 million years ago. The first species to acquire an adaptive immune system based on immunoglobulin-type receptors must have arisen after the appearance of the direct ancestors of present-day jawless fish (lampreys and hagfish), which are the most highly evolved living species that lack the ability to generate large families of variable immunoglobulin-type lymphocyte receptors (arrow). ACP, alternative complement pathway; CCP, classic complement pathway; Ig, immunoglobulin; LCP, lectin-activated complement pathway; LRR, leucine-rich repeat domain; MAC, membrane attack complex; MHC, major histocompatibility complex; PRR, pattern recognition receptor. (Adapted from Sunyer JO, Zarkadis IK, Lambris JD: Complement diversity: a mechanism for generating immune diversity? Immunol Today 19:519, 1998.)
receptors that are unrelated to immunoglobulins but also provide the basis for an adaptive immune response have been discovered in primitive jawless fish such as lampreys and hagfish. This finding shows that at least two different strategies for the creation of an adaptive immune system emerged at the dawn of vertebrate evolution about 500 million years ago, and it emphasizes the importance of adaptive immunity for survival and further evolution of the vertebrate lineages.

Given this key role of adaptive immunity in the evolution and survival of vertebrates, it is surprising that all invertebrate animals, and possibly some of the lowest vertebrate species as well, completely lack the ability to generate lymphocyte populations bearing large families of clonally diverse antigen receptors. In these animals, protection against pathogen invasion depends entirely on innate immunity, elements of which appear to exist in all animals and plants and must have evolved with the earliest multicellular forms of life. In many cases, components of the innate immune system are significantly conserved in structure and function in animals from the lowliest invertebrates to the most complex vertebrates. This preservation of innate immune mechanisms, with their functions largely intact, over such vast evolutionary distances is a clear indication of their importance, even in animals that have developed sophisticated adaptive immune responses.

PATHOGEN RECOGNITION BY THE INNATE IMMUNE SYSTEM

Some mechanisms of innate immunity are constitutive, meaning that they are continuously expressed and are not significantly modulated by the presence or absence of infection. Examples include the barrier functions provided by epithelial surfaces continuously exposed to microbial flora, such as those of the skin and intestinal and genital tracts. In contrast, the inducible mechanisms of innate immunity involve increased production of mediators and upregulation of effector functions that eliminate microorganisms. Induction occurs as a result of exposure to a wide variety of microbes and represents a less specific form of immune recognition than that associated with the specific antibodies and T cells that mediate adaptive immunity. The basic principle underlying this form of response is a process known as pattern recognition. This recognition strategy is based on the detection of commonly occurring and conserved molecular patterns that are essential products or structural components of microbes.

PAMPS AND DAMPS: PATTERNS FOR INNATE IMMUNE RECOGNITION

Pathogen-Associated Molecular Patterns

The general name given to the targets of innate immune recognition in microbes is pathogen-associated molecular patterns (PAMPs). These structural features or components distinctive for microorganisms are not normally found in the animal host. The best-known example of a PAMP is bacterial lipopolysaccharide (LPS), a ubiquitous glycolipid constituent of the outer membranes of gram-negative bacteria. Another important example is the peptidoglycan structure present as the basic cell wall component in nearly all bacteria. These structures may vary partially from one bacterium to another, but the basic elements are conserved, thus providing the possibility of recognizing a broad array of pathogens by sensing a single or a relatively small number of PAMPs. Many PAMPs that serve as targets of recognition for the innate immune response are now known to be associated with bacteria, fungi, and viruses.

In addition to allowing direct recognition of molecules produced by various microorganisms, the innate immune system is able to respond to the patterns of host-derived molecules released by cells undergoing necrotic death. The molecules recognized are generally referred to as damage-associated molecular patterns (DAMPs), and include multiple different families of proteins, as well as nonproteinaceous substances such as uric acid microcrystals. Thus, the response to DAMPs can be an indirect response to microbial invasion, or it can be triggered by other types of tissue damage such as ischemia to result in sterile inflammation.

Pattern Recognition Receptors

Recognition of PAMPs and DAMPs is mediated by a collection of germline-encoded molecules known collectively as pattern recognition receptors (PRRs) (Table 18-2). These receptors are host proteins that have evolved, through millions of years of natural selection, to possess precisely defined specificities for particular PAMPs or DAMPs expressed by microorganisms. The total number of PRRs present in complex vertebrates such as humans is estimated to be several hundred—a number limited by the size of the genome of any animal and the number of genes it can dedicate to immune protection. The human genome, for example, is estimated to contain approximately 20,000 to 35,000 genes, most of which are not related directly to the immune system. This demonstrates one of the strong points of contrast between innate and adaptive immune systems, because the latter can possess in the range of 10^14 different somatically generated receptors for foreign antigens in the form of antibodies and T-cell receptors. With its much more limited array of receptors, the innate immune system uses the strategy of targeting highly conserved PAMPs that are shared broadly by large classes of microorganisms. Because most pathogens contain PAMPs, this strategy allows the generation of at least partial immunity against most infections.

PRRs are expressed by many cell types, some of which are specialized effector cells of the immune system (e.g., neutrophils, macrophages, dendritic cells, lymphocytes), and others of which are not generally regarded as part of the immune system (e.g., epithelial and endothelial cells). Unlike the T and B cell receptors used for adaptive immune recognition, expression of PRRs is not clonal, which means that all receptors displayed by a given cell type (e.g., macrophages) have identical structure and specificity. When PRRs are engaged by recognition of their associated PAMPs or DAMPs, effector cells bearing the PRRs are triggered to perform their immune effector functions immediately, rather than after undergoing proliferation and expansion, as in the case of adaptive immune responses. This accounts for the much more rapid onset of innate immune responses.
In recent years, considerable progress has been made toward identifying many of the important PRRs involved in the induction of innate immunity. These receptors can be classified into three functional classes: secreted, endocytic, and signaling PRRs (see Table 18-2). In addition, many of the known PRRs can be classified into structurally defined families on the basis of a few characteristic protein domains. Among these, the best known include proteins with calcium-dependent lectin domains, scavenger receptor domains, and leucine-rich repeat domains.

### Pattern Recognition Receptors of the Lectin Family

Calcium-dependent lectin domains are common modules of secreted and membrane-bound proteins involved in the binding of carbohydrate structures. A well-characterized PRR belonging to this class is the mannan-binding lectin (MBL), also known as soluble mannose-binding protein, which represents a secreted PRR that functions in initiation of the complement cascade (Figure 18-2).^{55} This protein is synthesized primarily in the liver on a constitutive basis, although its production can be increased as an acute phase reactant following many types of infection. MBL binds to carbohydrates on the outer membranes and capsules of many bacteria, as well as fungi, some viruses, and parasites. Although mannose and fucose sugars bound by MBL can also be found on the surfaces of normal mammalian cells, they are present at too low a density or in the wrong orientation to efficiently engage the lectin domains of MBL. In contrast, the coats of many microorganisms contain an array of these sugars, which allows strong binding of MBL. Thus, in this case, the spacing and orientation of specific carbohydrate residues constitute the PAMP that triggers the activation of innate immunity by MBL. MBL functions as one of a small number of secreted PRRs that can initiate the lectin pathway of complement activation. At least two other soluble proteins with lectin activity in human plasma, known as ficolins (ficolin/P35 and H-ficolin), can also activate this pathway following their interaction with bacterial polysaccharides.^{10}

Several of the soluble lectin-type PRRs also play an important role in the opsonization of microbes by binding to their surfaces and directing them to receptors on phagocytic cells. Among these are two pulmonary surfactant proteins, SP-A and SP-D, which similarly recognize and bind to their surface sugar codes of microbes in the respiratory tract.^{11} These molecules are similar in structure to MBL, having both collagen-like and lectin domains, and together they constitute a family of soluble PRRs known as collectins. Another family of soluble PRRs that performs a similar function in plasma is the pentraxins, so called because they are formed by the association of five identical protein subunits.^{12} This family includes the acute phase reactants C-reactive protein (CRP) and serum amyloid P protein (SAP), along with a number of so-called long pentraxins, which have an extended polypeptide structure with homology to the classic short pentraxins (i.e., CRP and SAP) only at their carboxy-terminal domains. Long pentraxins are expressed in a variety of different tissues and cells, and their specific functions are mostly unknown. However, the long pentraxin PTX3 has been shown to play an important, nonredundant role in resistance to fungal infection in mice, and recent studies indicate that PTX3 is essentially a functional ancestor of antibodies that recognizes microbes and promotes their clearance through complement activation and phagocytosis.^{13}

---

**Table 18-2  Pattern Recognition Receptors (PRRs)**

<table>
<thead>
<tr>
<th>Receptor Class</th>
<th>Examples</th>
<th>Prominent Sites of Expression</th>
<th>Major Ligands</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secreted PRRs</td>
<td>Collectins, Mannan-binding lectin, Ficolins, Surfactant proteins (SP-A, SP-B), Pentraxins, Short pentraxins (CRP, SAP), Long pentraxins, Lectin-family receptors, Macrophage mannose receptor, DEC-205, Dectin-1, Scavenger receptor A, MARCO, Complement receptors, CD11b/CD18 (CR3), CD21/35 (CR2/1)</td>
<td>Plasma</td>
<td>Carbohydrate arrays typical of bacterial capsules, fungi, and other microbes</td>
<td>Complement activation, Opsonization</td>
</tr>
<tr>
<td>Endocytic PRRs</td>
<td>Toll-like receptors, CARD/NOD proteins, Pyrin domain proteins</td>
<td>Macrophages, dendritic cells, some endothelia, epithelia, and smooth muscle cells</td>
<td>Cell wall polysaccharides (mannans and glucans), LPS, LTA, and opsonized cells and particles</td>
<td>Pathogen uptake by phagocytes, Delivery of ligands to antigen-processing compartments, Clearance of cellular and extracellular debris</td>
</tr>
<tr>
<td>Signaling PRRs</td>
<td>Toll-like receptors, CARD/NOD proteins, Pyrin domain proteins</td>
<td>Macrophages, dendritic cells, epithelia</td>
<td>Multiple conserved pathogen-associated molecular patterns (LPS, LTA, dsRNA, lipoproteins, flagellin, bacterial DNA, others)</td>
<td>Activation of inducible innate immunity (antimicrobial peptides, cytokines, reactive oxygen or nitrogen intermediates), Instruction of adaptive immune response</td>
</tr>
</tbody>
</table>

CARD, caspase activation and recruitment domain; CR, complement receptor; CRP, C-reactive protein; DEC-205, dendritic and epithelial cells, 205 kD; dsRNA, double-stranded RNA; LPS, lipopolysaccharide; LTA, lipoteichoic acid; MARCO, macrophage receptor with collagenous structure; NOD, nucleotide-binding oligomerization domain; SAP, serum amyloid P protein; SP, surfactant protein.
In addition to these soluble proteins, a large number of membrane-bound glycoproteins with lectin domains are known to exist; some of them participate in innate immunity by serving as endocytic PRRs for the uptake of microbes or microbial products.14,15 (Figure 18-3). One of the most extensively studied of these is the macrophage mannose receptor (MMR).16 Although originally identified on alveolar macrophages and known to be expressed on macrophage subsets throughout the body, this receptor is also expressed on a variety of other cell types, including certain endothelia, epithelia, and smooth muscle cells. The MMR is a membrane-anchored, multilectin domain–containing protein that mediates the binding of a broad range of pathogens, leading to their internalization via endocytosis and phagocytosis. Although the major function of the MMR appears to be directing the uptake of its ligands, evidence suggests that this receptor may be capable of signaling to modify macrophage functions following receptor engagement.17 Another member of this receptor family, the β-glucan binding cell-surface lectin known as dectin-1, has a role in the modulation of inflammation in a mouse model of infection-induced arthritis.18

Pattern Recognition Receptors of the Scavenger Receptor Family

The scavenger receptor family contains a broad range of structurally diverse cell surface proteins that are expressed most prominently on macrophages, dendritic cells, and endothelial cells (see Figure 18-3). Although they were originally defined by their ability to bind and take up modified serum lipoproteins, they also bind a wide range of other ligands, including bacteria and some of their associated products. Multiple members of this family have been implicated as PRRs for innate immunity. These include the scavenger receptor A (SR-A) and a related molecule called the macrophage receptor with collagenous structure (MARCO).20 Both of these molecules contain a scavenger receptor cysteine-rich domain in the distal ends of their membranes and a collagen-like stalk with a triple-helical structure. Both are known to bind bacteria, and SR-A also binds well-known PAMPs such as lipoteichoic acids and LPS.21,22 Mice that have been made deficient in SR-A by targeted gene disruption show increased susceptibility to infections caused by a variety of bacteria, thus providing strong evidence of the role of scavenger receptors in protective immunity, most likely through the activation of innate immune mechanisms.23,24 Members of the class B scavenger receptor family, including CD36 and SR-BI/CLA-1, have also been found to recognize a variety of pathogen-derived molecules.19 Although these members of the scavenger receptor family clearly function as endocytic PRRs in the uptake of microbes, their potential to serve as signaling receptors has not yet been established. However, several scavenger receptors play a co-receptor role in the signal transduction process mediated by members of the Toll-like receptor (TLR) family (discussed later), most likely by capturing specific ligands and transferring them to adjacent TLRs.19

Pattern Recognition Receptors with Leucine-Rich Repeat Domains

Leucine-rich repeat domains (LRRs) are structural modules found in many proteins, including PRRs involved in signaling the activation of innate immunity. Molecules in this class include, most notably, the family of mammalian Toll-like receptors (TLRs), which are membrane-bound signal-transducing molecules that play a central role in the
Part 3

The first member of the Toll family to be discovered was the *Drosophila* Toll protein, which was identified as a component of a signaling pathway that controls dorsoventral polarity during development of the fly embryo. The sequence of Toll showed it to be a transmembrane protein with a large extracellular domain containing multiple tandemly repeated LRRs at the N-terminal end, followed by a cysteine-rich domain and an intracellular signaling domain (Figure 18-4). A role for Toll in immune responses was suggested by the observation that its intracellular domain shows homology to the mammalian interleukin-1 receptor (IL-1R) cytoplasmic domain. This association was later confirmed in studies showing that Toll was critical for the antifungal response in the fly, linking this pathway for the first time to innate immunity.

The sequence of Toll eventually led to a search for similar proteins in mammals; this effort has been richly rewarded, yielding a family of 10 Toll-like receptors (TLRs) in humans and 12 in mice. Among these, TLR1 through TLR9 are conserved between mice and humans, TLR10 is present only in humans, and TLR11 through TLR13 are expressed only in mice. All these molecules contain large extracellular domains with multiple LRRs, as well as intracellular signaling domains known as Toll/IL-1R, or TIR, domains. Many of these TLRs have been linked to innate immune responses against various PAMPs of different microorganisms.

**Toll-like Receptor 4 and the Response to Lipopolysaccharide.** The first human TLR to be identified was the molecule now designated TLR4, which is a major component in the response to one of the most common of all PAMPs—bacterial LPS. Earlier studies on the response to LPS had identified two proteins—CD14 and LPS-binding protein—as molecules involved in the binding of LPS to the surface of LPS-responsive cells. However, these molecules did not possess any potential for transducing signals into the cell, so it was unclear how LPS binding would lead to the activation of cellular responses associated with gram-negative bacterial infection. The answer was provided by positional cloning studies of the LPS gene in the LPS-hyporesponsive C3H/HeJ mouse. This study revealed a single amino acid substitution in the signaling domain of TLR4. Specific deletion of the TLR4 gene by targeted gene disruption in mice subsequently confirmed the essential role of this molecule in the response to LPS, because TLR4 knockout mice have almost no response to LPS and are highly resistant to endotoxic shock. Biochemical studies provide further support for TLR4 as a component of the LPS receptor; they show that LPS bound to the surface of cells is in close contact with both CD14 and TLR4, as well as another protein called MD-2, which appears to perform an accessory function in the binding of LPS to the receptor complex. Additional studies have elucidated the downstream elements in the signaling pathways that connect TLR4 to the activation of genes associated with inducible innate immunity (see Figure 18-4).

Studies of Toll signaling pathways in *Drosophila* have identified the transcription factor nuclear factor κB (NFκB) as one of the key effectors of gene activation following the engagement of Toll. This basic pathway in the fly appears to be largely conserved in TLR signaling in higher animals, including mammals.

**Other Pathogen-Associated Molecular Patterns Recognized by Toll-like Receptors.** The search for ligands that lead to signaling through various TLRs has demonstrated that this family of PRRs is collectively responsible for innate immune responses to an extraordinary array of PAMPs. In addition to its central role in the signaling of responses to LPS, TLR4 is involved in responses to multiple different self- and non–self-ligands. The antimitotic agent and cancer chemotherapy drug paclitaxel (Taxol) has been shown to mimic LPS-induced signaling in mouse cells through a pathway that requires both TLR4 and MD-2. Other foreign ligands of TLR4 include the fusion protein (F protein) of respiratory syncytial virus and the heat shock...
protein 60 (HSP60) of chlamydia.\textsuperscript{42} TLR4 can signal in response to mammalian HSP60, a protein expressed at increased levels and most likely released by stressed or damaged cells.\textsuperscript{33} This represents a variation of the pattern recognition principle in which the pattern is an endogenous molecule released by damaged host cells that serves as a DAMP, rather than a PAMP produced directly by a pathogen. Other examples of recognition of DAMPs by TLR4 include responses to oligosaccharide breakdown products of tissue hyaluronans and responses to the extra-domain A region of fibronectin produced by alternative RNA splicing in response to tissue injury or inflammation.\textsuperscript{31,44}

The range of PAMPs recognized through TLR2 is probably even greater than for TLR4. TLR2 is known to be involved in signaling in response to multiple PAMPs of gram-negative and gram-positive bacteria, including such structures as bacterial glycolipids, bacterial lipoproteins, parasite-derived glycolipids, and fungal cell wall polysaccharides.\textsuperscript{10} TLR2 does not function independently in responding to these PAMPs; rather, it forms heterodimers with TLR1 or TLR6. This ability to pair with other TLRs appears to be unique to TLR2, because other TLRs that have been studied carefully (e.g., TLR4, TLR5) most likely function only as monomers or homodimers. Other TLRs with currently defined ligands are TLR5 (involved in the response to bacterial flagellin), TLR3 (double-stranded RNA), TLR7 (single-stranded RNA), and TLR9 (unmethylated bacterial DNA).\textsuperscript{10} It is apparent that most, if not all, microbes contain multiple PAMPs that are recognized by different TLRs. For example, a typical bacterium expressing LPS also contains unmethylated DNA and thus generates signals not only through TLR4 but potentially through TLR9 as well. Because different TLRs are capable of activating distinct signaling cascades (see Figure 18-4), the ability of a single cell to detect several different features of a pathogen simultaneously with multiple TLRs may help the innate immune response to be more finely tuned to respond to a particular challenge.\textsuperscript{45}

**CARD and Pyrin Domain Proteins.** A large number of cytosolic proteins that have structural similarities to membrane-bound TLRs and that function as sensors for PAMPs of intracellular pathogens and regulators of innate immune responses have been identified. Many of these proteins contain LRR domains and have been classified on the basis of their incorporation of a CARD or pyrin domain. The nomenclature and classification schemes for this growing family of innate immune sensors and regulators are still evolving, and it has been proposed that they should be grouped and classified as members of a single family designated the CATERPILLER (CARD, R [purine]-binding, pyrin, lots of leucine repeats) family.\textsuperscript{46} However, the most recent literature shows a trend toward referring to this group of proteins as the NLR family, an acronym that stands for either “nucleotide-binding domain, leucine-rich repeat proteins”\textsuperscript{47} or “Nod-like receptors.”\textsuperscript{50} The first intracellular microbial sensors in this family to be described were the
Nod1 and Nod2 proteins, which contain LRR domains linked to a central nucleotide binding and oligomerization (NOD or NACHT) domain and an N-terminal CARD domain. As in the case of TLRs, the LRR domains of these proteins appear to be involved in the recognition of pathogen-derived molecules and a variety of host components that function as DAMPs, and their CARD domains are linked to downstream signaling for the activation of innate immunity. Although originally implicated in response to bacterial LPS, it is now well accepted that both Nod1 and Nod2 are primarily involved in the recognition of murine peptide monomers released from bacterial cell wall peptidoglycans. Signals resulting from the recognition of peptidoglycan components by Nod1 and Nod2 lead to activation of the NFκB pathway, as in the case of TLR signaling. However, other signaling pathways also appear to be engaged, such as activation of procaspase-1 and caspase-9 by CARD domain interactions, leading to increased production of IL-1β and cell death through a process referred to as pyroptosis.

The pyrin domain-containing proteins represent a major subgroup of NLRs that are believed to signal in response to microbial invasion or cellular stress. The prototype member of this family is pyrin, which is the product of the gene that is mutated in those with familial Mediterranean fever. Although pyrin itself lacks an LRR domain, numerous other members of this family contain an LRR linked to a central NOD domain and an N-terminal pyrin domain. These include cryopyrin (also known as NLRP3 or NALP3), which is mutated in patients with a range of hereditary inflammatory diseases referred to collectively as cryopyrin-associated periodic syndromes. Cryopyrin, together with these multiple related proteins, constitutes a large set of related proteins known as NLRP (NLR-PYRIN domain) or Nalp (NACHT-LRR-PYRIN domain-containing proteins) family. The human genome contains 14 genes encoding NLRP proteins, the precise functions of which are still largely unknown. However, several NLRP proteins, particularly NLRP3 and NLRP1, have been identified as key components in the formation of intracellular complexes known as inflammasomes. These cytosolic protein complexes serve as activating platforms that are involved in the activation of caspases—intracellular proteases required for the processing of inflammatory cytokines such as IL-1β and IL-18. Direct recognition of specific PAMPs by NLRP proteins remains to be established, although initial studies implicate these proteins as direct or indirect sensors of various stimuli, including constituents of bacteria (peptidoglycan, bacterial RNA, exotoxins), viruses (double-stranded RNA), and uric acid crystals.

**EFFECTOR MECHANISMS OF INNATE IMMUNE RESPONSES**

The ability to recognize pathogens through PRRs allows activation of numerous antimicrobial effector mechanisms by the innate immune response. These responses lead to the killing of pathogens through production of effector molecules with direct microbicidal activities, including the membrane attack complex of complement, a variety of antimicrobial peptides, and the caustic reactive oxygen and reactive nitrogen intermediates generated within phagocytic cells. In invertebrates, these mechanisms represent virtually the entire protective response against microbial invaders. However, in most vertebrates, including mammals, innate immune recognition also has profound effects on triggering and programming the adaptive immune response that follows somewhat later. This ability of the innate immune system to instruct the adaptive response has major implications for the development of long-term protective immunity to infection and may play a critical role in mechanisms leading to autoimmunity.

**CELL TYPES MEDIATING INNATE IMMUNITY**

Many types of cells have the ability to mount at least a limited response to PAMPs, but the most effective cell types in this regard are the specialized phagocytes, such as macrophages, neutrophils, and dendritic cells. Upon recognition of microbial stimuli, these cells have the ability to upregulate nicotinamide adenine dinucleotide phosphate (NADPH) oxidase by assembling the components of this enzyme complex on phagosomal membranes, leading to an oxidative burst that produces microbicidal superoxide ions. Many phagocytic cells also increase their expression of inducible nitric oxide synthase (iNOS, or NOS2) upon contact with various PAMPs. This leads to the production of reactive nitrogen intermediates, including nitric oxide and peroxynitrite, which have potent direct antimicrobial activities. These responses are synergistic because the antimicrobial activity of the phagocyte oxidase system is frequently enhanced by the expression of reactive nitrogen intermediates.

**Innate-like Lymphocytes**

A number of distinct lymphocyte subsets also play important roles in innate immune responses. One group of such lymphocytes, the natural killer (NK) cells, appears to be a true member of the innate immune system. These lymphocytes do not express receptors generated by somatic recombination and thus depend on germline-encoded receptors for signaling their responses against pathogen-infected cells. NK cells participate in the early innate response against virally, and probably bacterially, infected cells through expression of cytotoxic activity and secretion of cytokines.

Several other subsets of lymphocytes belonging to the T and B cell lineages have been identified as participants in the rapid response against pathogens to which the host has not previously been exposed. Although these cells express clonally variable, somatically rearranged antigen receptors (T cell antigen receptors or membrane immunoglobulins) and thus could be classified as components of the adaptive immune system, their manner of functioning is much more characteristic of innate than adaptive immunity. These innate-like lymphocytes (ILLs) may represent remnants of the earliest primitive adaptive immune system, and they appear to have been conserved to varying degrees because they continue to make specialized contributions to host immunity.

Among the currently recognized ILLs are two B cell populations, known as the B1 and marginal zone B cell
Antimicrobial Peptides

Antimicrobial peptides are the key effector molecules of inducible innate immunity in many invertebrates and are being increasingly recognized as important elements of innate immunity in higher animal species, including mammals.\(^64\) They are evolutionarily ancient components of host defense that are widely distributed throughout all multicellular organisms in the animal and plant kingdoms. More than 1,000 such peptides have been identified, and their diversity is so great that it is difficult to categorize them. However, at a structural and mechanistic level, most of these peptides share several basic features. They generally are composed of amino acids arranged to create an amphipathic structure with hydrophobic and cationic regions. The cationic regions target a fundamental difference in membrane design between microbes and multicellular animals, which is the abundance of negatively charged phospholipid head groups on the outer leaflet of the lipid bilayer. The preferential association of antimicrobial peptides with microbial membranes leads to membrane-disrupting activity, most likely involving the interaction of the hydrophobic regions of the peptide with membrane lipids.\(^65\)

Antimicrobial peptides produced in response to engagement of various PRRs account for most of the inducible immunity against microbes noted in many invertebrate animals and plants. Although these peptides are probably less central to host immunity in most vertebrates, evidence indicates that they make important contributions to immunity in more highly evolved animals, including mammals.\(^66\) In humans, active antimicrobial peptides, such as the α- and β-defensins, are constitutively or inducibly produced in skin and epithelia of the gastrointestinal and respiratory tracts.\(^67\) These molecules most likely act as natural preservatives of epithelia that are colonized or frequently exposed to microbial flora. Because the acquisition of resistance against these agents by sensitive microbial strains is extremely unusual, antimicrobial peptides are of great interest as templates for the development of new antimicrobial pharmaceuticals.\(^67,68\)

**INFLUENCE OF INNATE MECHANISMS ON ADAPTIVE IMMUNITY**

In addition to functioning as a first line of defense against invading pathogens, a critical feature of the innate immune system in higher animals such as mammals is its effect on activating the adaptive immune system. In fact, it is now clear that in most situations, the adaptive immune system mounts a response to a pathogen only after the pathogen has generated signals via PRRs of the innate immune system. This principle serves as the basis for the adjuvant effect, which is the observation that antibody and T cell responses are efficiently generated against protein antigens only if these are introduced together with a nonspecific activator of the immune system, which is generically known as an adjuvant. Most adjuvants are in fact extracts or products of bacteria, and it is clear that in most or all cases, adjuvant effects result from activation of the innate immune response.\(^69\)

Innate immune responses can prime or potentiate the adaptive immune response in many ways (Figure 18-5). In the case of T cell responses, one extremely important and well-recognized mechanism involves the upregulation of co-stimulatory molecules. T cells require at least two signals to become activated from a naïve resting state. One signal is provided through the T cell antigen receptor by its binding to a specific peptide ligand presented by an MHC class I or II molecule. The second signal is provided by one of several co-stimulatory ligands that are expressed by specialized antigen-presenting cells such as dendritic cells (see Chapter 13). The best studied of these are the molecules of the B7 family—B7-1 (CD80) and B7-2 (CD86)—which engage the activating receptor CD28 on the surface of the T cell. Expression of B7 family co-stimulatory molecules on the surface of antigen-presenting cells is controlled by the innate immune system, such that these molecules are

---

1. These are involved in the spontaneous production of natural antibodies, which are largely germline-encoded immunoglobulins that are reactive to commonly expressed microbial determinants. In addition, both of these B cell populations generate rapid T cell–independent responses following bacterial challenges and thus contribute to the first line of immune defense that precedes the onset of adaptive immunity.
2. Among the T cells, two populations of ILLs have been identified and characterized in detail: γδ T cells and NK T cells. The γδ T cells express somatically rearranging receptors that use a limited number of variable region genes and are thought to recognize a narrow spectrum of foreign or self-ligands.\(^61\) In humans, the specificities of two subsets of γδ T cells have been at least partially defined. One of these, the major circulating population expressing the Vδ2 gene product, responds rapidly and without prior immunization to a variety of small alkyl phosphate and alkyl amine compounds that are produced by many bacteria. Another subset, characterized by its expression of the Vδ1 gene product, responds to major histocompatibility complex (MHC) class I–related self-molecules of the MHC class I chain–related A and B (MICA/B) and CD1 families.\(^61\) These molecules may serve as markers of cellular stress and are upregulated on cells in the context of infection or inflammation, leading to the activation of Vδ1-bearing γδ T cells.
3. A similar principle appears to be involved in the functioning of NK T cells, which are so named because of their co-expression of an αβ T cell antigen receptor and a variety of receptor molecules that are typically associated with NK cells.\(^61\) Similar to γδ T cells, NK T cells have somatically rearranged antigen receptors that use a limited array of V genes and most likely recognize a narrow range of foreign or self-antigens. A major population of NK T cells is reactive with the MHC class I–like CD1d molecule, and these ILLs appear to be activated by recognition of a variety of lipid or glycolipid ligands that can be presented by CD1d. Recently, several bacterial glycolipids have been identified as specific antigens that stimulate NK T cells, suggesting that these cells may be rapid responders that contribute to innate antibacterial immunity.\(^65\) A wide variety of mouse disease models have shown that NK T cells also make significant contributions to the development of adaptive immune responses and may play a particularly important role in immunoregulation to prevent autoimmunity.\(^62,63\)
Instruction on the adaptive immune response by the innate immune system. When an antigen-presenting cell (APC) comes into contact with pathogen-bearing pathogen-associated molecular patterns (PAMPs), responses are triggered via innate immune mechanisms that dramatically alter the ability of the APC to stimulate an adaptive (T cell–mediated) immune response. For example, signals generated by contact with PAMPs such as lipopolysaccharide (LPS) with Toll-like receptor 4 (TLR4) lead to the activation of transcription factor nuclear factor κB (NFκB), which enters the nucleus of the APC and assists in switching on genes for cytokines (e.g., interleukin [IL]-1, -6, and -12) and a variety of chemokines and co-stimulatory molecules (e.g., the B7 family members CD80 and CD86). In addition, binding of the pathogen to endocytic pattern recognition receptors (PRRs) such as the mannose receptor leads to delivery of the pathogen to endosomes (Endo) and lysosomes (Lys). There, the protein antigens of the pathogen are partially degraded to generate antigenic peptides that can be presented by major histocompatibility complex (MHC) class II molecules for recognition by the T cell antigen receptors (TCRs) of specific T cells. These effects of pattern recognition by the innate immune system lead to expression of the signals required for activation of quiescent antigen-specific T cells and the subsequent generation of specific antibodies. (Adapted from Medzhitov R, Janeway C Jr: Innate immunity, N Engl J Med 343:338, 2000.)

Recent studies have shown that innate immune signaling through TLRs has a major impact on the responses of phagocytic antigen-presenting cells; it also provides an important second signal for immunoglobulin production by B cells. In the case of phagocytic cells, uptake of microbes by phagocytosis and subsequent maturation of the phagosome are stimulated by concurrent TLR signaling. In dendritic cells, which are the major antigen-presenting cells for the priming of T cell responses, TLR signaling has a major impact on whether antigens from phagocytosed microbes are effectively presented on MHC class II molecules. For B cells responding to foreign antigens, it has been demonstrated that concurrent signaling through TLRs is necessary for the efficient stimulation of T cell–dependent differentiation into plasma cells and subsequent antibody secretion. This concept is also relevant for T cell responses to autoantigens, including several prominent nuclear antigens that are targets of autoantibodies in rheumatic disease.

Innate immune responses also trigger the production of many cytokines and chemokines, which enhance the development of adaptive immune responses and change the nature of the adaptive response generated. For example, contact between dendritic cells and PAMPs such as LPS or bacterial lipoproteins leads to the production of IL-12 as a result of signaling through TLRs. This cytokine acts on antigen-specific T cells to promote their differentiation into T helper type 1 cells, which are associated with the production of interferon-γ and other effecter mechanisms that favor the clearance of bacterial pathogens. In the case of myeloid lineage dendritic cells, signaling through TLRs (and potentially other PRRs) induces a process known as maturation, which is associated with increased expression of antigen-presenting and co-stimulatory molecules that enables the efficient priming of naïve antigen-specific T cells.

This requirement for the innate immune response to “switch on” the expression of molecules required for the priming and differentiation of T cell responses helps ensure that proinflammatory adaptive immune responses occur mainly in the setting of a relevant infectious challenge. After activation, helper T cells control other components of adaptive immunity, such as the activation of cytotoxic T cells, B cells, and macrophages. Innate immune recognition, therefore, appears to control all major aspects of the adaptive immune response through the initial recognition of infectious microbes by PRRs. The discovery of self-molecules that act as DAMPs further extends this paradigm to include innate immune responses that are triggered by tissue damage. This extended view, sometimes referred to as the “danger model,” helps explain why certain self-ligands produced or released in the setting of infection or tissue damage can function in essentially the same manner as the PAMPs associated with microorganisms.

**DISEASE ASSOCIATIONS INVOLVING INNATE IMMUNITY**

Given the obvious role of the innate immune response in virtually all types of infectious disease, one might expect that gross defects in the mechanisms of innate immunity occur relatively rarely and in association with clinical immunodeficiency. In fact, increasing evidence indicates that mutations that inactivate various innate immune pathways can lead to increased pathogen sensitivity in both laboratory mice and humans. Because many of the pathways leading to innate immunity are amplified during recurrent or prolonged activation of the immune system, they must also participate in mediating tissue damage in chronic inflammatory disease. In addition, certain self-molecules that are produced or released at increased levels as a result of inflammation, including heat shock proteins, nucleic acids, and microcrystals of monosodium urate or calcium pyrophosphate, may act as DAMPs. These may signal through TLRs or other PRRs to stimulate adjuvant-like effects that increase the potential for autoreactive lymphocytes to be activated.

Perhaps a more surprising finding has been that some defects in innate immunity are associated with a markedly increased predisposition to autoimmune disease. Several different mechanisms have been proposed to explain this paradoxical association. Mechanisms of the innate immune response play an important role in the clearance of self-antigens released from necrotic or apoptotic cells, resulting in a noninflammatory clearance of self-antigens that tends...
to favor tolerance rather than the stimulation of immune responses.\textsuperscript{85} Failure of such clearance may lead to excessive exposure to self-antigens, triggering normally silent autoreactive lymphocyte clones to expand and differentiate into effector cells. This may account for the development of lupus-like autoimmunity in mice with targeted deletion of the gene for the short pentraxin SAP, which, along with other components of the innate immune system, appears to play a significant role in the clearance of DNA/chromatin complexes.\textsuperscript{86} Reduced levels of serum mannos-binding lectin in humans also appear to be a risk factor for the development of systemic lupus erythematosus, possibly because of the role of this soluble PRR in facilitating the clearance of apoptotic cells.\textsuperscript{87}

Deficiencies of early components of the classic pathway of complement activation have been strongly associated with lupus-like autoimmunity in both humans and mouse models.\textsuperscript{88-92} This may be the result of alterations in the clearance of apoptotic cells or other sources of self-antigens, resulting in increased stimulation of normally silent autoreactive lymphocytes.\textsuperscript{93,94} An alternative, but nonexclusive, mechanism relates to involvement of the complement system, particularly the early components C1 and C4, in facilitating the induction of self-tolerance by the adaptive immune system by increasing the localization of autoantigens such as double-stranded DNA and nucleoproteins within the primary lymphoid compartment.\textsuperscript{95,96} Thus, a deficiency of C1 or C4 appears to result in failure to delete or functionally inactivate autoreactive B cell clones as they arise during lymphopoiesis in the bone marrow.\textsuperscript{95,97} Studies carried out in mouse models suggest that this tolerance-inducing mechanism is partially disrupted in animals that are deficient in a variety of other components of innate immunity, including SAP and the complement receptors CD21/CD35.\textsuperscript{96,95}

Multiple examples of links between defects in signaling receptors of the innate immune system and chronic inflammatory diseases have emerged from studies of the CARD and pyrin families of cytosolic PRRs.\textsuperscript{90,97} The first association of this type was provided by genetic mapping studies that identified the Nod2 protein as the product of the IBD1 locus, which contributes to disease susceptibility in a subset of patients with Crohn’s disease.\textsuperscript{98-101} This soluble PRR of the CARD family normally functions by inducing cytokine production in response to bacterial peptidoglycan, but mutant alleles associated with increased risk of Crohn’s disease are defective in this function.\textsuperscript{98} In this case, it may be failure of innate immunity to adequately control bacterial colonization or infection in the intestine that leads to the final expression of disease. Consistent with this view, a recent study has demonstrated diminished expression of a class of antimicrobial peptides (β-defensins known as cryptdins) in Paneth cells from the ileum of patients with Crohn’s disease and Nod2 mutations.\textsuperscript{102} Other findings suggest that defective signaling by mutant variants of Nod2 can result in reduced production of immunoregulatory cytokines such as IL-10, perhaps resulting in uncontrolled inflammation in the intestine.\textsuperscript{103} Other studies have established links between various members of the pyrin family and specific chronic inflammatory disorders. These include the causative association of mutations in pyrin with familial Mediterranean fever, and of cryopyrin with cryopyrin-associated periodic syndromes.\textsuperscript{86,47} These diseases and other chronic inflammatory or autoimmune disorders associated with specific deficiencies in innate immune mechanisms are frequently considered together as autoinflammatory diseases.\textsuperscript{104} Recognition that these diseases are frequently associated with dysregulation of inflammatory cytokine production, in particular IL-1β, has led to some striking therapeutic advances in the treatment of selected patients using systemic administration of the IL-1 receptor antagonist anakinra.\textsuperscript{105-107}

Deficiencies in at least two populations of ILLs—NK cells and NK T cells—have been associated with multiple autoimmune syndromes in both humans and mice.\textsuperscript{82,108-110} This is believed to reflect a significant role for these ILLs in regulating adaptive immune responses, although the precise mechanisms by which they act still are not fully understood. Given the complex interplay between innate immunity and adaptive immunity, it is extremely likely that associations between alterations in innate immunity and autoimmune diseases will continue to emerge. As for some of the examples cited here, a fuller understanding of these associations is likely to lead to new and successful therapies for autoimmune and autoinflammatory diseases.

**Future Directions**

The last two decades of research in immunology have seen a great emphasis on the fundamental role played by innate immune mechanisms in all immune responses. The innate immune system in humans represents the accumulation of many stages of evolution and natural selection that began with the most primitive organisms. Because of the ancient origins of the innate immune system, some of the most important discoveries in the field of innate immunity have come from studies of relatively simple animals such as flies and worms. Now that many of the pieces of this elaborate system have been discovered and categorized, continued research efforts are turning increasingly toward exploring the roles of various components of innate immunity in the human immune system. These efforts are very likely to yield insights into many currently unexplained diseases and may provide targets for new therapeutics.

**Connection to the Clinic**

- Innate immune responses serve as the foundation of all immunity; they participate at some level in all infectious, inflammatory, and autoimmune diseases.
- Drugs targeting specific innate immune effectors are beginning to emerge as useful therapeutic agents.
- Innate immune receptors are responsible for triggering the clinical syndromes associated with uric acid crystals and potentially other types of crystal-induced arthritis.
- Primary defects in specific innate immune molecules have been identified as the cause of a range of uncommon disorders known as autoinflammatory diseases.
- Many systemic autoimmune diseases are associated with genetic polymorphisms of molecules involved in innate immune responses.
- Research into innate immunity has provided important new insight into potential mechanisms for common autoimmune diseases such as Crohn’s disease and systemic lupus erythematosus.
References


The references for this chapter can also be found on www.expertconsult.com.
Adaptive Immunity and Organization of Lymphoid Tissues

MICHAEL L. DUSTIN

KEY POINTS

- Lymphocytes are born, and many self-reactive cells are deleted in primary lymphoid tissues.
- Immune responses are usually initiated in secondary lymphoid tissues.
- Tertiary lymphoid tissue can be generated at sites of inflammation and may promote tissue-specific autoimmunity.
- Self-tolerance is established by antigen recognition in primary lymphoid tissues or elsewhere in the absence of inflammation.
- Generation of an adaptive immune response is dependent on innate immune system activation.
- Dendritic cells sense both tissue-specific factors and innate immune stimuli in shaping T cell responses to antigens.

The adaptive immune system is so named because it can adapt to virtually any pathogen or toxin that enters the body. Although invertebrates defend themselves through innate immunity alone, vertebrates have all developed some form of adaptive immunity—the ability to generate novel receptors by genetic recombination mechanisms that can then be selected to recognize diverse macromolecules associated with rapidly evolving pathogens. A molecule that can be recognized by the adaptive immune system is known as an “antigen” (Table 19-1).

Mechanisms to distinguish self from foreign, as well as benign foreign from harmful foreign, macromolecules are critical processes in successful adaptive immunity. At the core of these mechanisms is the essential partnership of adaptive and innate immunity, first formulated by Charles Janeway, Jr. Although the adaptive immune system can recognize any foe and focus powerful effector mechanisms to destroy this foe, its ability to calculate the relative risks of mounting an immune response or becoming tolerant to a given recognized structure is guided by innate recognition of pathogen-associated molecule patterns (PAMPs) or inflammation triggered by tissue damage–associated molecular patterns (DAMPs).

The function of the adaptive immune system is tightly linked to its anatomy. In some respects T lymphocytes are cells without boundaries that can be found in almost any tissue at any time. T and B lymphocytes are readily monitored in patients because they are reliably found in the blood of normal individuals. However, the vast majority of the T and B lymphocytes are located in secondary lymphoid tissues where they search for antigens. There are three types of tissues that concentrate lymphocytes—the primary lymphoid tissues, where these cells are born; the secondary lymphoid tissues, where they search for antigens; and tertiary lymphoid tissues that form at sites of chronic inflammation. Secondary and tertiary lymphoid tissues have a characteristic and functionally important organization into B cell and T cell zones. In fact, all of the previously mentioned histologic findings are directly related to functional goals of the adaptive immune system, as is clarified subsequently.

This chapter addresses the basics of adaptive immunity and its partnership with innate immunity in the context of the anatomic sites in which these responses take place.

LYMPHOCYTE MIGRATION PARADIGMS FOR HOMING, INTERSTITIAL NAVIGATION, AND EGRESS

Because we talk about adaptive immunity in the context of secondary lymphoid tissues, this is a good time to review what is known about lymphocyte migration. It has been appreciated since the early 1960s that lymphocytes “recirculate” between secondary lymphoid tissues and the blood and more recently that, on activation, they take on different tissue-specific homing properties that are important for function. If one arbitrarily begins this circuit with a naïve T cell in the blood, then there are three distinct transitions that can be considered: (1) interaction with the vessel wall and extravasation; (2) interstitial locomotion in the tissue parenchyma; and (3) egress from the tissue parenchyma back into the blood, sometimes via lymph. General features of these three steps are addressed here, and relevant details will be commented on in different functional contexts in relation to adaptive immunity.

Multistep Paradigm for Extravasation

The movement of lymphocytes from blood to tissues is complicated by the high-flow rates in blood vessels. Springer
Any molecular structure recognized by the adaptive immune system. The ligand for B cell receptor (BCR)/antibody or T cell receptor (TCR)

Family of small secreted or shed proteins (typically 8 kDa) that bind to G protein–coupled receptors and activate or attract cells. A chemical (chemo-) that induces movement (chemo-kinetics)

Family of adhesion molecules that are specialized for adhesion in the context of migration and generation of contractile force/mechanical stabilization. Noncovalent heterodimers that are regulated rapidly by signaling from chemokine receptors and antigen receptors

Family of adhesion molecules that are specialized for mediating initial attachment of leukocytes to the vessel wall from flowing blood. Have amino-terminal calcium-dependent lectin domains and interact with carbohydrate ligands that can incorporate protein determinants

Table 19-1 Terms and Definitions

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Chemokine</th>
<th>Integrin</th>
<th>Selectin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any molecular structure recognized by the adaptive immune system. The ligand for B cell receptor (BCR)/antibody or T cell receptor (TCR)</td>
<td>Family of small secreted or shed proteins (typically 8 kDa) that bind to G protein–coupled receptors and activate or attract cells. A chemical (chemo-) that induces movement (chemo-kinetics)</td>
<td>Family of adhesion molecules that are specialized for adhesion in the context of migration and generation of contractile force/mechanical stabilization. Noncovalent heterodimers that are regulated rapidly by signaling from chemokine receptors and antigen receptors</td>
<td>Family of adhesion molecules that are specialized for mediating initial attachment of leukocytes to the vessel wall from flowing blood. Have amino-terminal calcium-dependent lectin domains and interact with carbohydrate ligands that can incorporate protein determinants</td>
</tr>
</tbody>
</table>

and Butcher established the current paradigm called the “multistep model” for T cell extravasation.

The first step is the initial tethering of the free-flowing leukocyte to the vessel wall. A special class of adhesion molecules known as selectins and their carbohydrate ligands mediates this step (see Table 19-1). The selectin family comprises three members: L-, E-, and P-selectin. L-selectin is expressed on naïve T and B lymphocytes and is essential for the entry of these cells into lymph nodes, but not the spleen. The ligand for L-selectin is a complex of sulfated sialic acid bearing complex carbohydrates linked to different protein backbones expressed on high endothelial cells in postcapillary venules in primary, secondary, and tertiary lymphoid tissues. E- and P-selectin are expressed on activated endothelial cells at sites of inflammation in diverse tissues. They bind to glycoprotein ligands expressed on leukocytes, which have terminal sialyl-Lewis-X blood group antigens and can also incorporate other structural modifications. The high affinity ligand for P-selectin is the protein backbone PSGL-1 with a stretch of sulfated tyrosines that make up part of the ligand. PSGL-1 can be expressed by lymphocytes without the necessary secondary modification to make it a ligand for P-selectin. Therefore a fusion of P-selectin to immunoglobulinFc that can be purified and fluorescently tagged is the best probe to determine if a leukocyte is competent to bind to P-selectin expressing endothelial cells. The genetic basis of forming selectin ligands is complex with requirements for expression of core proteins, specific sialotransferases, fucosyltransferases, and sulfotransferases. Defects in fucose metabolism are the basis for a rare genetic immunodeficiency/mental retardation syndrome called leukocyte adhesion deficiency type II (LAD-II). Selectins will only mediate leukocyte tethering and rolling, not arrest and extravasation.

Selectin-mediated tethering allows the leukocyte to bring G protein–coupled receptors close enough to the vascular wall to bind ligands attached to glycoproteins. Chemokine receptor signaling is critical to activate closely linked integrin family members to bind their ligands. This chemokine signal is pertussis toxin sensitive, indicating that it involves Gα coupled receptors. With respect to lymphocyte recirculation, there are three important chemokine receptors: CCR7, CXCR4, and CXCR5. CCR7, binding to ligands CCL19 and CCL21, is the most important chemokine system for entry of T cells into secondary lymphoid tissues. CXCR4 binding to its ligand CXCL12 also contributes to entry of T and B cells. CXCR5, binding to its ligand CXCL13, is the major system controlling entry of B cells into secondary lymphoid tissues. Following activation, CCR7 is downregulated on many effector T cells and other chemokine receptors are upregulated, allowing these cells to home to peripheral sites of inflammation. Activated endothelial cells in these tissues express ligands that selectively recruit subsets of activated T cells. In contrast, activated B cells either become memory cells that retain CXCR5 and CCR7 expression or differentiate into plasma cells that downregulate CXCR5 and CCR7 and upregulate CXCR4, targeting these cells to medullary cords via a novel migratory mechanism and bone marrow.

Integrin family members are the immediate recipients of chemokine signals to produce rapid arrest of rolling leukocytes and to initiate the extravasation process. The major integrin that mediates homing to secondary lymphoid tissues is LFA-1, which is composed of the αL and β2 subunits. LFA-1 is expressed only on leukocytes and binds to intercellular adhesion molecules (ICAMs), of which the best characterized are ICAM-1 and ICAM-2. ICAM-1 and ICAM-2 are the major ICAMs expressed on endothelial cells, with ICAM-1 displaying regulated expression in response to inflammatory mediators like tumor necrosis factor (TNF) and interferon (IFN)-γ.

The deficiency of the integrin β2 subunit is the basis of a rare genetic syndrome known as leukocyte adhesion deficiency type I (LAD-I). In this disease leukocyte extravasation at sites of inflammation is defective and patients are highly susceptible to bacterial infections of the skin and mucous membranes. Patients with LAD-I are developmentally normal and can be treated by bone marrow transplantation with a high success rate. A leukocyte adhesion deficiency type III in which multiple leukocyte integrins show defects in regulation of Rap1, a small G protein important in LFA-1 regulation, has also been described.

The structure of integrins reveals remarkable machinery for regulated adhesion. The inactive form is folded into a compact globular structure, in which the ligand-binding domain points toward the leukocyte surface. Following activation by chemokines, the integrin extends to two times its original height and projects the ligand binding site to greater than 20 nm from the leukocyte membrane, with orientation toward the endothelial cell surface. This dramatic change is closely coupled to cytoskeletal association, providing anchorage needed for arrest and cell spreading following ligand binding.

A second integrin expressed on naïve and activated lymphocytes called VLA-4 is composed of the α4 and β1 subunits and can play a small role in entry into lymph nodes but a major role in entry into inflamed sites. Its ligand is VCAM, also a member of the immunoglobulin superfamily regulated by inflammatory cytokines. Inhibition of VLA-4 by a monoclonal antibody (natalizumab) is the basis of an approved treatment for multiple sclerosis that decreases leukocyte entry into the central nervous system. Natalizumab therapy has been associated with rare cases of progressive multifocal leukoencephalopathy, a severe infection caused by reactivation of latent JC virus. Thus risks and benefits
need to be weighed carefully with immunosuppressive therapies even when only a single specific pathway is targeted. When T lymphocytes are activated to home to mucosal effector sites, they upregulate expression of the integrin β7, which also associates with α4 to form the gut-homing integrin α4β7, and bind a different immunoglobulin superfAMILY ligand called MAdCAM (for mucosal addressin cell adhesion molecule) expressed on endothelial cells in the gut.32

The extravasation process involves the movement of lymphocytes between or through endothelial cells.33 Endothelial junctional complexes include special adhesion molecules that need to be transiently disengaged to allow lymphocyte passage between endothelial cells. The transcellular pathway may be dominant in situations where the endothelial junctions are particularly sturdy, as in the brain, thymus, or lymph nodes. Both junctional and transcellular routes involved active processes in the leukocyte and endothelial cells, but this step is not thought to be regulated to control homing decisions.

To summarize, three types of receptor ligand pairs define the key regulated steps in lymphocyte homing. The compatibility of all three is required to gain entry into the tissue. If a compatible selectin-ligand pair is not available, the leukocyte will not be able to initiate adhesion to the endothelial wall and will flow past. If a chemokine receptor-chemokine pair is not available, it will be impossible to activate integrins, even if the selectin-ligand pair mediates rolling or tethering, and the cell will eventually release and remain in the blood. If the integrin-ligand pair is not compatible, the cells will not be able to arrest and extravasate, even if the selectin-ligand pair and chemokine-ligand systems are engaged. Thus, these molecular pairs can be thought of as a hierarchic area code for lymphocyte homing—the digits defined by the compatible interactions, each of which must be correctly engaged to allow entry.34

**Tissue Organization and Interstitial Migration**

Classical histology and mouse genetics have been used to establish the molecular mechanisms that account for the segregation of T cells and B cells within secondary lymphoid tissues. The T cell zone is defined by the production of CCL19 and CCL21 by stromal cells and the expression of CCR7 on T cells.17 Interestingly, these are the same signals that trigger the arrest of T cells on endothelial cells for tissue entry. T cell zones are also amply populated by conventional dendritic cells (DCs), which express CCR7 as they mature. DCs appear to form a network on a scaffold of reticular fibers.35 The parenchyma of lymph nodes and splenic white pulp nodules are crisscrossed by thick collagen bundles sheathed in fibroblastic reticular cells.36 The inner compartment of these fibers forms a network of conduits in the lymph node and spleen that allows DCs in the parenchyma access to the afferent lymph or blood, respectively. The B cell follicles depend on CXCL13 expressed by follicular stromal cells and CXCR5 expressed on B cells. In fact, the balance of CXCR5 and CCR7 expression by B cells controls their proximity to the boundary between T cell zone and B cell zone, where B cells can encounter helper T cells.21 B cell zones are also populated by variable numbers of follicular DCs, which are differentiated stromal cells, rather than cells of hematopoietic origin.

Two-photon laser scanning fluorescence microscopy has revealed the dynamics of immune cells in secondary lymphoid tissues. Fluorescently labeled T cells moved with an average speed of 12 μm/min in T cell zones, and B cells moved 30% slower in the follicles of acute organ-cultured lymph nodes of live mice.37 The paths taken by T and B cells appeared random, but subsequent analysis of fluorescent lymphocyte movement in the presence of differentially labeled stromal cells revealed that the lymphocytes were guided by a ramified stromal network and changed direction frequently as they encountered branches in the stromal scaffolding.38 The DC network is supported by the stromal network, ensuring that T cells contact DCs as they follow a random pathway through the stromal network.35 Thus the process by which a DC that has brought antigen from the periphery can show this antigen to many T cells is based on forming random contacts with thousands of T cells per hour. Due to its random nature, this model has been referred to as “stochastic repertoire scanning.”

The major signal that stimulates the rapid and random migration of T and B cells in lymph nodes is thought to be chemokines presented on the surface of stromal cells. The speed of B lymphocyte migration is significantly reduced in mice lacking Gαq subunit of G protein–coupled receptors, further supporting the model that chemokines may have a role in this process.39 Chemokines can act as chemoattractants but can also stimulate migration in a random fashion—called chemokinesis—under some conditions. Rapid T cell motility in vitro can be recapitulated by surfaces coated with only CCR7 ligands.40 Somewhat surprisingly, integrins such as LFA-1 contribute little to migration of DCs or T cells.40,41

Although the initial scanning process is random, once antigen-bearing DCs begin to interact with antigen-specific T cells, a number of nonrandom elements of migration are established. B cells that recognize antigen upregulate CCR7 and downregulate CXCR5 such that they are attracted to the boundary between the T and B cell zones.21 DCs that have been in contact with CD4+ T cells produce CCL3 and CCL4, which attract CD8+ T cells under inflammatory conditions.32 This directed migration of CD8+ T cells biases the scanning of the CD8+ T cell repertoire toward DCs that have received T cell help and thus have come into contact with interesting antigens. This process likely increases the efficiency of repertoire scanning.

**Immunologic Synapses Maintain Antigen-Specific Interactions with Dendritic Cells**

The most extreme change in lymphocyte migration during an immune response is the near full arrest referred to as an immunologic synapse.33 In vitro analysis has revealed elaborately organized structures underlying the arrest of T cell migration in contact with DCs bearing appropriate MHC-peptide complexes.44 In vivo imaging analysis has repeatedly revealed stable T cell–DC interactions as a common feature of both tolerance and immunity induced by high-affinity ligands.45 Although LFA-1-ICAM-1 interactions are not necessary for migration in lymph nodes, these adhesion molecules are required for stable synapses.46 Immunologic synapses have been proposed to lead to asymmetric cell divisions that generate effector T cell and memory T cell divisions even when only a single specific pathway is targeted.
precursors.\(^\text{47}\) Dynamic T cell–DC interactions referred to as kinapses may set up symmetric divisions that help maintain the differentiated state of these early precursors.\(^\text{58}\)

**Egress from Lymph Nodes and the Thymus: Sphingosine-1-Phosphate**

The recirculation of lymphocytes requires that they periodically cease their local scanning activity or activation process in secondary lymphoid tissues and exit to the lymph or blood. This is referred to as egress. Egress is also important for recently matured lymphocytes to leave the bone marrow or thymus to move to secondary lymphoid tissues. Insight into the molecular processes controlling egress from lymph nodes and the thymus were provided through investigation of the fungal metabolite derivative FTY720, which has been approved as a drug (fingolimod) for treatment of relapsing-remitting multiple sclerosis.\(^\text{49}\) FTY720 administration rapidly decreases T and B cell numbers in blood. FTY720 is phosphorylated by sphingosine kinase and acts as an agonist for a number of sphingosine-1-phosphate receptors that are expressed on lymphocytes and endothelial cells. Further insight was provided by the study of bone marrow chimeric mice in which fetal liver cells from embryonic lethal sphingosine-1-phosphate receptor 1 (SIPR1) knockout mice were used to reconstitute lethally irradiated syngeneic wild type mice.\(^\text{50}\) The SIPR1-deficient, recently matured lymphocytes are unable to exit the thymus. Furthermore, when the mature SIPR1-deficient lymphocytes trapped in the thymus were transferred into wild-type recipients, they were able to enter lymph nodes normally but could not egress from lymph nodes due to failure to counterbalance CCR7-dependent retention signals. These results suggested a T cell autonomous defect in egress from both the thymus and lymph node. Although FTY-720-P is an agonist of SIPR1, it has the effect of downregulating expression of the receptor such that the compound recapitulates the knockout phenotype. A parallel study with a reversible SIPR1 agonists and antagonists suggests that SIPR1 also has a role in controlling lymphatic endothelium permeability and this effect would contribute to arrest of egress is a similar manner to the T cell autonomous effects.\(^\text{51}\)

**B Cell Development in the Bone Marrow**

Immature B cells express surface BCRs composed of surface immunoglobulin noncovalently complexed to the signal transducing subunits Igα and Igβ.\(^\text{52}\) The majority of immature B cells are reactive to self-antigens and even display “polyreactivity,” meaning that they produce antibodies that bind not just one but many self-antigens.\(^\text{53}\) These auto-reactive cells are mostly lost from the repertoire in two checkpoints—the encounter of immature B cells with antigens in the bone marrow or by newly matured B cells on entry into secondary lymphoid tissues.\(^\text{53}\) Checkpoint 1 depends on antigen display in the bone marrow. If BCRs on immature B cells encounter surface immobilized antigens on bone marrow cells or extracellular matrix, they will undergo apoptosis. The second source is the blood. The bone marrow parenchyma is bathed in blood plasma antigens due to fenestrated endothelial cells lining bone marrow sinusoids. When soluble antigens bind to their BCRs, they may undergo apoptosis if the antigen is multivalent and cross-links the BCR, inducing strong signaling. If the antigen is monovalent, the B cell may become anergic—a form of nonresponsiveness that is induced by weaker signaling through the BCR. In the second checkpoint newly matured B cells that have egressed from the bone marrow will encounter antigens on a variety of cells. For example, DCs present intact antigens to B cells\(^\text{54}\) and because DCs sample many tissue antigens, this may provide a wider array of tissue-specific self-antigens to check against the BCR. These selection checkpoints decrease the frequency of self and polyreactive B cells by greater than 10-fold, but some polyreactive B cells persist in the repertoire. Polyreactivity may contribute to pathogen recognition, as recently shown for neutralizing antibodies to low-density envelope glycoprotein of HIV.\(^\text{55}\) Thus the adaptive immune system may always need to tolerate some self-reactivity to defend us against the universe of pathogens.

**T Cell Development in the Thymus**

T cell precursors arrive in the thymus and initiate TCR gene rearrangement in this microenvironment. The thymus is an epithelial organ formed during development from the third pharyngeal arch. Histologically, the tissue has a clearly distinguishable cortex, medulla, and vascular corticomедullary junction.\(^\text{7}\) Figure 19-1 is adapted from a recent paper examining the dynamics of T cell selection in the thymus.\(^\text{56}\) It shows fluorescence staining with *Ulex europaeus* antigen-1 (UEA-1) lectin, a marker for the thymic medulla, on the left side and major histocompatibility complex (MHC) class II antigen in the thymic cortex on the left side. The stepwise path of thymocyte migration is outlined on the left panel. Early CD4 and CD8 negative progenitors enter via postcapillary venules at the corticomедullary junction (step 1) and migrate to the subcapsular region of the cortex, where TCR gene rearrangement takes place and CD4 and CD8 become expressed on the surface of immature T cells, also called thymocytes (step 2). These cells then migrate

### PRIMARY LYMPHOID TISSUES: SITES WHERE T AND B CELLS ARE GENERATED AND SELF-TOLERANCE MECHANISMS ARE INITIATED

The critical event in the birth of T and B cells is the rearrangement of the antigen receptor genes to generate T cell antigen receptors (TCRs) and B cell antigen receptors (BCRs) that are expressed on the cell surface. Each T cell and B cell has a single antigen receptor as birth certificate and identification card. The life, death, or expansion of each of these cells controls the availability of this antigen-binding structure to the adaptive immune system. This is the heart of the clonal selection theory first proposed by Burnett in the 1950s. Self-reactive cells need to be deleted or retasked shortly after their birth. This process takes place in the bone marrow for B cells and the thymus for T cells. The anatomy of these two sites is different in keeping with the dramatically different recognition mode employed by B and T cells. First we discuss the bone marrow space, followed by the functional anatomy of the thymus.
randomly among thymic epithelial cells in the cortex sampling self-MHC-peptide complexes (step 3). If they express a TCR that recognizes self-MHC-peptide complexes with low affinity, they will undergo “positive selection.” Positively selected thymocytes then mature into CD4⁺ or CD8⁺ positive cells that migrate to the medulla under control of CCR7 ligands⁵⁷ (step 4). Medullary thymic epithelial cells express a transcription factor called AIRE that mediates expression of a number of tissue-specific genes by amplifying expression from poorly expressed chromatin regions.⁵⁸ Rare patients lacking expression of AIRE have a complex autoimmune syndrome characterized by polyendocrinopathy and other tissue-specific autoimmune diseases. The hypothesis is that AIRE-mediated transcription of otherwise tissue-specific genes in thymic epithelial cells promotes negative selection of some autoreactive T cells and may also promote generation of tissue antigen-specific regulatory T cells (step 5). Conventional positive selection requires Ras activation on intracellular membranes, whereas stronger Ras activation at the plasma membrane mediates negative selection.⁵⁹ Regulatory T cell generation requires signaling through protein kinase C-θ and c-Rel transcription factors.⁶⁰ The small percentage of thymocytes that express a TCR that pass both positive and negative selection downregulate CD69 and S1PR1 and exit the lymph nodes as naïve “conventional” T cells (step 6).

SECONDARY LYMPHOID TISSUES: SITES WHERE ANTIGEN FINDS RARE SPECIFIC T AND B CELLS

The frequency of naïve T cells that recognize any particular antigen is so low that it has been challenging to estimate. Recently, enrichment methods using antibody-coated magnetic beads have been developed to assist direct measurement of precursors by flow cytometry with high-avidity MHC-peptide–based probes known as tetramers.⁶¹ The number of cells specific for commonly used model antigens is approximately 500 in most mouse strains when pooling cells from all secondary lymphoid tissues (spleen and lymph nodes). There are approximately 500 million total CD4⁺ T cells in these tissues, such that antigen-presenting DCs need to make contact with a million irrelevant T cells to find one specific T cell. Thus a small number of antigen-positive DCs early in an infection must have access to highly concentrated swarms of T cells in secondary lymphoid tissues to sample enough T cells to find a few specific precursors to activate and expand. As described earlier, this search process is initially random but may become more directed and efficient as a response progresses. Although naïve T cells recirculate in an unbiased manner through secondary lymphoid tissues, the antigen-carrying cells, primarily DCs, have a high degree of regional bias in their movement, such that they are thought to relay information about both innate immune stimulation and tissue of origin of antigens.

Once antigen-specific T cells come into contact with DCs they form immunologic synapses; integrate signals through the T cell receptors and co-stimulatory ligands, which represent part of the innate immune system contribution to T cell activation; and divide more than 20 times with short cell cycle times of approximately 6 hours.⁶² Expansion of CD8⁺ T cells, which give rise to cytotoxic effector cells, is greater in general than for CD4⁺ T cells, which give rise to various types of helper T cells. After expansion, which peaks around day 7 to 10, the infectious agent is often eradicated and most of the effector T cells undergo apoptosis. The flulike symptoms associated with viral infections are due to cytokines produced by the dividing and differentiating T cells.

Because primary adaptive responses take a week or longer to develop, the host is dependent on innate immune mechanisms like natural antibodies, neutrophils, interferons, and natural killer cells to control the infection until sufficient numbers of effector T cells are generated. Thousands of the expanded T cells that survive after the pathogen is destroyed become memory cells.⁶³ Although the process of asymmetric division may contribute to establishment of memory and effector cells, linear models in which memory cells develop from effector cells are supported by fate mapping experiments demonstrating that functional memory can arise from effector T cells that expressed granzyme B, a cytotoxic effector molecule.⁶⁴

There are two subsets of memory T cells: (1) central memory cells that express L-selectin, CCR7, and LFA-1 and recirculate via secondary lymphoid tissues and (2) effector memory cells that lack L-selectin and CCR7 but may express P- and E-selectin ligands and other chemokine receptors such as CXCR4, CCR5, CCR4, and/or CCR9.⁶⁵ These effector memory cells migrate to peripheral sites of inflammation and are equipped for rapid effector function.

Figure 19-1 Section of mouse thymus stained with Ulex europaeus antigen-1 (UEA-1) lectin (left) or anti-major histocompatibility complex (MHC) class II (right) visualized with immunofluorescence microscopy. The UEA-1 staining is strongest in the medulla but also highlights the capsule and stroma of the cortex. The cortical thymic epithelial cells are strongly positive for MHC molecules, which allow negative and positive selection of thymocytes. (Images courtesy Richard Lewis, Stanford University.)
if they encounter antigen. Memory cells respond rapidly to recurrence of the same infection and together with antibodies rapidly eradicate that same agent if it is encountered a second time. These memory cells may also cross-react with other pathogens, and depending on the degree of cross-reaction, this type of response can result in rapid clearance of a new pathogen or sometimes an impaired response.

Antigens from Blood Are Detected Most Efficiently in the Spleen and Liver (Portal System)

The spleen is a large visceral organ that filters approximately 5% of the cardiac output. The red pulp is an important location for removal of aged red blood cells from the circulation. The red pulp also contains many macrophages that specialize in this clearance process and DCs that come into direct contact with naïve T cells that are in the blood. The function of these red pulp DCs is not known. Most attention has been focused on the white pulp nodules and the marginal zone as sites of T cell–DC interaction and antigen capture, respectively (Figure 19-2, left). Blood flows into the spleen via an artery that splits into arterioles that empty into venous sinuses in the marginal zone and red pulp (Figure 19-2, right). The blood is then re-collected into a venous system that drains to the liver, joining with the portal tract. The arterioles are surrounded by sheaths of T cells (the periartrial lymphoid sheath [PALS]) that make up the T cell zones of the white pulp. Bridging channels connect the red pulp, where lymphocytes leave the blood, to the PALS, where they migrate in a pertussis toxin–sensitive manner. B cell follicles and the marginal zone then surround the PALS. Macrophages, DCs, and marginal zone B cells line the marginal zone, where they have direct access to blood antigens. DCs that pick up antigens in the marginal zone migrate to the PALS within 9 hours. A reticular fiber network connects the marginal zone to the PALS and allows soluble antigens from the blood to reach resident DCs in the PALS. Thus the spleen provides multiple opportunities to mount primary and recall responses to particulate or soluble antigens in the blood.

As mentioned earlier, the spleen drains into the liver. Immune responses in the liver are poorly understood, but this is an important site because many pathogens colonize the liver. Two blood circulations supply the liver: the portal vein with deoxygenated blood from the gut and spleen and the hepatic artery with oxygenated blood. These circulations mix in the liver sinusoids, a low-pressure network of blood spaces between sheets of hepatocytes connecting the portal tract and the central collecting vein. Thus most of the liver parenchyma appears as plates of hepatocytes alternating with blood carrying sinusoids (Figure 19-3, left). The liver is rich in a type of sessile macrophage called a Kupffer cell and patrolling lymphocytes, particularly natural killer T cells (Figure 19-3, right). The major antigen-presenting cells in the liver are not Kupffer cells but the perivascular Ito cells. The liver also plays an important role in the clearance of B cells mediated by therapeutic anti-CD20 antibodies that are used to treat patients with rheumatoid arthritis.

Antigens from Mucosal Surfaces Are Detected Most Efficiently in Peyer’s Patches and Mesenteric Lymph Nodes

The mucosal-associated lymphoid tissues include the tonsils, Peyer’s patches, lamina propria, cryptopatches, and appendix. Peyer’s patch and lamina propria DCs have different

Figure 19-2 Organization of the spleen. Left, Tissue section from a mouse injected with low-molecular-weight (blue) and high-molecular-weight (red) fluorescent dextrans. The low-molecular-weight dextran labels the red pulp (blue) and the high-molecular-weight dextran labels macrophages in the marginal zone (red) such that the white pulp nodule containing approximately \(10^7\) lymphocytes per cm\(^2\) is dark. The white pulp nodule forms around a central arteriole from which smaller arterioles branch to the red pulp sinuses. B and T cells are segregated into follicles (B cells) and the periarteriolar lymphoid sheath (T cells). (Scale bar 0.2 mm.) Right, Image of a live mouse spleen during injection of low-molecular-weight fluorescent dextran. The blood carrying the dextran passes through the white pulp in small blood vessels (arrows) emanating from the central arteriole (not shown) and connected to red pulp sinuses that rapidly fill with blood, which fills the marginal sinus but not the white pulp itself. In this image approximately 0.3% of the T cells in the PALS are labeled with a fluorescent dye. These areas are packed with T and B cells. Marginal zone macrophages and B cells have direct access to blood to capture pathogens. White pulp lymphocytes are not in direct contact with blood and need antigen presentation by cells that migrate from the marginal zone or antigens that are delivered to dendritic cells via reticular fibers. (Scale bar 0.1 mm.) (Courtesy Janelle Waite, New York University.)
mechanisms for sampling the contents of the gut lumen. Peyer's patches are composed of large B cell follicles with smaller T cell zones (Figure 19-4, top and bottom right). They have high endothelial venules that allow efficient entry of naïve T and B cells, as well as memory cells with gut homing phenotypes. The large size of the follicles in the Peyer's patches causes a domelike effect with the epithelium protruding into the lumen. Some of the microvilli-laden absorptive epithelial cells in this dome region are replaced by smooth-surfaced M cells. These cells act as relay points for entry of enteric pathogens into the dome region of Peyer's patches where a population of CCR6+ DCs efficiently presents pathogen-related antigens to induce rapid local T cell responses \(^\text{72}\) (Figure 19-4, bottom left). In contrast

---

**Figure 19-3**  Microcirculation and immune cells in liver. *Left*, Intravital imaging of blood space in mouse liver. Confocal imaging of hepatocytes (green autofluorescence) and blood space (red fluorescent dextran). The flow pattern is toward the center of the image panel. *Right*, Intravital imaging of NKT cells (bright green) and SIGN R1+ Kupffer cells in liver sinusoids (dark spaces) between hepatocytes (faint green autofluorescence). The Ito cells would also be lining the sinusoids, but on the other side of the endothelial cells in the space of Disse. (Courtesy Tom Cameron, New York University.)

**Figure 19-4**  Gut-associated lymphoid tissues. *Top right*, Immunofluorescence image of Peyer's patch with B cells in green and T cells in red. \(^\text{102}\) *Bottom right*, A similarly oriented view of Peyer's patches from mouse in which CCR6+ dendritic cells also express GFP due to targeting of gfp to CCR6 locus. \(^\text{72}\) *Left*, Images of CD11c+ dendritic cells (DCs) in the tips of microvilli in the terminal ileum after exposure to *Salmonella* bacteria. The DCs extend processes through the epithelial layer to assist bacterial uptake. \(^\text{103}\) The CCR6+ DCs rather than the CX3CR1+ lamina propria DCs appear to be the most important for the T cell response to *Salmonella* antigens.
to the Peyer’s patches, the core of the villi in the terminal ileum is populated by activated T cells, plasma cells, DCs, and macrophages. The DCs migrate to mesenteric lymph nodes, where they present antigens to stimulate production of regulatory T cells and effector T cells to balance gut tolerance and immunity.\(^{13}\) CX3CR1\(^+\) macrophages actively sample the gut contents and then may present antigens to effector and regulatory T cells that patrol the lamina propria\(^{14}\) (Figure 19-4, bottom left). The gut presents a barrier between the host and billions of commensal bacteria and food antigens. There are also many pathogens—bacteria, viruses, protozoans, and worms—that exploit this niche. This is a rapidly developing area of research and is touched on again in the context of T cell differentiation.

**Antigens from Other Tissues and Solid Organs Are Detected in Peripheral Lymph Nodes**

An extensive network of peripheral lymph nodes filters lymph from the skin, visceral organs, and nervous system. Lymph is composed of fluid that leaves blood vessels and then must be collected from the tissues in afferent lymphatic vessels, passes through at least one lymph node, exits the lymph node as efferent lymph, and is returned to the blood via the thoracic duct. The afferent lymphatics of a lymph node connect to the capsule of the node and drain into the subcapsular sinus, which contains many macrophages (Figure 19-5). The floor of the subcapsular sinus covers the lymph node parenchyma and is the point of origin of the reticular fiber network that connects to the high endothelial venules and medullary cords, where cells move from the parenchyma into the efferent lymph. Lymph does not come into direct contact with cells in the parenchyma, but cells lining the reticular fibers and a space between the high endothelial venules and the reticular fibroblast sheath are exposed to lymph fluid.\(^{36}\) The parenchyma is divided into T and B cell zones defined by CCL19/21 and CXCL13 producing stroma (see Figure 19-5). DCs migrate from peripheral tissues in a CCR7-dependent manner and join networks of DCs in the T cell zones with scattered cells in the follicles.\(^{35}\) Emigrant DCs have been reported to array around high endothelial venules, where they are efficiently encountered by newly extravasated T and B cells. Antigen-positive DCs have been shown in experimental models to stop both B and T cell migration during antigen encounter.\(^{45,54}\) These immunologic synapses, discussed earlier, play an important role in antigen transfer for B cells and for priming of proliferative response in T cells. It has been noted that different populations of DCs such as dermal DCs and Langerhans cells actually populate different subregions of the T cell zones, but the significance of this is not clear.

---

**Figure 19-5** Lymph node schematic and dendritic cell morphologies. Schematic: B, B cell follicles; T, T cell zone; M, Medullary cords. Other structures are labeled. Lymph enters via the afferent lymphatics and exits via the efferent lymphatics. T cells enter the lymph node via the high endothelial venules and exit via medulla to the efferent lymphatic. Intravital microscopy–based images of dendritic cells in lymph nodes of CD11c-YFP marker mice are linked to a schematic of the lymph node. (All scale bars 50 µm.) (From Lindquist RL, Shakhar G, Dudziak D, et al: Visualizing dendritic cell networks in vivo, Nat Immunol 5:1243-1250, 2004.)
Peripheral Tolerance Induction under Steady-State Conditions

Lymph nodes are associated with peripheral tolerance induction. DCs' presentation of antigens in the steady state induces proliferation of T cells that are then deleted or anergic by 7 days. This process may be particularly important to tissue-specific antigens that are not present in the thymus in the steady state. It is known that low-affinity self-antigens can escape peripheral tolerance induction by this mechanism. Low-affinity self-antigen–specific T cells that are then activated in the context of infection by a pathogen with strong innate stimulation may then induce tissue-specific autoimmunity.

Regulatory T Cells Reduce Autoreactivity by Inhibiting Immunologic Synapse Formation

Regulatory T cells are CD25+ IL-2 dependent T cells that express the transcription factor FoxP3 and have the ability to suppress immune responses to tissue-specific self-antigens. These cells represent another layer in redundant mechanisms to prevent autoimmune attack by adaptive immunity. One way in which regulatory T cells function is to block immunologic synapse formation between T cells and autoantigen-presenting DCs. This mechanism appears to operate on low-affinity self-antigens that are the most likely type of self-reactive T cell to escape other peripheral tolerance mechanisms. Prevention of long-lived T cell–DC interactions may reduce proliferation and cytokine production by autoreactive T cells.

Changes in the Lymph Node during Infection/Vaccination

Infection or vaccination in tissues induces a strong reaction in draining lymph nodes. Innate signals trigger production of inflammatory cytokines that lead to increased blood flow, increased adhesion molecule and chemokine expression to increase T and B cell entry, and suppression of T and B cell exit by downregulation of S1PR1. DCs in reactive lymph nodes express higher levels of co-stimulatory ligands such as CD80 and CD86 and promote robust proliferation, survival, and differentiation of antigen-specific T cells. T cells that are activated in the lymph nodes regain expression of S1PR1 after 3 to 4 days and a new repertoire of homing molecules and migrate to effector sites.

Tissue Environment of Immature Dendritic Cells Determines T Cell Imprinting

When T cells are activated in lymph nodes, the repertoire of homing molecules expressed by the activated effector T cells is determined in part by the origin of the DC. DCs arise from a common monocyte-like precursor that migrates into tissues via the blood. DCs that drain from the gut produce retinoic acid from vitamin A. Following maturation and migration to draining lymph nodes, these DCs secrete retinoic acid, which induce activated T cells to express gut-homing chemokine receptors like CCR9 and gut-specific integrins like α4β7. Because gut-associated postcapillary venules express MAdCAM and present CCL25 (a CCR9 ligand), these effector T cells will tend to home to the gut. In the absence of retinoic acid produced by gut-derived DCs, the signals induced by skin-derived DCs favor expression of ligands for E- and P-selectin and CCR4 on T cells. Because endothelial cells of skin-associated postcapillary venules express P-selectin and present CCL17, these T cells will tend to home to inflamed skin. Recently, it has been shown that DCs in the skin metabolize vitamin D to generate a signal for T cell expression of CCR10, which allows these cells to migrate to the epidermis in response to CCL27. Although DCs will strongly skew T cells to home back to the sites from which the DCs migrated, the expression of homing receptors and chemokines on lymphocytes also has a stochastic component, which means that these effector cells and memory cells will also show up in diverse peripheral sites scattered throughout the animal—giving effector T cells their ability to appear in any tissue at any time. Overall, these results show that DCs sense their tissue environment and innate immune activation signals in the process of shaping T cell responses.

Germinai Center Reactions: Sites of Antibody Affinity Maturation and Class Switch Recombination

Low-molecular-weight antigens that enter the subcapsular sinus are directly accessible to B cells. Particulate antigens are captured by subcapsular macrophages and transferred to B cells, which then mediate their transport to follicular dendritic cells in a nonantigen specific process requiring complement receptors. B cell activation by antigens in conjunction with co-stimulation via complement receptor or other forms of innate immune stimulation can lead to immediate proliferation of the B cells and the formation of plasma cells producing IgM antibodies from daughter cells. Specific T cell help promotes the formation of germinal centers within the B cell zone of any secondary lymphoid tissues. Germinal centers are roughly spherical collections of hundreds of antigen-specific B cells that undergo interaction with follicular T helper cells in the light zone and proliferation in the dark zone (Figure 19-6). The light zone is populated by stromal follicular DCs, which unlike conventional DCs are of nonhematopoietic origin, and follicular helper T cells. Follicular DCs use complement receptors to hold immune complexes on their surface for sampling by antigen-specific B cells. The follicular helper T cells provide help to B cells that maintain or increase their affinity for antigens and can also provide cytokine signals to promote class switching. Intravitral microscopy studies demonstrate that germinal centers are dynamic open structures in which antigen-specific B cells are continuously in motion and follicular DCs are accessible to interaction with B cells having diverse receptor specificity. Thus antigen-specific B cells can be recruited into germinal center reactions at any time in the process and can compete openly with B cells that were present earlier. Interactions between follicular helper T cells and centrocytes in the light zone are also highly dynamic and depend on the SAP adapter protein. T cells help control interzonal migration and affinity maturation.
of inflammatory cytokines plays a key role in this process. This normal process can also be recapitulated in the adult because chronic inflammation induces production of TNF and lymphotoxin in normal tissues, leading to the induction of stromal cells to form organized follicles and T cell zones within the inflamed tissues. Induction of stromal cells to produce CCL19/21 and/or CXCL13 is probably important in this process because transgenic expression of CXCL13 in ectopic locations in mice leads to formation of fully developed B cell follicles in these tissues.99 Thus B cells may have a particular capacity to induce tertiary lymphoid tissues. The efficacy of anti-TNF therapies in autoimmune disease may relate to the suppression of such tissues.100

Tertiary lymphoid tissues are often associated with autoimmune diseases including rheumatoid arthritis and the local infiltration of naïve T cells in an area with high concentrations of tissue-specific self-antigens, and innate stimulation may promote progression of disease by recruiting new T and B cell specificities into the autoimmune process.101 Breaking this cycle may be a key target of anticytokine and anti-B cell therapies that have been remarkably effective and are discussed in later chapters.

FOUR MAJOR TYPES OF EFFECTOR T CELLS

CD8+ effector T cells are generally cytotoxic and involved in killing of infected cells. CD8+ T cell responses can be initiated without CD4+ T cells’ help, but CD4 T cell help is required for effective CD8+ T cell memory.102 Important transcription factors for effector CD8 T cell differentiation are Tbet and eomesodermin.103 Three major types of effector CD4+ T cells are recognized. The first defined axis for CD4 T cell differentiation was between IFN-γ-producing T cells expressing the master transcription factor Tbet and IL-4-producing T cells using the master transcription factor GATA3.104 The IFN-γ-producing cells are referred to as Th1 cells, which help inflammatory cytotoxic T cell and macrophage responses, and the IL-4-producing cells, called Th2, help antiparasitic B cell responses leading to IgE production and eosinophil infiltration. The major cytokines that initiate Th1 responses are IFN-γ and IL-12 in the absence of IL-4, whereas IL-4 is the major cytokine that initiates the Th2 program. Recently it has been appreciated that a third type of effector T cell produces IL-17 and is important in many inflammatory diseases. Differentiation of these Th17 cells requires the nuclear hormone receptor RORγt.105 The cytokine conditions that lead to development of Th17 cells are transforming growth factor (TGF)-β plus IL-6, with IL-23 for maintenance.106 The role of TGF-β places Th17 cells on the same axis with induced regulatory T cells, with the deciding factor being the presence of IL-6 or IL-1β to favor Th17 over Treg induction.107 These are highly proinflammatory T cells that trigger recruitment of neutrophils to combat extracellular bacterial infections. Recently, the generation of gut-associated Th17 cells in mice has been linked to a single species of segmented filamentous bacteria.108 The presence of the commensal bacteria improves the clearance of intestinal pathogenic bacteria that require a Th17 response, but it also leaves the host more susceptible to autoimmune arthritis.109

Because most somatic mutations destroy the BCR or lower its affinity, there is a large amount of apoptosis in the germinal center in addition to proliferation to provide substrates for mutations.

TERTIARY LYMPHOID TISSUES: GENERATED AT SITES OF CHRONIC INFLAMMATION

Tertiary lymphoid tissues resemble secondary lymphoid tissues in many respects but are formed in the adult in response to chronic inflammation in locations where such tissues do not exist in steady-state conditions. The induction of tertiary lymphoid tissues can be compared with the formation of lymph nodes during normal development. Normal lymph node development involves the colonization of connective tissues in characteristic vascular nexuses by RORγt-dependent, CD4+, lymphotoxin-positive lymphoid tissue inducer cells.110 These cells use surface lymphotoxin and TNF to induce local stromal cells to express ICAM-1 and VCAM-1 and produce CCL19/21 or CXCL13, leading to development of lymph node stroma with a reticular fiber conduit system that is integrated into the developing lymphatic vessel system. Thus steady-state presentation of antigen to lymphoid tissue inducer cells.

Figure 19-6  Germinal center in living mouse lymph node. A germinal center was induced by vaccination after introduction of fluorescent antigen-specific B cells and antigen. Follicular dendritic cells were labeled with FDC-M1 antibody conjugated to a red fluorescent dye. Reticular fibers, which serve as antigen conduits into the follicle, are labeled blue. The germinal center (GC) is outlined and divided into a dark zone (DZ) containing aggregated centroblasts (green) and a light zone (LZ) containing loosely aggregated centrocytes, follicular dendritic cells (FDC-M1+), and follicular T cells (not shown). The surrounding follicle is densely packed with bystander B cells (not shown), which often enter and traverse the LZ. Green cells in the medullary region are plasma cells produced in the germinal center reaction. (From Schwickert TA, Lindquist RL, Shakhar G, et al: In vivo imaging of germinal centres reveals a dynamic open structure, Nature 446:83-87, 2007.)
SUMMARY

Adaptive immune responses have a degree of flexibility that is unparalleled in molecular recognition systems. This flexibility has a dangerous side that is constrained by anatomy and strong coupling to innate immunity. Autoimmune disease may result when weak points in tolerance mechanisms intersect with infection or tissue damage, leading the adaptive system to attack its own organs in a self-amplifying process of tissue destruction, inflammation, and inappropriate anatomic adaptation such as tertiary lymphoid tissue genesis. These processes are highly relevant to rheumatoid diseases. For example, a mouse model for autoimmune rheumatic disease called the KRN transgenic mouse develops a T cell– and B cell–dependent joint disease. Strategies to break these cycles by attacking innate or adaptive immune system components nonspecifically have met with success, but more specific strategies may reduce negative consequences of general immunosuppression.

References


The references for this chapter can also be found on www.expertconsult.com.
Autoimmunity ranges from physiologic autoreactivity to overt autoimmune disease, in part reflecting the complexity of the immune system, the presence of multiple layers of tolerance mechanisms, and genetic heterogeneity.

Autoimmune diseases can be classified by extent of organ involvement (organ-specific to systemic), innate immune system requirements, and effector mechanisms, but each type of autoimmune disease has unique pathophysiologic characteristics.

Advances in defining the mechanisms underlying autoimmune diseases have been greatly facilitated by delineation of the innate and adaptive immune systems.

Susceptibility to autoimmune diseases is multifactorial involving genetic, environmental, gender, and other factors, with genetic predisposition usually playing a central role. Their contributions are typically heterogeneous, partial, and additive.

Animal models are critical for understanding the pathophysiology of autoimmune diseases, providing fundamental insights into genetic, mechanistic, and pathologic processes. Spontaneous animal models for most autoimmune diseases do not exist.

Progress in delineating pathways revealed many key players in autoimmune diseases that are of potential therapeutic relevance.

Autoimmune diseases represent a significant health burden for 3% to 9% of the general population, and rheumatology, perhaps more than any medical subspecialty, encompasses a broad array of such diseases involving a wide range of organ systems (Table 20-1). Rheumatologists consequently have a substantial interest in defining the causes and pathophysiologic processes related to autoimmunity and in applying this information to the clinic.

The immune system must effectively defend against a diverse universe of pathogens while simultaneously maintaining tolerance to self-antigens. Recent advances have begun to clarify how this equilibrium is established and maintained and, importantly, have identified many critical factors and processes involved in the pathophysiology of autoimmune diseases. This chapter seeks to provide a general overview of autoimmunity covering the definition of the term, general tolerance mechanisms for T and B lymphocytes, theories of how tolerance can be breached, and ways in which genetic and environmental factors have been implicated to break tolerance and produce disease. Emphasis is placed on manifestations and mechanisms related to rheumatologic pathoses.

**KEY POINTS**

- Autoimmunity ranges from physiologic autoreactivity to overt autoimmune disease, in part reflecting the complexity of the immune system, the presence of multiple layers of tolerance mechanisms, and genetic heterogeneity.
- Autoimmune diseases can be classified by extent of organ involvement (organ-specific to systemic), innate immune system requirements, and effector mechanisms, but each type of autoimmune disease has unique pathophysiologic characteristics.
- Advances in defining the mechanisms underlying autoimmune diseases have been greatly facilitated by delineation of the innate and adaptive immune systems.
- Susceptibility to autoimmune diseases is multifactorial involving genetic, environmental, gender, and other factors, with genetic predisposition usually playing a central role. Their contributions are typically heterogeneous, partial, and additive.
- Animal models are critical for understanding the pathophysiology of autoimmune diseases, providing fundamental insights into genetic, mechanistic, and pathologic processes. Spontaneous animal models for most autoimmune diseases do not exist.
- Progress in delineating pathways revealed many key players in autoimmune diseases that are of potential therapeutic relevance.

**DEFINITION AND CLASSIFICATION OF PATHOGENIC AUTOIMMUNITY**

Autoimmunity, the immune response against self, evokes the specter of “horror autotoxicus,” a term coined by Paul Ehrlich at the turn of the 20th century to depict the perceived disastrous consequences of this condition. In fact, autoreactivity is more nuanced, ranging from a low “physiologic” level of self-reactivity that plays an essential role in lymphocyte selection and maintenance of normal immune system homeostasis, to an intermediate level of autoimmunity including autoantibodies and tissue infiltrates unassociated with clinical consequences, to pathogenic autoimmunity associated with immune-mediated dysfunction or injury. From the clinical perspective, it is the transformation to pathogenic autoimmunity that demarcates significant from insignificant self-reactivity.

The diagnosis of an autoimmune disease is generally based on the presence of adaptive immune system–mediated pathology caused by self-reactive antibodies or T cells. For many common autoimmune diseases, more definitive evidence for an autoimmune etiology has come from studies showing transfer of disease by autoantibodies or self-reactive T cells and animal models exhibiting congruent characteristics. There are, however, no universally accepted criteria and some less well-characterized diseases, currently considered to be autoimmune, may turn out to have other causes.

An example of diseases that exhibit some characteristics of autoimmunity yet are distinct in their pathogenesis are the so-called autoinflammatory syndromes. These are mostly rare monogenic disorders typified by intermittent bouts of fevers, rash, serositis, and arthritis caused by defective control of basic inflammatory mechanisms. Included are familial Mediterranean fever, the cryopyrinopathies, hyperimmunoglobulinemia D with recurrent fever, familial cold urticaria, and Blau syndrome. These disorders could be considered one end of a broader definition of autoimmunity, but because their pathophysiology is mediated entirely through the innate arm of the immune system, they are currently classified as a separate entity. Nonetheless, the possibility of a less stringent demarcation has been suggested by disorders such as Behçet’s syndrome, systemic juvenile rheumatoid arthritis, and Crohn’s disease, which appear to manifest both autoinflammatory and autoimmune features.

Autoimmune diseases are classified as systemic or organ-specific depending on the extent of their clinicopathology (Table 20-2). The systemic category includes systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), scleroderma, primary Sjögren’s syndrome, dermatomyositis, and systemic vasculitis. In this case, autoimmunity targets

---

**Table 20-1**

<table>
<thead>
<tr>
<th>Autoimmune Diseases</th>
<th>Organs Implicated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic lupus erythematosus (SLE)</td>
<td>Skin, joints, kidneys, central nervous system</td>
</tr>
<tr>
<td>Rheumatoid arthritis (RA)</td>
<td>Joints, lungs, eyes, pericardium</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>Skin, joints, lungs, kidneys</td>
</tr>
<tr>
<td>Primary Sjögren’s syndrome</td>
<td>Salivary and lacrimal glands, respiratory tract, joints</td>
</tr>
<tr>
<td>Dermatomyositis</td>
<td>Skin, muscles, joints</td>
</tr>
<tr>
<td>Systemic vasculitis</td>
<td>Skin, joints, kidneys, nervous system</td>
</tr>
</tbody>
</table>

---

**Table 20-2**

<table>
<thead>
<tr>
<th>Autoimmune Diseases</th>
<th>Organ Implicated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic lupus erythematosus (SLE)</td>
<td>Skin, joints, kidneys, central nervous system</td>
</tr>
<tr>
<td>Rheumatoid arthritis (RA)</td>
<td>Joints, lungs, eyes, pericardium</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>Skin, joints, lungs, kidneys</td>
</tr>
<tr>
<td>Primary Sjögren’s syndrome</td>
<td>Salivary and lacrimal glands, respiratory tract, joints</td>
</tr>
<tr>
<td>Dermatomyositis</td>
<td>Skin, muscles, joints</td>
</tr>
<tr>
<td>Systemic vasculitis</td>
<td>Skin, joints, kidneys, nervous system</td>
</tr>
</tbody>
</table>

---

**Table 20-3**

<table>
<thead>
<tr>
<th>Autoimmune Diseases</th>
<th>Pathophysiologic Mechanisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic lupus erythematosus (SLE)</td>
<td>Autoantibodies, immune complex deposition</td>
</tr>
<tr>
<td>Rheumatoid arthritis (RA)</td>
<td>Aberrant B cell activation, autoantibodies</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>Abnormal T cell responses, autoantibodies</td>
</tr>
<tr>
<td>Primary Sjögren’s syndrome</td>
<td>Autoantibodies, lymphocyte infiltration</td>
</tr>
<tr>
<td>Dermatomyositis</td>
<td>Autoantibodies, immune complex deposition</td>
</tr>
<tr>
<td>Systemic vasculitis</td>
<td>Autoantibodies, immune complex deposition</td>
</tr>
</tbody>
</table>

---

**Table 20-4**

<table>
<thead>
<tr>
<th>Autoimmune Diseases</th>
<th>Genetic Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic lupus erythematosus (SLE)</td>
<td>MHC, complement, immunoglobulin genes</td>
</tr>
<tr>
<td>Rheumatoid arthritis (RA)</td>
<td>MHC, IL-10, IL-6</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>MHC, TGF-β</td>
</tr>
<tr>
<td>Primary Sjögren’s syndrome</td>
<td>MHC, IL-10</td>
</tr>
<tr>
<td>Dermatomyositis</td>
<td>MHC, TGF-β</td>
</tr>
<tr>
<td>Systemic vasculitis</td>
<td>MHC, IL-10</td>
</tr>
</tbody>
</table>

---

**Table 20-5**

<table>
<thead>
<tr>
<th>Autoimmune Diseases</th>
<th>Environmental Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic lupus erythematosus (SLE)</td>
<td>UV exposure, infections</td>
</tr>
<tr>
<td>Rheumatoid arthritis (RA)</td>
<td>Tobacco, obesity</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>Smoking, infections</td>
</tr>
<tr>
<td>Primary Sjögren’s syndrome</td>
<td>Smoking</td>
</tr>
<tr>
<td>Dermatomyositis</td>
<td>Smoking, UV exposure</td>
</tr>
<tr>
<td>Systemic vasculitis</td>
<td>Smoking, infections</td>
</tr>
</tbody>
</table>

---

**Table 20-6**

<table>
<thead>
<tr>
<th>Autoimmune Diseases</th>
<th>Gender Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic lupus erythematosus (SLE)</td>
<td>Female prevalence</td>
</tr>
<tr>
<td>Rheumatoid arthritis (RA)</td>
<td>Female prevalence</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>Female prevalence</td>
</tr>
<tr>
<td>Primary Sjögren’s syndrome</td>
<td>Female prevalence</td>
</tr>
<tr>
<td>Dermatomyositis</td>
<td>Female prevalence</td>
</tr>
<tr>
<td>Systemic vasculitis</td>
<td>Female prevalence</td>
</tr>
</tbody>
</table>
ubiquitously expressed self-antigens and end-organ injury is typically mediated by autoantibodies and, less commonly, T cells. Contrastingly, in organ-specific diseases the self-antigens are typically cell or tissue specific in location or accessibility and end-organ damage can be mediated by antibodies and/or T cells. Some of the more notable examples in this group, which span virtually all organ systems, include Hashimoto’s thyroiditis, Graves’ disease, multiple sclerosis (MS), type 1 diabetes mellitus (T1DM), antiphospholipid syndrome (APS), pemphigus vulgaris, autoimmune hemolytic anemia, idiopathic thrombocytopenic purpura, and myasthenia gravis. It should be noted, however, that although the distinction of systemic and organ-specific disorders provides a conceptual framework, the pathophysiologies of autoimmune diseases are more diverse than might be implied by this simple classification. Autoimmune diseases can also be classified by hypersensitivity reaction type based on the mechanism of adaptive immune system–mediated injury. See later discussion in the chapter.

**ANIMAL MODELS OF AUTOIMMUNITY**

Much of what is known about the immune system and autoimmunity has been derived from studies in animals, particularly the mouse, which has an immune system and genome composition similar to humans. There are many well-characterized autoimmune models, and investigators can manipulate their genomes and immune systems, test interventions, and modify their environments. Animal models of autoimmune disease comprise three main types on the basis of their derivation: (1) spontaneous, (2) genetically modified, and (3) induced.

### Table 20-1: Examples of Rheumatologic Autoimmune Diseases

<table>
<thead>
<tr>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Juvenile rheumatoid arthritis</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>Neonatal lupus</td>
</tr>
<tr>
<td>Systemic sclerosis</td>
</tr>
<tr>
<td>CREST syndrome</td>
</tr>
<tr>
<td>Mixed connective tissue disease</td>
</tr>
<tr>
<td>Antiphospholipid syndrome</td>
</tr>
<tr>
<td>Vasculitis</td>
</tr>
<tr>
<td>Giant cell arteritis/polyartery rheumatica</td>
</tr>
<tr>
<td>Takayasu arteritis</td>
</tr>
<tr>
<td>Granulomatosis with polyangiitis</td>
</tr>
<tr>
<td>Churg-Strauss syndrome</td>
</tr>
<tr>
<td>Polyarteritis nodosa</td>
</tr>
<tr>
<td>Microscopic polyangiitis</td>
</tr>
<tr>
<td>Polymyositis/dermatomyositis</td>
</tr>
<tr>
<td>Relapsing polyarthritis</td>
</tr>
<tr>
<td>Sjögren’s syndrome</td>
</tr>
<tr>
<td>Behçet’s disease</td>
</tr>
<tr>
<td>Kawasaki’s disease</td>
</tr>
<tr>
<td>Sarcoidosis</td>
</tr>
</tbody>
</table>

**CREST, calcinosis, Raynaud’s phenomenon, esophageal dysfunction, sclerodactyly, and telangiectasia.**

The induced models encompass a wide variety of both systemic and organ-specific diseases. More commonly studied models of systemic disease include tetramethylpentadecane (TMPD, also called `pristane`)–induced autoimmunity, mercury-induced autoimmunity, and chronic graft-versus-host disease. All bear similarities to human SLE in producing antinuclear antibodies and immune complex deposits, but they differ in pathophysiology and strain susceptibility. The induced models of organ-specific diseases, a common approach is to immunize rodents with a self-antigen or closely related peptide or foreign antigen, plus a strong adjuvant, usually complete Freund’s. This approach makes it possible to induce autoimmunity in virtually all organ systems and to produce diseases mediated by cellular and humoral mechanisms. Some of the more
Table 20-2  Classification, Mechanisms, and Models of Autoimmune Diseases

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>AutoAg</th>
<th>Consequence</th>
<th>Hypersensitivity Type</th>
<th>Animal Model (Example)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Systemic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antiphospholipid syndrome</td>
<td>β2-GP1 (apoH)</td>
<td>Vascular thrombosis, recurrent miscarriages</td>
<td>II</td>
<td>(NZW × BXSB)F1</td>
</tr>
<tr>
<td>Microscopic angitis</td>
<td>p-ANCA (MPO)</td>
<td>Glomerulonephritis, leukocytoclastic vasculitis, mononeuritis multiplex, lung inflammation</td>
<td>III</td>
<td>Anti-MPO</td>
</tr>
<tr>
<td>Granulomatos with polyangiitis</td>
<td>c-ANCA (PR3)</td>
<td>Vascularitis of kidney, upper airway, and lungs</td>
<td>III</td>
<td>None</td>
</tr>
<tr>
<td>Cryoglobulinemia</td>
<td>Unknown</td>
<td>Cutaneous vasculitis, glomerulonephritis, skin lesions, arthritis, CNS lupus, and others</td>
<td>III</td>
<td>MRL-FasLpr</td>
</tr>
<tr>
<td>SLE</td>
<td>Nucleic acids</td>
<td>Glomerulonephritis, skin lesions, arthritis, CNS lupus, and others</td>
<td>III</td>
<td>MRL-FasLpr, BWF1, BXSB, NZM2410</td>
</tr>
<tr>
<td>Systemic sclerosis</td>
<td>Unknown</td>
<td>Fibrosis of multiple organs</td>
<td>III</td>
<td>Tsk/+ mice, bleomycin-induced</td>
</tr>
<tr>
<td>RA†</td>
<td>RF IgG immune complexes; citrullinated proteins; other joint antigens</td>
<td>Arthritis and associated manifestations rheumatoid nodules, rheumatoid lung, Felty’s syndrome</td>
<td>III</td>
<td>CIA, PGIA, AA, SKG, K/BxN, BXD2 mice</td>
</tr>
</tbody>
</table>

**Organ-Specific**

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>AutoAg</th>
<th>Consequence</th>
<th>Hypersensitivity Type</th>
<th>Animal Model (Example)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graves’ disease</td>
<td>TSH receptor</td>
<td>Stimulation of receptor; hyperthyroidism</td>
<td>II</td>
<td>EAT</td>
</tr>
<tr>
<td>Myasthenia gravis</td>
<td>ACh receptor</td>
<td>Receptor blockade/modulation; neuromuscular paralysis</td>
<td>II</td>
<td>EAMG</td>
</tr>
<tr>
<td>Autoimmune hemolytic anemia</td>
<td>RBC surface Ag</td>
<td>C and FcyR-mediated cell destruction; anemia</td>
<td>II</td>
<td>NZB</td>
</tr>
<tr>
<td>Idiopathic thrombocytopenic purpura</td>
<td>Platelet integrin GpIIb/IIa</td>
<td>Thrombocytopenia; purpura, bleeding</td>
<td>II</td>
<td>(NZW x BXSB)F1</td>
</tr>
<tr>
<td>Goodpasture’s syndrome</td>
<td>Type IV collagen* and other basement membrane Ags (Dsg3)</td>
<td>Pulmonary hemorrhage, glomerulonephritis</td>
<td>II</td>
<td>Anti-CIV, antilaminin</td>
</tr>
<tr>
<td>Pemphigus vulgaris</td>
<td>Epidermal cadherin (Dsg3)</td>
<td>Bullous skin lesions</td>
<td>II</td>
<td>Anti-Dsg3</td>
</tr>
<tr>
<td>Neonatal lupus</td>
<td>Ro/La</td>
<td>Cutaneous LE, heart block</td>
<td>II</td>
<td>None</td>
</tr>
<tr>
<td>T1DM</td>
<td>Pancreatic β-cell antigens</td>
<td>Pancreatic islet inflammation, diabetes</td>
<td>IV</td>
<td>NOD, BB</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>CNS antigens; MBP, PLP, MOG in animal models</td>
<td>Progressive CNS inflammation and paralysis</td>
<td>IV</td>
<td>EAE, Théller’s virus infection</td>
</tr>
</tbody>
</table>

*Most likely hypersensitivity type.
†Both systemic and organ-specific.
AA, adjuvant arthritis; ACh, acetylcholine; ANCA, anti-neutrophil cytoplasmic antibody; anti-CIV, antitype IV collagen; anti-Dsg3, antidesmoglein 3; β2-GP1, β2-glycoprotein 1; CIA, collagen-induced arthritis; EAE, experimental autoimmune encephalomyelitis; EAMG, experimental autoimmune myasthenia gravis; EAT, experimental autoimmune thyroiditis; MBP, myelin basic protein; MOG, myelin oligodendrocyte glycoprotein; MPO, myeloperoxidase; PGIA, proteoglycan-induced arthritis; PLP, proteolipid protein; PR3, proteinase 3; RF, rheumatoid factor; TSH, thyroid stimulating hormone; other abbreviations are mouse strains.

**COMPOSITION OF THE INNATE AND ADAPTIVE IMMUNE SYSTEMS**

The immune system comprises two major arms, innate and adaptive. The former is classically the initiator of immune responses to pathogens because of its capacity, on the one hand, to be activated by diverse pathogen-associated molecular patterns (PAMPs) expressed by microbes and, on the other hand, to activate the adaptive immune system through the production of proinflammatory factors and the presentation of antigens to lymphocytes.10,12 The adaptive response adds more protection through clonal recognition of an almost limitless diversity of antigenic specificities, immunologic memory of previous antigenic encounters, and additional cellular and antibody-mediated mechanisms to eliminate or neutralize microbes and other pathogens. Both

commonly studied organ-specific models developed by this approach include collagen-induced arthritis (CIA), proteoglycan-induced arthritis (PGIA), and experimental autoimmune encephalomyelitis (EAE), but there are many others to thyroid, eye, gonad, nerves, neuromuscular junction (acetylcholine receptor), muscle, heart, adrenal gland, bladder, stomach, liver, inner ear, kidney, and prostate tissues. In certain susceptible strains of rats, a progressive inflammatory erosive arthritis called adjuvant arthritis can also be induced by intradermal injection of complete Freund’s adjuvant or by mineral oil components of Freund’s adjuvant such as TMPD. Other induced models of arthritis include streptococcal cell wall and antigen-induced arthritis.14 Recently, arthritis characterized by synovial hyperplasia and ankylosis was reported in HLA-DR4 transgenic mice immunized with human citrullinated fibrinogen.79
systems use a large variety of immune-related molecules, cell types, and interactions among the various components to provide layers of protection that contribute to overall resistance to commensal and invading organisms. Similarly, prevention of self-reactivity is also effected by multiple inhibitory or deleting processes that, in toto, afford layers of safeguards against autoimmunity. The innate and adaptive immune systems are highly interconnected with activating and inhibitory feedback mechanisms operating in both directions.

**Innate Immune System Composition**

The main functions of the innate immune system include (1) the clearance of foreign and potentially detrimental self-materials through the recruitment, activation, and effector activity of immune cells; (2) activation of the complement pathways, and (3) triggering of the adaptive immune response (see also Chapter 18). A key feature of the innate, and to some extent the adaptive, immune system is that it is activated by the engagement of pattern-recognition receptors (PRRs) that recognize common structures on pathogens and also a few self-molecules such as certain damage-associated molecular patterns (DAMPs) (e.g., heat shock proteins), released during cellular stress, injury, or death. PRRs can be classified by their location into intracellular, membrane-bound, and secreted types, and interactions among the various components to provide layers of protection that contribute to overall resistance to commensal and invading organisms. Similarly, prevention of self-reactivity is also effected by multiple inhibitory or deleting processes that, in toto, afford layers of safeguards against autoimmunity. The innate and adaptive immune systems are highly interconnected with activating and inhibitory feedback mechanisms operating in both directions.

**Table 20-3 Partial List of Autoimmune Disease Models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Autoimmune Disease</th>
<th>Species</th>
<th>Notable Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spontaneous Variant</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRL-Fas&lt;sup&gt;lpr&lt;/sup&gt;</td>
<td>SLE</td>
<td>Mouse</td>
<td>Fas mutation, defective apoptosis, lymphoproliferation, human counterpart ALPS</td>
</tr>
<tr>
<td>(NZB × NZW)&lt;sup&gt;F1&lt;/sup&gt;</td>
<td>SLE</td>
<td>Mouse</td>
<td>Female predominance</td>
</tr>
<tr>
<td>NZB</td>
<td>SLE, AlHA</td>
<td>Mouse</td>
<td>Anti-RBC</td>
</tr>
<tr>
<td>BXSB</td>
<td>SLE</td>
<td>Mouse</td>
<td>Lupus and inflammatory arthritis</td>
</tr>
<tr>
<td>(NZW × BXSB)&lt;sup&gt;F1&lt;/sup&gt;</td>
<td>SLE, APS, ITP</td>
<td>Mouse</td>
<td>Yaa mutation (duplication of X chromosome containing TLR7 gene on Y chromosome)</td>
</tr>
<tr>
<td>BXD2</td>
<td>SLE, RA</td>
<td>Mouse</td>
<td>Anticardiolipin, antiplatelet</td>
</tr>
<tr>
<td>SKG</td>
<td>RA</td>
<td>Mouse</td>
<td>ZAP-70 mutation</td>
</tr>
<tr>
<td>NOD</td>
<td>T1DM</td>
<td>Mouse</td>
<td>MHC (H-2&lt;sup&gt;D&lt;/sup&gt;) similar to T1DM-predisposing HLA</td>
</tr>
<tr>
<td>BB</td>
<td>T1DM</td>
<td>Rat</td>
<td>Lymphopenia</td>
</tr>
<tr>
<td><strong>Genetically Modified</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1q ko</td>
<td>SLE</td>
<td>Mouse</td>
<td>Defective clearance of apoptotic cells</td>
</tr>
<tr>
<td>FcγRIIIb ko</td>
<td>SLE</td>
<td>Mouse</td>
<td>Impair regulation of B cell and antigen-presenting cells</td>
</tr>
<tr>
<td>BAFF Tg</td>
<td>SLE</td>
<td>Mouse</td>
<td>Enhanced survival of B cells</td>
</tr>
<tr>
<td>TLR7 Tg</td>
<td>SLE</td>
<td>Mouse</td>
<td>Enhanced survival and activation of B cells and DCs</td>
</tr>
<tr>
<td>Sanroque mice (roquin gene)</td>
<td>SLE</td>
<td>Mouse</td>
<td>R&lt;sub&gt;c&lt;/sub&gt;H&lt;sub&gt;1&lt;/sub&gt; M199R mutation increases ICOS expression and Tfh cell expansion</td>
</tr>
<tr>
<td>miR-17-92 Tg</td>
<td>SLE</td>
<td>Mouse</td>
<td>miRNA-induced autoimmunity</td>
</tr>
<tr>
<td>K/BxN, TCR Tg</td>
<td>RA</td>
<td>Mouse</td>
<td>Arthritis mediated by anti-GPI</td>
</tr>
<tr>
<td>MBP-specific TCR Tg mice</td>
<td>MS</td>
<td>Mouse</td>
<td>Spontaneous autoimmune encephalomyelitis</td>
</tr>
<tr>
<td>Double Tg: anti-GP TCR and insulin promoter-GP (GP=LCMV glycoprotein)</td>
<td>T1DM</td>
<td>Mouse</td>
<td>Ignorant Tg cells are activated by infection with LCMV causing insulin and diabetes</td>
</tr>
<tr>
<td><strong>Induced</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPD (pristane)-induced autoimmunity</td>
<td>SLE</td>
<td>Mouse</td>
<td>IFN-ζ and TLR7 dependent</td>
</tr>
<tr>
<td>Hg-induced autoimmunity</td>
<td>SLE</td>
<td>Mouse, rat</td>
<td>IFN-γ dependent</td>
</tr>
<tr>
<td>Chronic graft-versus-host disease</td>
<td>SLE</td>
<td>Mouse, rat</td>
<td>Parent into F1</td>
</tr>
<tr>
<td>Collagen-induced arthritis</td>
<td>RA</td>
<td>Mouse, rat</td>
<td>Autoimmunity to type II collagen</td>
</tr>
<tr>
<td>PG-induced arthritis</td>
<td>RA</td>
<td>Mouse</td>
<td>Autoimmunity to proteoglycan</td>
</tr>
<tr>
<td>Adjuvant arthritis</td>
<td>RA</td>
<td>rat</td>
<td>Inflammatory arthritis induced by complete Freund’s adjuvant or mineral oil</td>
</tr>
<tr>
<td>EAE</td>
<td>MS</td>
<td>Mouse, rat</td>
<td>Autoimmunity to MBP, MOG, or PLP depending on MHC haplotype</td>
</tr>
<tr>
<td>EAT</td>
<td>Thyroiditis</td>
<td>Mouse, rat</td>
<td>Autoimmunity to thyroglobulin</td>
</tr>
<tr>
<td>EAMG</td>
<td>MG</td>
<td>rat</td>
<td>Autoimmunity to acetylcholine receptor</td>
</tr>
</tbody>
</table>

AIHA, autoimmune hemolytic anemia; ALPS, autoimmune proliferative syndrome; APS, antiphospholipid syndrome; EAE, experimental autoimmune encephalomyelitis; GPI, glucose-6-phosphate isomerase; ko, gene knockout; ITP, idiopathic thrombocytopenic purpura; MBP, myelin basic protein; MS, multiple sclerosis; mus, mouse; MOG, myelin oligodendrocyte glycoprotein; PLP, proteolipid protein; T1DM, type 1 diabetes mellitus; Tg, transgenic; TMPD, 2,6,10,14-tetramethylpentadecane.
The complement system comprises an immune surveillance system that interacts with, and promotes elimination of, modified host cells, cellular debris, and foreign entities (see also Chapter 23). This system is tightly interconnected with inflammatory, immune, and other pathways and because of this, often plays important roles in both autoimmune and autoimmune diseases.

Activation of the complement cascade occurs through the classical, alternative, and lectin pathways, which function in complement-mediated lysis (C5b-C9 attack complex), release of proinflammatory mediators (C3a, C5a), opsonization (C1q, mannose binding lectin), and B cell activation (iC3b, C3dg; activation of B cells via CR2). Complement is regulated by soluble factors (C1 esterase inhibitor, sMAP and MAP-1, C2 receptor inhibitor trispanning, factor H, factor H-like protein 1, and C4b-binding protein), as well as cell surface decay accelerators such as CR1 and DAF (CD55), inhibitors of the terminal complement complex (CD59, vitronectin, clusterin), carboxypeptidase-N, and CD46. Major functions are the killing and elimination of pathogens, initiation and amplification of inflammation, safe removal of apoptotic cells and material, interaction with TLR signaling and coagulation pathways, and in facilitating immune complex clearance, angiogenesis, and tissue regeneration. Importantly, complement enhances and modulates humoral and cellular immune responses by several mechanisms including C3dg opsonization of antigen that promotes both costimulation of B cells and 

Adaptive Immune System Composition

The main constituents that distinguish the adaptive from the innate immune response are the B and T lymphocytes, which express an expansive array of diverse antigen receptors (see also Chapter 19). In aggregate, the pool of lymphocytes is capable of recognizing virtually all antigens and, by the process of selection, lymphocytes with beneficial specificities can be expanded and retained. These properties allow the immune system to detect a manifold galaxy of foreign antigens and to respond rapidly and specifically to secondary challenges. T and B lymphocyte antigen receptors are created by the random recombination of a large number of variable region (V-D-J) gene segments and diversity-producing end-joining mechanisms in primary lymphoid organs. Furthermore, peripheral B cells can also increase specificity by somatic hypermutation and selection of higher affinity clones. The clonal nature of lymphocytes, bestowed by uniquely rearranged antigen receptors, is key to the exquisite discriminating power of the immune system. Clonality also extends to other characteristics such as cytokine production, surface receptors, and cell signaling molecules.

T cells play a central role in the adaptive immune response through elaboration of cytokines such as IFN-γ and TNF, IL-4, and IL-17, by providing essential signals to B cells in T cell–dependent humoral responses, as direct killers of cells, and by promoting inflammation in target
TOLERANCE MECHANISMS

Increasing insights into mechanisms of self-nonself discrimination have emerged over the past 6 decades in parallel with growing knowledge of the immune system. More than 50 years ago, Burnet and Medawar advanced the critical concept that tolerance was imposed by clonal deletion of self-reactive lymphocytes during early development (i.e., central tolerance).

Later, with the discovery that mature B cells undergo somatic hypermutation in the periphery, it was hypothesized by Bretscher and Cohn that the production of autoantibodies might be impeded by the need for both B and T cell compartments to breach tolerance. In 1975, while studying allogeneic responses, Lafferty and Cunningham posited that activation of T cells involved the passing of a second signal that need not be antigen-related, thereby implicating costimulation from antigen-presenting cells as a critical factor in lymphocyte activation.

In 1987, the nature of the costimulation, or two-signal, requirement was further defined when Jenkins and Schwartz showed that engagement of antigen receptor alone without a second signal resulted in functional inactivation of T cells. A novel mechanism for self-nonself discrimination that incorporated the innate immune system was then advanced by Janeway in 1989 when he hypothesized that antigen-presenting cells critical for T cell activation remain quiescent unless activated by the engagement of PRRs by microbial products. This concept was further extended by Matzinger in 1994 to a “danger model” that includes activation of the innate immune system by both foreign and endogenous factors associated with tissue stress and damage. These models have laid the foundation for the current, more complex, view of self-nonself discrimination in which tolerance is imposed by both innate and adaptive immune systems through layers of mechanisms occurring at various stages throughout lymphocyte development and activation (Table 20-4).

Clone-Specific Self-Nonself Recognition

In contrast to innate immune cells, which are activated primarily by hardwired microbial PRRs, lymphocytes are unrestricted in specificity and therefore self-nonself discrimination must be implemented at the clonal level. To achieve this, T and B cells use several mechanisms that can be grouped into three general strategies. First, the type of response is controlled by developmental stage. For example, immature lymphocytes respond to strong antigen receptor stimulation by cell death, whereas a similar signal in mature cells leads to activation. Through this mechanism self-reactive clones are eliminated from the nascent lymphocyte repertoire before they can cause injury. Second, activation of mature lymphocytes requires—in addition to antigen receptor engagement—a second co-stimulatory signal that, if absent, results in anergy or cell death. For the most part, this requirement limits reactivity to self because co-stimulatory signals are largely provided by activated cells of the innate immune system. Third, lymphocytes are fine-tuned in various ways by a fairly extensive list of modulating factors, which is necessary for controlling self-reactive clones.

For example, defects affecting a broad range of surface receptors on lymphocytes, including those with pro-survival (IL-7R, BAFFR, IL-2R); proapoptotic (TNFRs, FasL, TRAIL); costimulatory (CD28, CD40, TLRs); differentiating (IL-12R, IL-4R, IFNγR, IL-23R, retinoic acid R, transforming growth factor (TGF)-βRs, SAP/SLAM family members, OX40, ICOS/ICOSL); inhibitory (FcγRIIb, CD22, CTLA4, PD-1); antigen receptor signal modulating (CD19, CD45); and activating (SAP/SLAM family members) activities have been shown to influence the development of autoimmunity. Collectively, these self-nonself recognition mechanisms provide the basic cellular means by which the innate and adaptive immune systems render T and B cell clonotypes tolerant to self-antigens and resistant to autoimmune disease.

Innate System and Tolerance

Given its vital role in initiating and modulating adaptive immune responses, it is not surprising that the innate immune system strongly influences both tolerance and autoimmunity. Indeed, although its contributions have yet to be fully explicated, several ways in which self-tolerance is influenced by the innate arm have been defined. First, as noted previously, activation of the innate immune system under normal circumstances typically requires engagement of microbial PRRs, which endows the immune system with a direct and simple way to distinguish foreign-
self-antigens. The importance of this mechanism is illustrated by the finding that overexpression of TLR7 by spontaneous gene duplication or transgenic approaches promotes systemic autoimmunity. This occurs because certain PRRs, such as TLR7 and TLR9 that equally sense both foreign and self-nucleic acids, avoid significant exposure to endogenous nucleic acids by virtue of their location in subcellular compartments. However, in the case of TLR7, overexpression makes it possible for normally substimulatory amounts of endogenous RNA to activate immune cells, thereby bypassing the usual requirement for microbial exposure. Second, some cells of the innate immune system actively suppress adaptive immune system activation under certain conditions. For example, immature and, under some circumstances, even mature dendritic cells have been shown to promote tolerance by inducing CD4 T regulatory (Treg) cells and other mechanisms. Third, another critical function of the innate immune system is the rapid noninflammatory clearance of apoptotic cells. Failure can result in an increased supply of self-antigenic material including nucleic acids, secondary necrosis, and release of proinflammatory factors, leading to systemic autoimmunity. Accordingly, deficiencies in several key apoptotic cell clearance molecules are associated with autoimmunity including (1) the Tyro3-Axl-Mer receptors on phagocytes that bind through Gas6 or protein S, the exposed phosphatidylserine (PS) on apoptotic cells; (2) the milk fat globule-EGF factor 8 (MFG-E8) protein that bridges the αvβ3 integrin on phagocytes and PS on apoptotic cells; and (3) natural IgM or C1q that binds to and enhances clearance of apoptotic cells.

Fourth, several complement components have been directly implicated in autoimmunity. For example, SLE is associated with deficiencies of proximal components of the classical pathway including C1q, C4, and C2. The mechanism for this is not certain, but both defective clearance of apoptotic material/immune complexes and a shift in the activation threshold of lymphocytes have been suggested. As another example, deficiency of CD55 (or decay-accelerating factor), a cell surface protein that restricts complement activation, is associated with enhanced T cell responses and exacerbation of neuroinflammation and lupus in animal models. Thus at many levels, the innate immune system plays a critical role in maintaining tolerance and controlling autoimmunity.

**T Cell Tolerance**

T cells are critical players in both achieving and fine-tuning tolerance to a high degree of specificity. Researchers have identified several mechanisms and divided them into three main areas: central tolerance wherein T cells first acquire their antigen receptor, peripheral tolerance wherein T cells encounter self-antigens not present in the thymus, and post-activation regulation wherein activated and expanded T cell clones are returned to their resting state. Central tolerance, as alluded to earlier, is imposed on T cells with

---

**Table 20-4  Multiple Tiers of Tolerance**

<table>
<thead>
<tr>
<th>Type</th>
<th>Cell Type</th>
<th>Site</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central Compartment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central tolerance</td>
<td>T cells</td>
<td>Thymus</td>
<td>Primarily deletion, anergy, possibly editing</td>
</tr>
<tr>
<td></td>
<td>B cells</td>
<td>Bone marrow</td>
<td>Editing, anergy, deletion</td>
</tr>
<tr>
<td>Peripheral Compartment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immature B cell tolerance</td>
<td>Transitional 1 (T1) B cells</td>
<td>Periphery</td>
<td>Deletion, anergy upon activation</td>
</tr>
<tr>
<td>Peripheral anergy</td>
<td>T and B cells</td>
<td>Secondary lymphoid organs and peripheral tissue</td>
<td>Inadequate signal induces cell inactivation</td>
</tr>
<tr>
<td>Ignorance</td>
<td>T cells; maybe B cells</td>
<td>Peripheral and secondary lymphoid organs</td>
<td>Insufficient self-antigen or co-stimulation</td>
</tr>
<tr>
<td>Inaccessible self-antigen</td>
<td>T and B cells</td>
<td>Secondary lymphoid organs and site of inflammation</td>
<td>Sequestration, crypticity</td>
</tr>
<tr>
<td>Regulation</td>
<td>T and B cells</td>
<td>Site of inflammation and secondary lymphoid organs</td>
<td>Suppression by regulatory cells via intercellular signals and cytokines</td>
</tr>
<tr>
<td>Clonal deletion following</td>
<td>T and B cells</td>
<td>Site of inflammation and secondary lymphoid organs</td>
<td>Apoptosis caused by a decrease in survival factors</td>
</tr>
<tr>
<td>activation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytokine deviation</td>
<td>T cells</td>
<td>Site of inflammation and secondary lymphoid organs</td>
<td>Differentiation toward less pathogenic T subsets</td>
</tr>
<tr>
<td>Postsomatic hypermutation</td>
<td>B cells</td>
<td>Germinal center</td>
<td>Insufficient CD4 T cell help, deletion (via Fas)</td>
</tr>
<tr>
<td>Tissue resistance</td>
<td>B and T cells</td>
<td>Peripheral tissues</td>
<td>Inhibitory intercellular signals and cytokines</td>
</tr>
<tr>
<td>Innate Mechanisms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRR engagement required for activation</td>
<td>Innate cells</td>
<td>Site of inflammation</td>
<td>Simple mechanism for self-nonself discrimination</td>
</tr>
<tr>
<td>Suppression of adaptive immune responses</td>
<td>Immature and mature DCs</td>
<td>Site of inflammation</td>
<td>Delivery of inhibitory signals and activation of Treg</td>
</tr>
<tr>
<td>Clearance of apoptotic cells</td>
<td>Complement, phagocytes</td>
<td>Secondary lymphoid organs and Peripheral tissues</td>
<td>Removal of potential proinflammatory material and self-antigens</td>
</tr>
<tr>
<td>Complement-mediated effects on adaptive responses</td>
<td>Lymphocytes, innate cells</td>
<td>Secondary lymphoid organs and peripheral tissue</td>
<td>Modulation of activation</td>
</tr>
</tbody>
</table>
The main purpose of this page is to discuss the mechanisms that maintain immune tolerance, particularly in the context of autoimmunity. It explains how T cells, which are crucial for immune responses, are regulated to prevent autoimmunity.

In the periphery, multiple mechanisms for avoiding autoreactivity have been identified. Among these, a common explanation is that most tissue-associated self-antigens are not accessible to trigger a response, a situation that can be caused by low abundance, specific characteristics of the self-antigen, or location. This mechanism is supported experimentally by the finding that T cells expressing a transgenic TCR to certain tissue-specific antigens are not deleted or activated, nor do they cause autoimmune disease.

Another peripheral mechanism is the aforementioned two-signal paradigm wherein T cell activation requires both TCR engagement and a co-stimulatory signal usually provided by CD28. Because the two ligands for CD28, CD80 and CD86, are primarily expressed at high levels on activated professional antigen-presenting cells, presentation of self-antigen by quiescent antigen-presenting cells would lead to tolerance. Indeed, immature DCs promote tolerance in this manner by constitutively presenting low doses of self-antigen on MHC, resulting in cell death or anergy of the corresponding T cells.

Peripheral tolerance of T cells is also maintained by active suppression via immunoregulatory cells of the immune system, among which CD4+ Tregs are the best characterized. They constitute a distinct αβ T cell subset generated in the thymus (natural Treg, nTreg) or in the periphery from naïve or mature CD4+ T cells exposed to TGF-β (induced Treg, iTreg). Both are developmentally induced by the Foxp3 (forkhead boxP3) transcription factor. They typically express high levels of the IL-2 receptor component, IL-2Ra (CD25), and require IL-2 for survival. Tregs participate in every adaptive immune response, are critical for maintaining the proper level of immune response, and are activated at the same time as conventional T cells. They are thought to suppress the magnitude of the response by (1) initially downregulating DC function and then inhibiting T cell activation by competing for IL-2, (2) producing immunosuppressive cytokines such as TGF-β, IL-10, and IL-35, and (3) by cell-cell inhibitory interactions that lead to cell killing or the induction of negative signals. These inhibitory actions suppress T cells that are in proximity to Treg cells regardless of their antigen specificity.

Researchers have described other T cells with regulatory activity in autoimmunity including Tr1, CD8 Treg, Qa-1/HLA-E-restricted CD8+ T cells, and γδ T cells, but they are less well characterized.

Tissues themselves also employ mechanisms that suppress self-reactivity and contribute to establishing immune privilege. These comprise three general categories. First, certain tissues are decorated with cell surface inhibitory molecules such as the proapoptotic FasL and TRAIL, lymphocyte inhibitory and Treg-promoting PD-L2, and complement regulatory proteins, CD55 and CD46, that can potentially eliminate or impede the activation of autoreactive T cells. Second, soluble inhibitors of inflammation and immune activation are secreted by particular tissues. Notably, in the aqueous humor of the eye, there is a broad spectrum of such factors that include TGF-β, α-melanocyte-stimulating hormone, vasointestinal peptide, calcitonin gene–related peptide, somatostatin, macrophage inhibitory factor, and complement inhibitors. Third, lymph node resident stromal cells have been shown to induce tolerance of CD8+ T cells recognizing peripheral tissue-restricted self-antigens. Thus, it has been proposed that stromal cells in lymph node and tissues may provide a means to eliminate T cells that bind to tissue-restricted antigens not expressed in the thymus. Fourth, the anterior chamber of the eye elicits a unique type of altered immune response through a complex multistep process termed anterior chamber-associated immune deviation (ACAID) that leads to a dampened and less tissue-destructive response. Although ACAID was long thought important for tolerance, it was recently argued that its primary function may be to modulate the immune response so that the eye can respond to infection without damaging its integrity.

Another possible peripheral mechanism is immune deviation wherein polarization away from a predisposing cytokine pattern, such as from a Th1 to a Th2 profile, suppresses the development of autoimmunity. Similarly, activation of NKT cells with α-GalCer, which induces IFN-γ production, is associated with dampening of the adaptive Th1 and Th17 effector responses and protection from experimental autoimmune uveitis. In these models, autoreactive T cells are activated but do not produce the proinflammatory factors necessary for tissue damage.

In addition to central and peripheral tolerance, the immune system must also avert autoimmunity by suppressing or eliminating T cells following their activation or expansion. This regulation is mediated by several processes, including upregulation of inhibitory receptors such as CTLA4, expression of proapoptotic receptors like Fas, and release of intracellular proapoptotic factors such as Bim. Deficiencies of such mediators that control the magnitude of T cell responses are associated with severe expansion of lymphocyte populations and with varying degrees of autoimmunity.
B Cell Tolerance

B cells are required not only for antibody production but also serve as potent antigen-presenting cells for T cells and follicular DCs, and can act in regulation as well. \(^1\) Moreover, depletion of B cells with rituximab has shown promise even in autoimmune diseases considered to be mediated by T cells such as T1DM \(^1\) and multiple sclerosis. \(^1\) Therefore, there is substantial interest in defining both mechanisms of B cell tolerance and the specific role of B cell tolerance in autoimmune disease.

Before discussing specific tolerance mechanisms, however, it should be emphasized that the fate of B cells after antigen receptor (BCR) cross-linking is highly dependent on developmental stage, context and strength of signal, and the nature of the antigen, probably more so than for T cells because B cells are subjected to less stringent selection during central tolerance. Several checkpoints important for controlling autoreactive B cells and for maintaining self-tolerance have been identified and include many central and peripheral mechanisms similar to those described for T cells, as well as a few additional ones.

Central tolerance of B cells takes place in the bone marrow during preB to immature B cell transition as they express rearranged immunoglobulin (Ig) genes on their surface. \(^1\) It appears that the dominant mechanism for B cells with high affinity to membrane-bound self-antigen is receptor editing (replacement of L-chain) and, to a lesser extent, deletion, while soluble self-antigens induce both receptor editing and anergy. \(^1\) Anergic B cells are detectable in the periphery as an IgD\(^-\)IgM\(^-\) population \(^1\) or in mice as splenic transitional 3 (T3) B cells. \(^1\) They are short-lived at least in part because they downregulate the BAFF receptor, which is required for their survival, putting them at a competitive disadvantage with other immature B cells, and they are less able to enter B cell follicles.

In the periphery, the earliest tolerance checkpoint occurs at the transitional 1 (T1) B stage over a 2-day interval before maturation to T2 and later naïve B cell subsets. \(^1\) T1 B cells are the immediate bone marrow emigrant population, retain an immature phenotype, and are BAFF-dependent for survival. Importantly, they undergo apoptosis and not activation when stimulated, which results in deletion of B cells recognizing peripheral self-antigens not expressed in the bone marrow compartment. Thus, this mechanism, which is unique to B cells, represents in essence an extension of central tolerance to the periphery.

Other peripheral tolerance mechanisms are achieved through many of those described earlier for T cells but differ qualitatively due to distinct differentiation pathways and differences in antigen recognition by B cells and T cells (i.e., BCRs can bind to virtually all tertiary structures while TCRs are restricted to recognizing self-MHC/peptide complexes on host cell surfaces). Accordingly, B cells can be ignorant of their corresponding self-antigen because of insufficient quantity or access \(^1\) or can undergo anergy and ultimately cell death if there is engagement of the BCR without co-stimulation (i.e., two signals). \(^1\)

Another notable checkpoint occurs during T cell–dependent immune responses as B cells undergo affinity maturation in germinal centers (GCs) and acquire new specificities, which may include self-reactivity. Evidence suggests that tolerance at this juncture is often defective in autoimmune diseases because most autoantibodies have acquired autoreactivity through somatic hypermutation and are class-switched, \(^1\) both indicative of GC maturation. Although recent studies have provided significant insights into GC processes involved in the selection of B cell clones with high affinity to foreign antigens, how tolerance of class-switched autoreactive B cells is achieved remains to be clarified. Nevertheless, the strongest evidence suggests (1) autoreactive B cells compete poorly for the cognate T cell help essential for GC B cell survival because the autoreactive BCR would bind less well to the original antigen, resulting in less internalization of antigen for processing and presentation to T cells \(^1\) and (2) B cells that acquire BCRs with high affinity to membrane antigens are deleted by a Fas-dependent mechanism. \(^1\)

THEORIES OF AUTOIMMUNITY

Development of autoimmune diseases is influenced by genetic and, to varying degrees, environmental, gender, and other factors, with current evidence supporting a model in which genetic predisposition is required (Figure 20-1). Therefore, theories of autoimmunity and loss of tolerance are closely intertwined with genetic influences. In addition, such theories must also explain how tolerance is breached when autoimmunity is induced in otherwise normal animals. Taking both these factors into account and applying a reductionist perspective, theories of autoimmune disease can be consolidated into two main mechanisms representing separate ends of a continuum—most diseases having some elements of both. On one end, loss of tolerance and the consequent autoimmune disease is caused by genetically imposed defects in central and/or peripheral tolerance mechanisms while, on the other end, autoimmunity arises from the conventional immune response to self-antigens for which tolerance is normally incompletely established (Table 20-5). In general, most systemic autoimmune diseases are caused by tolerance defects, whereas organ-specific diseases can be mediated by either mechanism.

**Figure 20-1**  Etiology of autoimmune diseases. Autoimmune diseases are usually caused by the additive effects of autoimmunity-promoting environmental, gender, and other factors superimposed on significant underlying genetic predisposition.
Table 20-5  Mechanisms of Autoimmunity

<table>
<thead>
<tr>
<th>Examples</th>
<th>Disease</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Defective Tolerance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Central Defects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIRE deficiency</td>
<td>APECED syndrome</td>
<td>Failure to delete autoreactive T cells because of reduced expression of peripheral antigens in thymus</td>
</tr>
<tr>
<td>ZAP-70 deficiency</td>
<td>Inflammatory erosive arthritis (mice)</td>
<td></td>
</tr>
<tr>
<td><strong>Peripheral Defects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAS/FASLG deficiency</td>
<td>Autoimmune lymphoproliferative syndrome (ALPS)</td>
<td>Defective apoptosis</td>
</tr>
<tr>
<td>Rc3h1 (M199R) mutation</td>
<td>Lupus (mice)</td>
<td>Increased ICOS on Tfh cells promotes their expansion</td>
</tr>
<tr>
<td>TREX1 (DNase III) deficiency</td>
<td>Aicardi-Goutières syndrome, chilblain lupus</td>
<td>Accumulation of intracellular DNA induces IFN-α production</td>
</tr>
<tr>
<td>FOXP3 deficiency</td>
<td>IPEX syndrome</td>
<td>Absence of Treg cells</td>
</tr>
<tr>
<td>PD-1 deficiency</td>
<td>Lupus, myocarditis (mice)</td>
<td>Defective peripheral tolerance of T cells</td>
</tr>
<tr>
<td><strong>Activation of Nontolerant Lymphocytes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penetrating injury</td>
<td>Sympathetic ophthalmia</td>
<td>Release of self-antigen in an inflammatory milieu</td>
</tr>
<tr>
<td>Coxsackie B virus infection</td>
<td>T1DM (mice)</td>
<td>Infection-mediated release of self-antigen in an inflammatory milieu</td>
</tr>
<tr>
<td>Immunization with self-antigen and strong adjuvant</td>
<td>EAE (mice)</td>
<td>Activation of ignorant T cells</td>
</tr>
<tr>
<td>Citrullination of proteins</td>
<td>RA</td>
<td>Generation of neo self-antigens</td>
</tr>
<tr>
<td>Altered structure of collagen IV caused by sulfilimine bonds</td>
<td>Goodpasture’s syndrome</td>
<td>Formation of conformational neo self-antigens</td>
</tr>
<tr>
<td>Cross-reactivity of group A streptococcal and cardiac antigens</td>
<td>Rheumatic fever</td>
<td>Molecular mimicry</td>
</tr>
<tr>
<td>Lymphopenia caused by disease associated IL-21 production</td>
<td>T1DM (NOD mice)</td>
<td>Lymphopenia-induced homeostatic proliferation</td>
</tr>
</tbody>
</table>

**Defective Tolerance**

Although one can infer that loss of tolerance underlies autoimmunity, the specific tolerance defects causing common autoimmune diseases have been difficult to delineate, presumably because of modest defects at multiple checkpoints. Nevertheless, studies of monogenic human autoimmune disease and animal models have identified a wide range of defects in the various layers of central and peripheral tolerance. Such defects are caused by diverse genetic abnormalities and are mediated by a variety of lymphoid and nonlymphoid cell types. A few representative examples follow.

Defects in central tolerance have long been suspected to cause autoimmunity because of its well-documented role in eliminating nascent self-reactive lymphocytes, but until recently solid evidence for this was lacking. A breakthrough came from the discovery that mutations in a transcription factor, AIRE (for autoimmune regulator), caused autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), a rare inherited disease associated with T cell and autoantibody-mediated autoimmune destruction of multiple endocrine organs.\(^\text{130,131}\) AIRE-deficient mice developed a similar syndrome, which was caused by reduced expression of thousands of peripheral tissue genes in the thymic medullary epithelium and consequently failure to eliminate T cells specific for those gene products. Thus it appears that the main function of AIRE is to prevent autoimmunity by deleting T cells recognizing peripheral tissue self-antigens.

Another example of altered thymic selection leading to autoimmunity, but in this instance caused by a defect in T cells, is the aforementioned SKG arthritis model. Here, a function-imparing mutation in the C-terminal SH2 domain of ZAP-70, a Syk family tyrosine kinase activated by the T cell receptor complex ζ-chain, reduces TCR signaling, leading to defective conventional T cell and nTreg development in the thymus, and presumably enhanced positive selection of autoreactive T cells.\(^\text{132}\) Interestingly, using several ZAP-70 mutants, researchers showed that slight differences in ZAP-70-mediated signaling strength affected arthritis susceptibility.\(^\text{132,133}\)

For B cells, major defects in central tolerance leading to autoimmunity have been more difficult to prove and no AIRE-like equivalent has been discovered. Nevertheless, defective central tolerance of B cells in autoimmunity has been suggested by the finding of a higher frequency of naïve mature B cells with self-reactivity in patients with SLE and RA.\(^\text{134-136}\)

Moving to the periphery, an example of defective peripheral tolerance promoting autoimmunity is FAS deficiency.\(^\text{137}\) FAS is a proapoptotic surface receptor that plays a critical role in maintaining immunologic homeostasis by eliminating undesired cells. Defects in FAS cause autoimmune lymphoproliferative syndrome (ALPS), also called Canale-Smith syndrome, and in mice, lymphoproliferative (lpr) disease,
both of which exhibit massive enlargement of secondary lymphoid organs—caused mainly by the accumulation of a normally rare so-called “double negative” subset of T cells that lack both CD4 and CD8 co-receptors—and a variety of autoimmune manifestations. Lpr mice are defective in eliminating B cells that acquire self-reactivity in the periphery, a process that typically occurs in the GC during somatic hypermutation. Similar abnormalities have been described in humans and mice with defects in the ligand for FAS, FASLG. Peripheral tolerance is also breached by overexpression of the costimulatory molecule ICOS on T follicular helper (Tfh) cells in Samroque mice. These mice have a point mutation in the Rc3h1 gene, a RING-type ubiquitin ligase, that impairs its ability to degrade ICOS mRNA. Increased expression of ICOS promotes the expansion of Tfh cells and germinal centers, production of IL-21, and autoimmune. An example of a lymphocyte-extrinsic cause for loss of tolerance and autoimmunity is deficiency in the 3′ repair exonuclease 1 (TREX1, DNase III). Loss-of-function mutations in TREX1 have been implicated in the Aicardi-Goutières syndrome, a rare progressive encephalopathy associated with elevated IFN-α levels in cerebrospinal fluid, and chilblain lupus, a rare form of SLE characterized by painful, bluish-red inflammatory skin lesions typically affecting areas exposed to cold. Autoimmunity is thought to be caused by the intracellular accumulation of ssDNA derived from endogenous retroelements normally degraded by TREX1, resulting in the activation of intracellular DNA sensors and consequent over production of IFN-α. The importance of DNA disposal in lupus has been further illustrated by the association of DNase I deficiency with development of lupus in humans and mice.

Defective regulation can also result in loss of tolerance and autoimmune disease as illustrated by the absence of Tregs. In humans, monogenic deficiency of the FOXP3 gene, which is required for Treg development, is associated with the IPEX (immune dysregulation, polyendocrinopathy, enteropathy X-linked) syndrome, a severe systemic autoimmune disease associated with diarrhea, eczematous dermatitis, and endocrinopathy, usually fatal within the first year. T1DM, autoimmune cytopenias, and nephritis are among other less common autoimmune manifestations of this syndrome. Similar findings are observed in scurfy mice that have a spontaneous function-impairing mutation of Foxp3.

**Autoimmunity Caused by Activation of Intolerant or Partially Tolerant T Cells**

Another theory is that autoimmunity develops through the conventional activation of self-reactive T cells that have not been deleted in the thymus and remain oblivious to the corresponding self-antigen after emigration. Such ignorant T cells, commonly found in the periphery of both normal humans and animals, can be activated by antigen presented by professional antigen-presenting cells in the context of an innate inflammatory milieu. Once activated, T cells can gain access to virtually all tissues and, when activated again locally, can elaborate proinflammatory factors causing tissue damage. Breach of tolerance by this mechanism is most often associated with organ-specific diseases, presumably because tissue-specific antigens are less likely to be expressed in the thymus. This theory is supported by the finding that type 1 diabetes is prevented in BB rats and NOD mice by intrathymic transplantation of pancreatic islets or inhibited in NOD mice by intrathymic administration of GAD, a major β-islet autoantigen in this model. Similarly, EAE can be prevented by intrathymic administration of either the immunizing antigen, myelin basic protein, or its major encephalitigenic epitope, and autoantibody production can be delayed in lupus-prone mice following intrathymic injections of polynucleosomes. Taken together, these data support the concept that autoimmunity can be caused by the activation of lymphocytes to self-antigens for which there was incomplete central tolerance. This is similar to the mechanism underlying the APECED syndrome caused by AIRE deficiency, but in this instance central tolerance is not known to be defective.

Breach of tolerance by ignorant lymphocytes has been shown to depend on many factors including the (1) nature of the self-antigen, (2) extent of exposure to antigen, (3) antigen receptor affinity, (4) frequency of autoreactive lymphocytes, (5) types and levels of co-stimulatory molecule expression, (6) cytokine and chemokine profiles, and (7) presence of inflammation. Autoimmunity is thought to be caused by the activation of lymphocytes to self-antigens beyond the initiating one. Under conditions favorable for antigen presentation and co-stimulation promotes epitope spreading and the activation and expansion of naïve autoreactive lymphocytes. This is similar to the mechanism underlying the APECED syndrome caused by AIRE deficiency, but in this instance central tolerance is not known to be defective.

Similar to immune responses in general, key factors for the initiation of autoimmunity are innate inflammatory and co-stimulatory factors that promote the initial activation and expansion of naïve autoreactive lymphocytes. This is thought to contribute to the autoimmune response by the release of self-antigens following cell damage or death, increased MHC/peptide expression, upregulation of co-stimulatory factors, and activation of professional antigen-presenting cells. Indeed, moderate to severe tissue necrosis is often associated with some evidence of autoreactivity, although this rarely progresses to autoimmune disease. In autoimmune diseases, chronic cell and tissue injury and the continual release of self-antigens under conditions favorable for antigen presentation and co-stimulation promotes epitope spreading and the activation of an expanding repertoire of lymphocytes recognizing autoantigens beyond the initiating one. This process is thought to account at least in part for the usual progressive course of autoimmunity. Overall, these and other findings support the theory that under certain conditions such as infection or trauma, self-antigens are released, which in the presence of inflammatory factors and an activated innate immune system, can lead to triggering and expansion of...
self-antigen-recognizing lymphocytes and autoimmune disease.

In addition to the release of sequestered self-antigens, there are several other ways in which the initial activation of nontolerant lymphocytes has been postulated to occur. Some investigators have suggested that the initial response may be directed toward certain cryptic determinants \(^{167,168}\) based on “crypticity,” the hierarchy of dominant and cryptic epitopes within a protein caused by differences in binding affinity to the MHC, protein processing, type of antigen-presenting cells, and the overall repertoire of epitope-specific T cells.\(^{166}\) It was therefore posited that during thymic selection, T cells engaging the few dominant epitopes are eliminated, whereas those recognizing the less antigenic, but more abundant, cryptic epitopes are spared. These T cells then emigrate to the periphery, where they can be activated by the corresponding cryptic epitope under certain conditions.

Another mechanism by which self-antigens can activate nontolerant lymphocytes is through the production of neo self-antigens following post translational or chemical modifications. A prominent example of this is the formation of citrullinated proteins caused by the deamination of arginine residues by peptidylarginine deaminase (PAD) enzymes. Several citrullinated proteins are not only major targets of autoantibodies in a subset of RA but are thought to play a significant role in disease pathogenesis.\(^{169-171}\) Deficiency of L-isoaspartate O-methyltransferase (PIMT), which catalyzes the repair of isoAsp proteins formed by the spontaneous conversion of aspartic acid to its isoaspartyl derivative, has also been shown to result in an accumulation of isoAsp proteins and the development of lupus in a mouse model.\(^{172}\) Furthermore, in the PL/J model of EAE, acetylation of the encephalitogenic Ac1-11 peptide of MBP is required for T cell activation even though unmodified peptide binds to MHC.\(^{173}\) These and other studies suggest that protein modifications can generate new and/or cryptic epitopes either directly by creating new structures such as citrulline or indirectly by altering MHC binding or modifying sites of peptide processing.\(^{174}\)

Neo self-antigens can also arise from changes in overall structure. An example of this is the formation of immune-genic immune complexes from nonantigenic soluble monomeric IgG, which can induce rheumatoid factors, antibodies to the Fc portion of complexed IgG.\(^{175,176}\) Likewise in Goodpasture’s disease, a configuration change in type IV collagen due to sulfilimine bonds produces a neotarget for pathogenic autoantibodies, a mechanism coined conformeropathy.\(^{177}\)

Another potential mechanism for triggering autoimmunity in susceptible individuals is lymphopenia-induced homeostatic expansion of T cells.\(^{178,179}\) Expansion occurs because of greater availability of survival-promoting cytokines (IL-7, IL-15), which, when combined with low-affinity engagement of the TCR to self-peptide/MHC, induces low-grade proliferation without full activation. It was therefore hypothesized that the requirement for self-reactivity could result in the preferential expansion of autoreactive T cells and consequently autoimmunity.\(^{175}\) The presence of lymphopenia in certain autoimmune diseases such as SLE and RA, autoimmune manifestations in some primary immunodeficiencies with lymphopenia, and several experimental autoimmune models support this.

In addition to self-antigens, foreign antigens with sufficient sequence or structural similarity can cross-activate nontolerant T (and B) lymphocytes, a mechanism termed molecular mimicry.\(^{180}\) For T cells, several findings support this possibility: (1) cross-reactivity requires only short peptide lengths of 8 to 15 amino acids; (2) T cell recognition is highly degenerate depending on only a few key amino acid residues, and it is possible to have mimotopes with no identical amino acids at any position; (3) it is estimated that a single T cell can react with \(10^4\) to more than \(10^5\) different peptides; (4) MBP-specific T cells cloned from MS patients can be stimulated by diverse microbial peptides; and (5) infection with a modified Theiler’s virus expressing a foreign cross-reactive peptide (Haemophilus influenzae protease IV protein that shares only 6 of 13 amino acids) induced T cells against a myelin protein, proteolipid protein (PLP), resulting in autoimmune CNS disease.\(^{186}\) To date, however, there is no compelling evidence connecting a specific microbial T cell mimotope to any autoimmune disease.\(^{187,188}\)

In contrast, there is fairly good evidence for molecular mimicry affecting self-reactive B cells for a few autoimmune diseases. The best examples include cross-reactivity of (1) bacterial adhesin FimH with LAMP-2 (lysosomal membrane-2) in ANCA-positive pauci-immune focal necrotizing glomerulonephritis; (2) group A streptococcal carbohydrate epitope, N-acetyl glucosamine, and M protein with cardiac myosin in rheumatic fever; and (3) Campylobacter jejuni lipo-oligosaccharide with ganglioside GM1 on peripheral nerves in the acute motor axonal neuropathy subtype of Guillain-Barré syndrome because of an identical determinant, [Gal β1–3 GalNAc β1–4 (NeuAc α2-3) Galβ]\(^{189}\). Overall, however, although molecular mimicry is an attractive hypothesis, supportive evidence for most autoimmune diseases is lacking. Whether this is because of multiple mimotopes from diverse sources or to perhaps the considerable plasticity of the T cell receptor remains an open question.

**IMMUNOLOGIC MECHANISMS OF TISSUE INFLAMMATION AND DYSFUNCTION**

The same effector mechanisms used by the immune system to neutralize pathogens are exploited in autoimmunity to inflict a wide range of deleterious effects on self-molecules, cells, and tissues. These have been broadly grouped into hypersensitivity types II to IV, which respectively encompass antibody-, immune complex-, and T cell–mediated processes (see Table 20-2).

In Type II reactions, pathologic autoantibodies bind to self-antigens primarily located on cell surfaces or in tissues and mediate autoimmune disease by three general mechanisms: (1) altering the function of the target antigen, (2) promoting cell injury or death, and (3) inducing inflammation. Blocking or enhancement of self-molecule function represents a special kind of type II hypersensitivity response in which autoantibodies alone are sufficient to effect autoimmune manifestations. Examples include the agonist anti-TSH receptor antibodies in Graves’ disease that stimulate...
thyrocyte growth and production of thyroid hormone, the antiacetylcholine receptor antibodies that block neuromuscular transmission in myasthenia gravis, and anti-β2-glycoprotein I antibodies in antiphospholipid syndrome that alter the regulation of anticoagulant activity. 193

Direct cell injury or death is mediated by the binding of IgM or IgG to surface antigens leading to cell lysis directly by complement activation or to phagocytosis via interaction of deposited C3 fragments with CR1 and CR3 receptors. Bound IgG also promotes phagocytosis through its interaction with Fc receptors. Examples include autoimmune hemolytic anemia, idiopathic thrombocytopenia, and autoimmune neutropenia.

Finally, antibodies bound to tissue antigens promote inflammation by activating complement, which generates the chemoattractant C5a and leukocyte-activating C3 fragments, and by FcγR binding, which activates immigrating leukocytes such as neutrophils and macrophages, as well as tissue resident mast cells and basophils. These cells produce proinflammatory factors that further expand the inflammatory response by recruiting and then activating additional circulating leukocytes.

Abnormal deposition of immune complexes of IgG antibody and soluble antigen in tissues cause type III responses. Such complexes, which also contain bound C3 complement fragments, are normally removed from the circulation by complement receptors on RBCs and by complement receptors and FcγRs on mononuclear phagocytes and platelets. However, this clearance can be overwhelmed under certain conditions such as overabundant production or immune complexes composed of excess antigen wherein less antibody coverage reduces both complement deposition and aggregation of Fc regions, leading to less efficient clearance. Once deposited in tissues, immune complexes initiate, through complement activation and FcγR binding, the same inflammatory cascades as type II responses. SLE and RA are autoimmune diseases mediated by this mechanism.

Type IV hypersensitivity encompasses cell and tissue injury mediated by activated T cells through their cytolytic activity in the case of CD8 T cells or the production of proinflammatory factors primarily by the CD4 subset. Apart from direct evidence for this mechanism in animal models using approaches not feasible in human studies, indirect evidence supporting this mechanism includes a higher frequency of autoreactive T cells with effector function in patients with autoimmune diseases, immunopathologic findings similar to T cell–mediated autoimmune models, and inhibition with T cell–blocking agents such as cyclosporin A. 191 T1DM and MS are examples of type IV reactions.

It should be mentioned, however, that the mechanism for some diseases is not always readily apparent. In RA, for example, the mechanism responsible for injury or dysfunction probably involves more than one hypersensitivity type, 195,196 whereas in others such as SLE, different clinical manifestations are mediated by different mechanisms (e.g., antineuronal antibody-mediated CNS pathology is a type II process, whereas glomerulonephritis is type III). 197–200 Finally, for some diseases such as dermatomyositis and systemic sclerosis, the type of mechanism mediating tissue injury remains to be defined.

PATHOPHYSIOLOGY OF AUTOIMMUNE RHEUMATIC DISEASES

Although the general principles underlying loss of tolerance and autoimmunity provide a useful conceptual foundation, the actual pathophysiologies—suggested by the few better-defined autoimmune diseases—are likely to also employ specific and unique mechanisms. Two notable examples of this are the pathophysiology of antinuclear antibody production in SLE and arthritis in RA. Brief discussions in the context of their broader significance to autoimmune diseases follow (see also Chapters 69 and 79).

Recent findings suggest a model of SLE pathophysiology that provides a mechanism for why antinuclear antibodies are virtually always present in lupus despite substantial genetic and clinical heterogeneity. 196 First, autoreactive B cells activate when self-reactive BCRs bind to nucleosomes or RNPs and internalize them into the endolysosome compartment, where the released nucleic acids engage TLR7 and TLR9 and provide a second signal. Such activated B cells act as potent antigen-presenting cells for T lymphocytes, and following class-switch recombination they produce IgG autoantibodies. Next, immune complexes formed by the binding of these autoantibodies to nucleic acid–containing material activate pDCs and DCs following their internalization via FcγRIIA (FcγRII in mice) and then engagement of TLRs by nucleic acids. Elaboration of lupus-promoting cytokines such as type I IFN and BAFF by pDCs and DCs, as well as enhanced antigen presentation, are thought to further drive loss of tolerance, activation of autoreactive B cells, and autoantibody production, thereby resulting in an amplification loop. Thus in lupus-susceptible individuals, confinement of nucleic acid–binding TLRs to endosomes is not enough of a barrier to block their activation by normally innocuous amounts of self-nucleic acids. This mechanism provides an explanation for the high prevalence of autoantibodies in SLE that either bind to nucleic acid or nucleic acid–complexed self-antigens such as DNA, nucleosomes, RNP, and myeloperoxidase (ANCA) or that exhibit cross-reactivity to such antibodies. Importantly, this also raises the possibility that other autoimmune diseases might also be mediated by specific PRRs.

The pathophysiology of RA also involves common and specific pathways that together promote inflammatory synovitis in roughly two phases, 194,196,201,202 which has been best described for the HLA-DR1*401 (DR4) anti cyclic citrullinated peptide (CCP) subset of RA patients with greater disease severity. The initial phase is postulated to involve the activation of T cells and B cells and the production of autoantibodies, including RF and anti-CCP. Targets for anti-CCP include citrillinated fibrinogen, vimentin, and α-enolase, and these antibodies are thought to play a major role in disease pathogenesis, although the mechanism remains controversial. The specific triggers of the anti-CCP response are also not known, although it is postulated that release of peptidylarginine deaminase from apoptosing granulocytes and monocytes may be a contributing factor. The resulting immune complexes and activated T cells lead to stimulation of macrophages, synovial fibroblasts, endothelial cells, mast cells, and osteoclasts, and the eventual production of proinflammatory factors such as TNF, IL-1, IFN-γ, chemokines, matrix metalloproteinases,
osteopontin, and many others that contribute to synovitis, pannus formation, bone erosion, and cartilage destruction. This later inflammatory and tissue damaging phase is largely mediated by activated fibroblast-like synoviocytes, which produce a wide array of proinflammatory mediators that promote recruitment and activation of circulating and resident immune cells. It has also been suggested that the spread of arthritis to unaffected joints could, in fact, be mediated by transmigration of these activated fibroblast-like synovial cells.203 The pathobiology of RA provides an additional example of the collaboration of the innate and adaptive arms of the immune system in autoimmune disease but also has two unique features; first, the major autoantigen, citrullinated protein, is a neoantigen specific for this disease, and second, tissue damage is ultimately mediated by fibroblast-like synoviocytes.

**GENETICS OF AUTOIMMUNE DISEASES**

Over the past few years, greater insight into the genetic landscape responsible for autoimmune disease susceptibility has come from both human and animal studies (see also Chapter 21). This progress was greatly facilitated by the availability of genomic sequences, improved definition of genetic variations and haplotypes among human populations, consortia with collections of patients and controls in the thousands, and numerous major technical and analytical advances.204,205 In particular, genetic studies have progressed from testing a few specific candidate polymorphisms to genome-wide family analyses of hundreds of cases, and even larger scale genome-wide association studies (GWAS) involving thousands,206–208 making it possible to capture common disease-predisposing variants with modest effects. Combined, these approaches have identified more than 30 candidate genes in SLE and RA, more than 10 in systemic sclerosis, and a few in Kawasaki’s and Behcet’s diseases.208-214 (see www.genome.gov/gwastudies for additional information). These candidates span the gamut of innate and adaptive immune systems but also include some loci with genes of unknown immunologic function. For example, in SLE, probably the best defined at the genetic level of any rheumatic disease, there are candidate genes involved in antigen presentation (HLA-DR3); B and T cell receptor signaling (PTPN22, BANK1, BLK), CD4 T helper cell regulation (OX40L or TNFSF4); T cell–mediated regulation (PDCD1); cytokine signaling (STAT4); interferon and TLR7/9 signaling (IRF5, TNFAIP3, IRAK1, IRF7, TYK2); Fc receptor function, one of which has been implicated in the transport of nucleic acid–containing immune complexes to TLR7/9-containing endosomes (FCGR2A); neutrophil function (ITGAM); clearance of self-antigens (C1Q, C2, C4); clearance of intracytoplasmic DNA (TREX1); and also several loci containing genes with no known connection to the immune system or lupus. Together these provide clues to specific pathways involved in SLE, which indeed has dovetailed well with more extensive genetic studies in mice encompassing more than 120 genes.7

From GWAS and other studies including those in animal models, several general conclusions can be made about genetic susceptibility in the more common autoimmune diseases. First, autoimmune diseases are associated with a large number of susceptibility genes shown to impact a wide range of immunologic, cellular, and end-organ functions in ways that enhance, modify, or even suppress relevant pathophysiologic processes.

Second, there is considerable genetic heterogeneity at both the individual and population levels regardless of whether the phenotype is relatively uniform as in RA or diverse as in SLE. Furthermore, although there are a large number of predisposing genes, having only a subset of these genes is sufficient for disease development. To what extent this heterogeneity is due to defects in a few common pathways or to numerous unique pathways remains unclear.

Third, the vast majority of candidate genes or loci have only modest effect sizes with most odds ratios less than 1.5, although in some diseases, notably SLE, a few rare variants (e.g., C1Q deficiency associated with > 90% incidence of SLE and TREX1 mutations leading to chilblain lupus) have been identified that are highly penetrant. Another salient finding derived from GWAS analyses is that for most autoimmune diseases, HLA alleles consistently have the highest or among the highest effect sizes. This accords well with the central role of antigen presentation and T cells in directing the adaptive immune response to specific antigens. Overall, however, for most candidate variations, defining the mechanism and proving a role in autoimmune disease will be hampered by their low effect sizes.

Fourth, some of the variant genes and loci are shared among autoimmune diseases, suggesting common underlying mechanisms.215 Noteworthy examples are the association of PTPN22 with a wide range of autoimmune diseases including T1DM, RA, SLE, JLA, Graves’ disease, systemic sclerosis, myasthenia gravis, generalized vitiligo, and granulomatosis with polyangitis (formerly Wegener’s granulomatosis), but not multiple sclerosis216 and STAT4 with RA, SLE, systemic sclerosis, and Sjögren’s syndrome.217,218 This supports a role for genetic factors in the known occurrence of multiple different autoimmune diseases in some families.

Fifth, common single-nucleotide polymorphism (SNP)-defined variants account for only a portion of overall heritability in autoimmune diseases (i.e., 20% to 60%)204,205,219 of which the HLA region typically accounts for a substantial part. Several reasons for the missing heritability have been suggested: (1) failed detection because of inadequate SNP coverage or the presence of disease-promoting non-SNP genomic variations such as copy number variants, (2) a large number of common variants with modest to marginal effects (odds ratio < 1.1 to 1.2) that are undetectable despite large study sizes given that the statistical power is reduced by both smaller effect size and lower variant frequency in the population, and (3) uncommon variants (1% to 5%) or rare disease-associated risk alleles (<1% frequency). Any of these possibilities will present significant challenges to further defining genetic susceptibility. It should also be noted that most common SNP variants associated with diseases do not affect coding regions, making it difficult to determine if the SNP variant is the actual change associated with disease, and to decipher its effects on gene function and autoimmunity.

Finally, information gleaned from genetic studies to date has not been useful for identifying individuals at risk, and it is thought that this will remain so because of the small effect sizes and relatively high frequency of known risk
alleles, combined with incomplete coverage of heritability. Nevertheless, these data have helped define disease-relevant pathways and within that context may lead to the identification of potential therapeutic targets.

**GENDER AND AUTOIMMUNITY**

A significant female gender bias in autoimmunity was identified early on and was considered an important clue to disease pathogenesis. The strength of this predilection, however, varies among autoimmune disorders with female prevalence in the 80% to 95% range for thyroiditis, SLE, Sjögren’s syndrome, and antiphospholipid syndrome, in the 60% to 75% range for RA, scleroderma, myasthenia gravis, and MS, but close to 50% for T1DM and autoimmune myocarditis. Why some autoimmune diseases are affected more by gender than others is not known, but the lack of common or distinguishing characteristics in those diseases with higher female predominance suggests the possibility of multiple mechanisms.

Indeed, to differing extents, depending on the disease, both sex hormones and sex chromosomes have been implicated in this dichotomy. In addition, fetal-maternal chimerism has been suggested as another factor, but this remains controversial. In terms of sex hormones, there is substantial evidence from in vitro, animal, and clinical studies to support a role for both female and male sex steroids in modifying the incidence and severity of autoimmunity. For example, in SLE, both estrogens and prolactin have been shown to exacerbate disease. In animal studies, they promote loss of tolerance and expansion of B cell populations, which, for estrogens, was also shown to be associated with enhanced survival and greater expression of Bcl-2. Importantly, studies of lupus-prone mice involving combinations of castration and hormone replacement convincingly demonstrated a disease-promoting effect of female hormones on spontaneous systemic autoimmunity.

Evidence also supports a role for sex chromosomes in autoimmune disease susceptibility. Individuals with Klinefelter’s syndrome (XXY) have a higher than expected incidence of SLE but under-representation in Turner’s syndrome (XO). Furthermore, using genetic and gonad manipulation in mice that allowed direct comparison of the effects of one or two X chromosomes on autoimmunity, susceptibility to both EAE in SJL mice, and TMPD-induced lupus was shown to be more severe in mice with XX. Thus it appears that both hormonal and to some extent chromosomal influences play roles in female predilection to autoimmunity.

**MICROBIAL AND OTHER ENVIRONMENTAL TRIGGERS**

Considerable evidence indicates that environmental factors can exert varying degrees of influence on the development of autoimmune diseases. A clear causal effect has been documented for a few disorders, while the low concordance rate (20% to 50%) among monozygotic twins in common autoimmune diseases support the presence of a significant environmental component. Nevertheless, the specific environmental factors and the extent to which they contribute to disease induction and exacerbation remain largely unknown for the majority of autoimmune disorders. The reasons for this are likely manifold and, depending on the disease, may include unsuspected or unknown environmental culprits, multiple independent factors, additive factors contributing to cumulatively small portions of the overall environmental effect, a highly variable or prolonged interval between exposure and disease onset, and a low incidence of autoimmune disease following exposure. Moreover, another general impediment is the difficulty of proving causation with epidemiologic data. Because of these difficulties animal studies have been invaluable in allowing controlled experimentation, and much of what we understand mechanistically has come about through this approach.

Despite these limitations, a broad range of environmental factors have been implicated, albeit with varying levels of confidence. Notably, the most prominent are infection and exposure to microbial products, which epidemiologic and animal models suggest can either enhance or inhibit autoimmunity depending on the type of exposure and disease. Several mechanisms by which microbes are postulated to induce or exacerbate autoimmunity include (1) molecular mimicry, (2) bystander activation of autoreactive T cells by pathogen-activated antigen-presenting cells, (3) inflammation of the target tissue and the release of immunologically hidden self-antigens, and (4) the production of disease-promoting cytokines such as IFN-α. Inhibition of autoimmune diseases by pathogens was initially suggested by epidemiologic evidence in T1DM and multiple sclerosis consistent with the “hygiene hypothesis” (i.e., a lower incidence of infection is responsible for an increasing incidence of allergic disease in Western countries). This concept has since been supported by additional epidemiologic and experimental data, although the microbial pathogens, type of exposure, and mechanisms remain uncertain. A recent study, however, of T1DM-prone NOD mice, which get worse in clean conditions and better with exposure to microbial products, showed that alterations in gut microbiota, which can interact with and modulate the host immune and inflammatory apparatus, could account for a significant portion of the environmental effects. Many of these mechanisms do not require specific pathogens but can be mediated by a wide spectrum of organisms. Finally, there are animal models that develop autoimmune disease under germ-free conditions, thereby excluding the requirement for microbial environment or infection. In this case, one would infer that the trigger is provided by self, most likely cell damage–derived products. Examples include type 1 diabetes in NOD mice and BB rats, APECED in AIRE-deficient mice, IPEX in Foxp3 deficiency, and lupus in MRL-lpr mice. Interestingly, however, in MRL-lpr mouse the addition of a filtered diet containing less PAMPs reduces disease severity. This suggests that microbial products at least partially affect disease susceptibility in this model and points to the difficulty of completely excluding a role for microbial organisms.

In terms of the other nonmicrobial environmental factors, some of the general types include (1) drugs such as procainamide, gold salts, and interferons that typically cause mild autoimmunity that resolves after discontinuation; (2) trauma, of which sympathetic ophthalmia caused by penetrating injury to the globe is an example, and (3) diverse environmental agents such as adulterated oil (caused...
CONCLUSION

Since 1904, when Donath and Landsteiner reported the first evidence of self-reactivity in paroxysmal hemoglobinuria,241 the field of autoimmunity has progressed tremendously at both the clinical and basic science levels. Notably, recent progress has made it possible to begin to dissect the interrelationships between genetic susceptibility and pathophysiology, and indeed several novel pathways linking etiology with immunopathology have already been identified. Recent studies have also been instrumental in documenting the central role of innate responses in the development of autoimmune adaptive immune responses. Importantly, they suggest that PRR-mediated innate immune responses might be a general requirement for the initiation of autoimmunity and furthermore that the type of adaptive autoimmune response and type of autoimmune disease may be largely determined by the nature of the innate response. Also, of importance to patient care, has been the successful bench-to-bedside translation approach that has brought therapeutic targeting of specific immune-related molecules to the clinic. Despite these substantial areas of progress, however, many facets of autoimmunity are still incompletely understood and patient care thus remains suboptimal. For instance in the area of therapy, the highly sought objective of specifically eliminating autoreactive, but not foreign, antigen-recognizing lymphocytes has remained elusive despite considerable effort242,243 and current, more broadly immunosuppressing therapies are only partially effective and are associated with infection and cancer risks. Accordingly, there continues to be a critical need to investigate the basic processes involved in autoimmune diseases and to develop new and more effective ways to manage afflicted individuals.

With this in mind, there are several areas of research that are poised to address some of the current deficiencies. At the bench it can be anticipated that the pathophysiology of at least the common autoimmune diseases will be defined in increasing detail, and the critical corresponding cell types and molecular components identified. In animal studies, existing technology is at the point where all genes essential for the development of a particular autoimmune disease model can be identified. Genetic studies in both human and animal models should continue to help focus clinical efforts on relevant immune and cellular pathways, and new technologies are likely to make it possible to identify rare disease-affecting genetic variants. In the clinic, major benefits are likely to be derived from new biomarkers, new therapeutics, predictive risk factors both genetic and environmental, and the transformation to individualized care approaches. In general, scientific advances in genome sequencing, application of a systems biologic approach, and instrumentation are also likely to have positive impacts. Thus, the future should continue to see significant progress in many areas related to autoimmunity and importantly, an improved outlook for patients.

Acknowledgments

This is manuscript number 21411-IMM from The Scripps Research Institute. We thank Dr. Brian Lawson and Kat Occhipinti for editorial assistance. This work was supported by NIH grants from NIAMS, NIAID, and NIEHS.

Selected References


Full references for this chapter can be found on www.expertconsult.com.


Genetics of Rheumatic Diseases

PETER K. GREGERSEN

looking toward a future where genetic knowledge begins to be integrated into the daily practice of medicine.

MAJOR HISTOCOMPATIBILITY COMPLEX

The MHC, which encodes the human leukocyte antigens (HLAs), has been associated with susceptibility to many different diseases since the late 1970s. The chromosomal region containing the MHC was originally identified because of the ability of genes in this region to regulate transplant rejection and to control the immune responses of mice and guinea pigs to simple antigens, a series of observations that led to the 1980 Nobel prize. The HLA molecules and their counterparts in rodents were subsequently shown to be directly responsible for immune response differences between individuals and for determining the likelihood of graft rejection. A large number of different HLA genes exist within the MHC, and they exhibit an enormous degree of structural variability. In addition to influencing immune response patterns, many of these alleles are associated with susceptibility to a wide spectrum of autoimmune diseases, making the MHC an essential starting point for anyone wanting to understand the genetics of rheumatic diseases.

Human Leukocyte Antigen Molecules and Antigen-Specific T Cell Recognition

The primary function of HLA molecules is the presentation of antigenic peptides to T cells. In the case of α/β T cells, most antigen recognition events involve the formation of a trimolecular complex consisting of the HLA molecule, its bound peptide, and the α/β T cell receptor (see Chapters 13 and 19). When this recognition occurs in the appropriate context, it may result in signal transduction and activation of the T cell. The requirement for MHC molecules to present antigenic peptides to T cells is frequently referred to as “MHC-restricted” T cell recognition. In each individual, T cells are generally restricted to recognize antigens presented by the person’s own HLA molecules. The allelic variations among different HLA molecules are a major factor accounting for differences in the types of antigenic peptides to which an individual responds or in the types of T cells that are used in an immune response.

Human Leukocyte Antigen Class I and Class II Molecules

The original serologic and biochemical studies of HLA molecules revealed the presence of two major isotypes: HLA class I and HLA class II. The basic structural features of
domains. The overall size and shape of the cleft in the two classes of HLA molecules are almost superimposable. However, subtle differences exist, particularly at the ends of the cleft, and this allows for some differences in the sizes of antigenic peptides that are presented by HLA class II molecules, compared with class I. Direct analysis of peptides bound to HLA class II molecules has shown that their size commonly varies from 12 to 19 amino acids. Thus relatively longer peptides lie in the cleft of class II molecules and may extend beyond the ends of the cleft, whereas class I molecules contain much shorter peptides that are buried within the cleft at either end.

Finally, x-ray crystallographic analyses of the entire trimolecular complex consisting of an HLA molecule, its bound peptide, and the T cell receptor exist. Figure 21-3 shows an example of this structure for an HLA-DRB1*0401 allele presenting influenza Ha peptide to its cognate α/β T
Initially, experts considered that the predominant role in presenting antigen to CD8 T cells. This distribution reflects their distribution on all somatic cells, with the exception of red blood cells. This distribution reflects their predominant role in presenting antigen to CD8 effector or cytotoxic T cells.

Chapter 21 discusses this complex process in detail.

**Genetic Organization of the Human Major Histocompatibility Complex**

The human MHC extends over approximately 4 million base pairs on the short arm of chromosome 6 (6p21.3). The HLA class I and class II gene clusters are found in distinct locations, as indicated on the highly abridged genetic map shown in Figure 21-4. Only those genes that are traditionally associated with immune function are shown in Figure 21-4. The MHC, one of the most gene-rich regions in the human genome, has identified more than 200 genes.

The HLA class I α-chain genes are on the telomeric side of the MHC including the classical class I genes, HLA-A, HLA-B, and HLA-C; these three genes are also referred to as the class Ia genes. Researchers have defined several other class I genes including the HLA-G, HLA-E, and HLA-F loci, also known as class Ib genes. Initially, experts considered these class Ib genes to be nonfunctional because they exhibit limited polymorphism. However, some of these genes clearly have significant immune functions. For example, HLA-E functions as a ligand for natural killer (NK) cell receptors. In a surprising twist, the cell surface expression of HLA-E molecules often depends on binding peptides derived from the signal sequence of class I molecules, but they can also present antigenic peptides to αβ TCR-bearing CD8 T cells. Despite its limited polymorphism, alleles of HLA-E may still affect responses to some viral infections.

The HLA class II genes are situated centromeric to the class I region and have a somewhat more complicated organization. The three major subregions of the class II cluster are designated DR, DQ, and DP. Each of these subregions contains a variable number of α- and β-chain genes. Particularly in the case of the DR subregion, this variability has led to confusion regarding the nomenclature to describe these genes. Experts have agreed on an international standard for this nomenclature regularly update it (see www.ebi.ac.uk/imgt/hla/).

The DR subregion contains a single α-chain gene, designated DRA, which does not exhibit significant allelic variation. In contrast, the genes encoding the DRβ chains (DRB) are highly polymorphic and vary in number among different individuals in the population. The boxed area in Figure 21-4 shows several examples of common DR haplotypes. (A haplotype refers to a group of alleles at closely linked loci that are commonly inherited together.) Many of these DRB genes are nonfunctional pseudogenes (indicated by the symbol ψ), although all haplotypes contain at least one functional DRB1 gene, and many haplotypes contain a second functional DRB gene (DRB3, DRB4, or DRB5).

**Human Leukocyte Antigen Class I and Class II Isotypes: Functional Correlates**

HLA class II molecules have a restricted tissue distribution, generally limited to antigen-presenting cells of the immune system such as B cells, macrophages, dendritic cells, and some subsets of T cells. This reflects the fact that HLA class II molecules are primarily involved in presenting foreign antigens to CD4 T cells during the initiation and propagation of the immune response. However, the expression of HLA class II molecules can also be induced on a variety of other cell types by inflammatory cytokines such as interferon-γ, enabling these cells to engage in antigen presentation to CD4 T cells. In contrast, HLA class I molecules are widely distributed on all somatic cells, with the exception of red blood cells. This distribution reflects their predominant role in presenting antigen to CD8 effector or cytotoxic T cells.

Another major functional difference between class I and class II molecules is related to the source of peptide antigens that are found in the antigen-binding cleft. In general, class I molecules present peptide antigens derived from proteins that are actively synthesized within the endoplasmic reticulum, whereas HLA class II molecules present antigens that are taken up from outside of the cell by endocytosis. These differences are reflected in the antigen processing machinery and the different trafficking patterns of class I and class II molecules inside the cell. Chapters 10 and 19 discusses this complex process in detail.

**Figure 21-3** Ribbon diagram derived from the three-dimensional crystal structure of a the trimolecular complex of a human α/β T cell receptor (top, green), influenza Hα antigen peptide, and the MHC class II molecule, DRB1*0401. Note that the peptide is contained with the peptide-binding cleft of the HLA-DR molecule. The polymorphisms associated with the “shared epitope” are located on the α-helical rim (DRB1 chain) of the peptide-binding cleft, where they may interact with either the bound peptide antigen or the T cell receptor. The Protein Data Bank (PDB) website (www.rcsb.org/pdb/home/home.do) provides three-dimensional versions of this HLA structure (Entry code 1J8H).
The DQ subregion contains one pair of functional \(\alpha\) and \(\beta\)-chain genes, designated DQA1 and DQB1. These two genes encode all the known HLA-DQ molecules. Significant expression of the protein products of the DQA2 and DQB2 loci has not been reported. A similar situation exists in the DP subregion, in which only the DPA1 and DPB1 genes give rise to known protein products, the HLA-DP molecules.

In addition to the HLA class II molecules, several other genes distributed within the class II region are involved in peptide antigen processing. The TAP1 and TAP2 genes have been of particular interest because they exhibit a modest degree of polymorphism and are involved in delivering peptides for loading onto HLA class I molecules.\(^{17}\) Other proteins encoded in this region are involved in peptide loading onto class II molecules\(^{18}\) such as the DM molecule (encoded by the DMA and DMB genes) and the DO/DN heterodimer (encoded by DOB and DNA).

The “central” MHC is located in between the MHC class I and class II regions. It contains a number of genes involved in immune function including those for tumor necrosis factor (TNF) and lymphotoxin, as well as the complement components C4A, C4B, C2, and factor B. The central MHC is quite gene dense, and it contains a significant number of other genes with potential relevance to immune function and disease susceptibility. However, the role of many genes in central MHC and class I regions is still largely unknown, and the interpretation of genetic associations with this region is complicated by the strong linkage disequilibrium across the MHC (see later).

**Human Leukocyte Antigen Molecules Are Highly Polymorphic**

One of the most dramatic features of the HLA system is the extreme degree of polymorphism at most of these loci. The formal definition of polymorphism (Table 21-1) usually requires that the most common allele at the locus does not exceed a frequency of 98%. In contrast, at many HLA loci, it is uncommon for a single HLA allele to exceed a frequency of 50% in the population. The number of different alleles present in the population is much larger than in any other known polymorphic locus encoding functional genes. For example, at the HLA-A locus, more than 100 different alleles have been reported; at HLA-B, the number of reported alleles exceeds 200. A similar degree of allelic diversity is seen at the DRB1 locus and to a lesser degree at DQA and DQB.

For many years the naming of these various HLA alleles was a major source of confusion in the literature. The difficulty with the nomenclature stems in part from the different methods that have been used to define HLA polymorphisms. Originally, HLA class I alleles were detected through the use of alloantisera (see Table 21-1); the prefix “allo-” refers to genetic differences that exist between individuals of the same species. Alloantisera directed toward HLA molecules are commonly found in the context of pregnancy, in which the mother mounts an immune response against the “foreign” HLA molecules carried by the fetus (derived from the father). Anti-HLA responses are also seen after blood transfusion because the HLA molecules on the
donor cells are highly immunogenic. In the case of HLA class II alleles, differences were originally detected using mixed lymphocyte responses. When T cells from a responder are mixed with lymphocytes from another individual, differences in HLA class II alleles cause the responder’s T cells to proliferate. Data on mixed lymphocyte culture (MLC) typing dominated the early HLA literature, and it was the method first used to detect the HLA class II associations with RA. Subsequently, serologic methods were also employed to detect class II polymorphisms.

The current names of the HLA class I and class II alleles are attached to the specific DNA sequence and locus for each allele and are definitive. However, many older publications have used the serologically derived names for alleles. It is therefore important to have some concept of how these naming conventions are related. The modern definitive (sequence-based) allele names are derived from the older serologic names because the serologic techniques frequently detected whole groups of related alleles. Two examples of this are shown in Table 21-2. The designation of HLA-B27 was developed for people carrying an HLA-B allele that was recognized by the B27-specific alloantisera. However, sequencing of the HLA-B alleles carried by these individuals revealed the existence of at least 17 different alleles, 5 of which are listed in Table 21-2. A similar situation exists for HLA class II allele families such as HLA-DR4, also shown in Table 21-2. In this case, the DR4 allospecificity was already known to detect a number of different alleles that could be further discriminated on the basis of MLC typing. These have been defined and named by their sequence, as shown in Table 21-2. A full list of all HLA alleles at the major loci can be found at www.ebi.ac.uk/imgt/hla/.

Despite the precision of the molecular definition of HLA alleles, the old serologic names are often used in oral discussion because they are less cumbersome. For example, the DRB1*0301 allele is common in white populations (up to 10% in some populations) and is often referred to as simply DR3, after its original serologic designation. At least 16 distinct alleles are detected by such DR3 alloantisera, and therefore the term DR3 is imprecise. However, when used in the context of a discussion about white populations, DR3 is assumed to refer to the predominant DRB1*0301 allele.

### HUMAN LEUKOCYTE ANTIGEN ASSOCIATIONS WITH RHEUMATIC DISEASES

#### Population-Association Studies and the Calculation of the Odds Ratio, an Estimate of Relative Risk

The ideal way to establish whether a genetic variant (allele) confers risk for a disease is by performing a prospective cohort study. In this kind of study, a group of individuals

<table>
<thead>
<tr>
<th>Table 21-2</th>
<th>Comparison among Modern, Sequence-Based Nomenclature, and Older Naming Conventions for Class I and Class II Alleles Belonging to HLA-B27 and HLA-DR4 Serologic Groups*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HLA-B Locus: HLA-B27 Alleles</strong></td>
<td><strong>HLA-DR1 Locus: HLA-DR4 Alleles</strong></td>
</tr>
<tr>
<td>Definitive Nomenclature Based on DNA Sequence</td>
<td>Serologic Designation (Defined by Alloantisera)</td>
</tr>
<tr>
<td>B*0201</td>
<td>B27</td>
</tr>
<tr>
<td>B*0701</td>
<td>B27</td>
</tr>
<tr>
<td>B*0702</td>
<td>B27</td>
</tr>
<tr>
<td>B*0703</td>
<td>B27</td>
</tr>
<tr>
<td>B*0704</td>
<td>B27</td>
</tr>
</tbody>
</table>

*The list of alleles is incomplete. The B27 allele family contains at least 17 members, and the HLA-DR4 allele family contains at least 35 members. See www.ebi.ac.uk/imgt/hla/ for a complete list of all HLA alleles.

MLC, mixed lymphocyte culture.
Table 21-3  Contingency Table for Cohort and Case-Control Studies*

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>No Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Case-Control Study

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not Exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No Disease</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

*a, b, c, d, number of individuals observed in each category.

carrying (exposed to) the allele is compared with a matched control group that does not carry the allele. These two groups are followed over time (preferably over a lifetime) to see if disease develops more frequently in the exposed group. The results can be displayed in a contingency table (Table 21-3). By examining the upper half of Table 21-3, it is apparent that the fraction of exposed individuals who get the disease is \( a/(a+b) \), whereas the fraction of unexposed individuals who develop the disease is \( c/(c+d) \). The ratio of these two fractions is known as the relative risk (RR) = \( a/(a+b) \times c/(c+d) = (ac+ad)/(ac+bc) \). If the disease is rare in the population, \( ac \) is small and the RR is approximated by \( (a \times d)/(b \times c) \), also referred to as the cross-product.

In reality, such prospective cohort studies are usually impractical and therefore a retrospective case-control design is used. In this type of study, subjects are initially identified according to whether they have the disease and individuals without the disease are the controls. The data can be tabulated as in the lower half of Table 21-3. In this case, the cross-product or \( (a \times d)/(b \times c) \) is known as the odds ratio (OR). In practice, this quantity is often reported as the estimated RR because the cross-product is close to the RR when the disease is rare. An OR of 1 indicates that the genetic factor confers no risk for the disease. An OR less than 1 suggests that the genetic factor under study is negatively associated with the disease. (ORs of less than 1 are occasionally reported as the negative inverse value; an OR of +0.5 may also be reported as –2.0.) With the exception of HLA-B27–associated diseases, most HLA associations with rheumatic diseases have ORs of less than 10. Several examples of typical HLA associations with rheumatic and autoimmune disorders are shown in Table 21-4.

### Human Leukocyte Antigen Class I Associations: HLA-B27 and Spondyloarthropathies

One of the strongest and earliest \(^{21}\) reported HLA associations with the rheumatic diseases is the association of HLA-B27 with ankylosing spondylitis (AS). In white populations, more than 90% of patients with AS carry HLA-B27, in contrast to approximately 8% of normal individuals, giving estimated RR values of 50 to 100 or higher. The consistency of this finding across most ethnic groups lends support to the contention that the HLA-B27 alleles are directly involved in the pathogenesis of AS. \(^{22,23}\) HLA-B27 is also associated with reactive arthritis and with the arthritis seen in the context of inflammatory bowel disease. As shown in Table 21-4, the strength of these associations is lower in terms of estimated RR compared with ankylosing spondylitis.

The serologic specificity of HLA-B27 actually encompasses many distinct HLA class I alleles. These alleles differ from one another at a number of amino acid positions, most of which involve amino acid substitutions in and around the peptide binding pocket. This fact leads naturally to the question of whether there are differences among these B27 alleles in terms of disease association. Most data indicate that this is not the case, although there may be some exceptions in some populations. \(^{22}\) These exceptions may provide clues to the role of the HLA-B27 molecule in pathogenesis. Overall, however, it appears that most of the structural differences among the B27 alleles do not affect disease risk. \(^{23}\)

In recent years it has become apparent that increased risk for some severe drug reactions can be ascribed to HLA-B27. \(^{24}\) Including risk for allopurinol-associated Stevens-Johnson syndrome in Asian populations with HLA-B58. \(^{25}\) These data imply the need for specific class I HLA typing before treatment with some medications. \(^{26}\)

### Human Leukocyte Antigen Class II Associations with Autoimmune Diseases

A large number of HLA class II associations with autoimmune diseases were described over 2 decades ago. \(^{27}\) RA has received particularly intense scrutiny over the years, but the precise reasons for the HLA associations with this disease are still unknown. In the case of systemic lupus erythematosus (SLE) and related illnesses, many of the HLA class II alleles are associated with the presence of specific autoantibodies or clinical phenotypes. Interestingly, the recent data in RA indicate that the major HLA-DR associations are
with anti-CCP antibody positive disease, suggesting that control of autoantibody responses may be a primary mechanism underlying these associations in RA as well.

Rheumatoid Arthritis: HLA-DRB1 Associations and the “Shared Epitope”

Stastny reported the first associations of rheumatoid arthritis (RA) with HLA class II alleles in the 1970s. This was done using cellular and antibody reagents that are no longer routinely used for HLA typing; however, as discussed earlier, the nomenclature for HLA alleles still derives from these early typing methods. The DRB1*0401 allele (corresponding to the “Dw4” type in Stastny’s original report) was the first HLA polymorphism to be associated with RA. Numerous studies have generally confirmed that this allele is the most strongly associated with RA, at least in white populations. However, several other HLA-DRB1 alleles have also been associated with RA, although the strength of these associations varies. In some ethnic groups, RA is not associated with HLA-DR4 alleles, but rather with HLA-DR1 or HLA-DR10. Experts now widely accept that the following alleles are the major contributors to RA risk at the DRB1 locus: DRB1*0401, -0404, -0405, -0101, and -1001. In addition, minor variants of these alleles and others (e.g., DRB1*1402) may also contribute to susceptibility and DRB1*0901 is a susceptibility allele in Asians, where this allele is common. Most of these risk alleles share a common sequence, as shown in Table 21-5. This consensus amino acid sequence 70Q or K-R-R-A-A74 has been termed the shared epitope. This structural feature is located on the α-helical portion of the DRβ chain in a position where it may influence both peptide binding and T cell receptor interactions with the DRB1 molecule. (In the case of the DRB1*1001 risk allele, one amino acid varies from this consensus by a conservative change, with an R at position 70, as does DRB1*0901, which is commonly associated with RA in Asian populations) (Table 21-6).

A number of different hypotheses have been advanced to explain the shared epitope association with RA. Two of these follow directly from knowledge about the role of HLA molecules in antigen presentation and immune regulation. Thus it has been suggested that a particular peptide antigen, or set of related antigens, may be involved in the initiation or propagation of RA, and that shared epitope positive DRB1 alleles possess a unique, or enhanced, ability to bind or present these peptides to the immune system. It has been difficult to address this hypothesis directly because the identity of these putative disease-causing peptide antigens is unknown. In view of the strong association of the shared epitope alleles with anti-CCP antibodies, it is of interest that citrullinated peptides may have a particular affinity for DRB1*0401 alleles. A second major hypothesis posits that these risk alleles regulate the formation of the peripheral T cell repertoire, by acting to select for particular T cell receptors during thymic selection. There is elegant experimental evidence in humans to support a role for DR4 alleles in shaping the peripheral T cell repertoire. However, it is unclear whether this effect on the TCR repertoire is really related to disease susceptibility. Researchers have proposed a number of other interesting hypotheses, involving molecular mimicry, allele specific differences in intracellular trafficking, and regulation of nitric oxide production, but these require further experimental confirmation.

The shared epitope hypothesis has come under scrutiny, with some investigators proposing a direct role for HLA-DQ polymorphisms, in part based on studies in transgenic mice. As can be seen in Figure 21-4, the HLA DQ α and β chains are encoded just centromeric to DRB1, and alleles at this locus are in strong linkage disequilibrium with DRB1 alleles. The strong linkage disequilibrium between the DR and DQ loci makes it difficult to tease apart the effects of DR versus DQ solely on the basis of population genetic studies; the arguments for a DQ effect generally depend on showing the enrichment of relatively rare genotypes in the RA patient group compared with controls. Overall, a primary role for DQ alleles is not strongly supported by large HLA association studies that have examined this issue and is not supported by more recent dense single-nucleotide polymorphism (SNP) mapping efforts within the MHC. Rather, a possible additional role for the HLA-DP has been suggested.

Regardless of whether DQ or DP alleles are involved in RA susceptibility, it is quite clear that the shared epitope hypothesis is not a complete explanation for the HLA associations with RA. This is evident from the fact that not all SE-positive alleles carry the same degree of genetic risk and the strength of the association varies in different populations. In general, DRB1*0101 alleles carry lower levels of RR for RA than the DRB1*0401 and 0404 alleles, and yet DRB1*0101 is the major risk allele in some ethnic groups. The shared epitope itself does not appear to associate strongly with RA in African-American and some Hispanic

**Table 21-5** Amino Acid Substitutions That Compose the Shared Epitope at Positions 70 through 74 of DRB1 Alleles Associated with Rheumatoid Arthritis

<table>
<thead>
<tr>
<th>DRB1 Alleles</th>
<th>Amino Acid Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>0101</td>
<td>70 Arg 71 Ala 72 Ala 73 Ala 74 Arg</td>
</tr>
<tr>
<td>0401</td>
<td>70 Gln 71 Arg 72 Lys 73 Arg 74 Trr</td>
</tr>
<tr>
<td>0404</td>
<td>70 Ala 71 Lys 72 Lys 73 Lys 74 Arg</td>
</tr>
<tr>
<td>0405</td>
<td>70 Ala 71 Arg 72 Ala 73 Ala 74 Ala</td>
</tr>
<tr>
<td>0408</td>
<td>70 Ala 71 Lys 72 Lys 73 Lys 74 Arg</td>
</tr>
<tr>
<td>1402</td>
<td>70 Ala 71 Arg 72 Ala 73 Ala 74 Ala</td>
</tr>
<tr>
<td>1001</td>
<td>70 Ala 71 Arg 72 Ala 73 Ala 74 Ala</td>
</tr>
</tbody>
</table>

**Table 21-6** Genotype Relative Risks of DRB1 Genotypes for Rheumatoid Arthritis

<table>
<thead>
<tr>
<th>DRB1 Genotype</th>
<th>Relative Risk</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0101/DRX</td>
<td>2.3</td>
<td>10^-4</td>
</tr>
<tr>
<td>0401/DRX</td>
<td>4.7</td>
<td>10^-12</td>
</tr>
<tr>
<td>0404/DRX</td>
<td>5</td>
<td>10^-4</td>
</tr>
<tr>
<td>0101/0401</td>
<td>6.4</td>
<td>10^-4</td>
</tr>
<tr>
<td>0401/0404</td>
<td>31.3</td>
<td>10^-43</td>
</tr>
</tbody>
</table>

populations. Furthermore, certain combinations of DRB1 alleles carry especially high risk, as originally observed by Nepom. Thus the combination of DRB1*0401 with *0404 carries a RR of higher than 30 in Caucasian populations. This compares with RR values in the range of 4 or 5 for either allele alone. Table 21-5 summarizes some of these relationships. Recently, attempts have been made to formalize the gradient of risk conferred by the various shared epitope alleles. However, it remains unclear whether these effects are mediated by the HLA-DR molecules themselves or reflect the action of other genes on these haplotypes.

**HLA-DQ Associations with Autoimmune Diseases**

Many of the first HLA class II associations with autoimmune disorders were detected using alloantisera for HLA-DR alleles, as indicated in Table 21-4. However, as knowledge increased about the genetic organization of the class II region, it became apparent that for some diseases, the genetic associations are stronger with HLA-DQ alleles. For example, although juvenile diabetes does exhibit HLA associations with both HLA-DR4 and HLA-DR3, it is likely that a group of associated HLA-DQ alleles actually are responsible for these observations. As discussed later, the HLA associations with particular autoantibodies in systemic lupus also probably reflect the effects of HLA-DQ alleles.

The DQ subregion presents special challenges for the newcomer to HLA because the old serologic nomenclature does not usually have a simple correlation with a group of alleles at a single locus. Because most of the HLA correlations with autoantibodies in lupus involve the DQ loci, it is important to understand this at the outset. The problem arises because both the α and β chains are polymorphic in DQ molecules. The serologic specificity of DQ2 may detect one of three closely related DQB1 alleles: DQB1*0201, DQB1*0202, or DQB1*0203. This is analogous to the DR serologic specificity detecting a group of related DRB1 alleles (see Table 21-2). However, in the case of DQ, the DQ2 serologic specificity also detects these alleles on several different haplotypes that may encode quite different DQ α chains. (This is different from HLA-DR molecules, in which the DR α chain does not vary among different MHC haplotypes.)

The DQ subregion presents special challenges for the newcomer to HLA because the old serologic nomenclature does not usually have a simple correlation with a group of alleles at a single locus. Because most of the HLA correlations with autoantibodies in lupus involve the DQ loci, it is important to understand this at the outset. The problem arises because both the α and β chains are polymorphic in DQ molecules. The serologic specificity of DQ2 may detect one of three closely related DQB1 alleles: DQB1*0201, DQB1*0202, or DQB1*0203. This is analogous to the DR serologic specificity detecting a group of related DRB1 alleles (see Table 21-2). However, in the case of DQ, the DQ2 serologic specificity also detects these alleles on several different haplotypes that may encode quite different DQ α chains. (This is different from HLA-DR molecules, in which the DR α chain does not vary among different MHC haplotypes.)

The DQ subregion presents special challenges for the newcomer to HLA because the old serologic nomenclature does not usually have a simple correlation with a group of alleles at a single locus. Because most of the HLA correlations with autoantibodies in lupus involve the DQ loci, it is important to understand this at the outset. The problem arises because both the α and β chains are polymorphic in DQ molecules. The serologic specificity of DQ2 may detect one of three closely related DQB1 alleles: DQB1*0201, DQB1*0202, or DQB1*0203. This is analogous to the DR serologic specificity detecting a group of related DRB1 alleles (see Table 21-2). However, in the case of DQ, the DQ2 serologic specificity also detects these alleles on several different haplotypes that may encode quite different DQ α chains. (This is different from HLA-DR molecules, in which the DR α chain does not vary among different MHC haplotypes.)

**Population-Association Studies: What Do They Mean?**

Almost all of the studies on HLA and disease involve population associations that are detected by means of retrospective case-control studies. It is essential to understand the strengths and weaknesses of this approach to genetic analysis to judge the significance of these HLA associations. In general, there are three possible reasons for detecting an association between a particular allele and a disease, once acceptable statistical criteria are met (see later section).

First, the allele under investigation may be directly involved in the pathogenesis of the disease. This assumption actually underlies most of the foregoing discussion on HLA and rheumatic disease. The studies described reflect the search for a more precise definition of particular amino acid substitutions or unifying structural characteristics of disease-associated alleles. This effort derives from the idea that HLA alleles directly predispose to disease by virtue of their ability to control the immune response. As discussed earlier, this may involve a number of mechanisms including preferential peptide binding and the influence of MHC on thymic selection of the peripheral T cell repertoire.

A second reason that must be addressed with any new genetic association is the possibility that the result is an artifact of population stratification of patients and controls. The specific concern is that the control group may not be genetically matched to the disease group at loci that are unrelated to disease. This often results from a failure to
study a control group that is ethnically matched to the disease group. This is a major issue generally in genetic case-control studies, and several approaches to control for this have been proposed\(^6\) including the use of panels of genetic markers that specifically reflect ethnic background.\(^6\) These methodologies for correcting for underlying population are now widely accepted and indeed are often required for publication in leading genetics journals. It is generally not adequate to accept self-reported ethnicity as a basis for matching cases and controls.

Finally, a third (and common) reason for observing a genetic association is that the causative gene is actually in linkage disequilibrium with the marker allele being tested, be it a SNP or a particular HLA variant. Linkage disequilibrium is discussed in greater detail in the next section and refers to the fact that genetic variants at adjacent loci often tend to be found together more frequently than expected by chance. Linkage disequilibrium over long distances is a particularly prominent feature of the HLA region, particularly for certain haplotypes.\(^6\)\(^8\) A good example of how HLA associations can reflect linkage disequilibrium with a gene that is functionally unrelated to HLA is hemochromatosis. Early studies showed that certain HLA class I alleles such as HLA-A3 were highly associated with this disorder. However, it is now clear that the causative gene, \(HFE\), is actually more than 3 million base pairs distant from the HLA-A locus (toward the telomere in Figure 21-4). The HLA-A3 association is observed simply because the HFE C282Y allele (causative for hemochromatosis) is frequently found on the same haplotype (see Table 21-1) as the HLA-A3 allele in many white populations.

Because there are a number of genes with immunologic function within the MHC complex that may themselves be directly involved in predisposing to autoimmunity, these loci must always be considered as possible causative genes when considering the significance of a new HLA association with rheumatic disease. Indeed, recent studies of autoimmune diseases indicate that multiple genes within the MHC can contribute independently to disease risk. In the case of RA, a number of studies have indicated that a separate locus in the central MHC may associate with the disease, independently of the HLA-DRB1 locus.\(^7\)\(^-\)\(^9\) In addition, there is evidence that genes in the class I region may influence the risk conferred by certain HLA-DRB1*0404 haplotypes\(^7\)\(^2\) or may interact with other non-MHC genes.\(^7\)\(^3\) Similar analyses in type 1 diabetes,\(^7\)\(^4\) lupus,\(^5\)\(^6\) juvenile arthritis,\(^7\)\(^7\)\(^8\) multiple sclerosis,\(^7\)\(^9\) and myasthenia\(^8\) all point to the fact that multiple different genes within the MHC can contribute to disease susceptibility. This issue is currently a major focus of research efforts in autoimmune diseases, and it is highly likely that additional risk genes in the MHC will be defined in the next few years using the dense SNP mapping techniques discussed in the following sections.

**Linkage Disequilibrium**

The concept of linkage disequilibrium is central to understanding the significance of any genetic association including HLA associations with disease. Linkage disequilibrium exists when the frequency of two alleles occurring together on the same haplotype exceeds that predicted by chance. For example, a common MHC haplotype that exhibits linkage disequilibrium in the white population carries a certain combination of alleles, A*0101-B*0801-DRB1*03011, commonly referred to as the A1-B8-DR3 haplotype, and more recently designated the “8.1” haplotype.\(^6\)\(^9\) This haplotype is present in about 9% of the Danish population, a typical white Northern European group. To understand why this reflects the presence of linkage disequilibrium, consider the fact that the A1 allele is present in 17% of Danes and the B8 allele is present in 12.7% of Danes. They could be expected to be found together only 12.7% × 17% = 2.1% of the time, much less than what is observed (9%). This simple calculation between the expected and observed association between alleles is a measure of linkage disequilibrium, known as \(D\); in this case \(D = (0.09 - 0.021) = 0.069\). Table 21-7 provides the general calculation of \(D\) for a simple two-locus, two-allele situation. Because the magnitude of \(D\) is strongly influenced by the relative frequencies of the various alleles, normalized measures of linkage disequilibrium are used in practice including \(D'\) and \(r^2\), as described in Table 21-7.

It is useful to understand these measures of linkage disequilibrium because detailed maps of linkage disequilibrium are widely available online for the entire human genome, with easy-to-use visualization tools (see [www.hapmap.org](http://www.hapmap.org)). The lower portion of Figure 21-6 shows a visualization of LD using the \(D'\) measure for a region around the PTPN22 gene on chromosome 1. The \(D'\) value between any two markers is reflected by the heat map (red \(D' = 1\); white \(D' = 0\)). In this case, linkage disequilibrium extends well beyond the region around the PTPN22 gene, as indicated by the “8.1” haplotype, and more recently designated the “8.1” haplotype.

![Figure 21-6](image.png)

**Figure 21-6** Map of the region around the PTPN22 locus on chromosome 1p13 covering approximately 200,000 base pairs. The blue and yellow haplotype pattern in the central part of the figure was generated by looking at combinations of single nucleotide polymorphism (SNP) alleles in 90 white subjects from the HapMap Project. It is analogous to the presentation of linkage disequilibrium discussed in Table 21-7. Note that despite the large number of SNPs, a limited number of haplotype patterns are observed, generating a kind of bar code for each subject. The lower portion of the figure shows a heat map in which the intensity of red color reflects the degree of correlation (linkage disequilibrium \(LD\)) measured by \(D'\) among SNPs across the region (indicated by tick marks). Note that widely separated SNPs are highly correlated. Two markers associated with type 1 diabetes (and other autoimmune diseases) are shown at the top. Marker rs2476601 is likely to be the causative variant in this region and results in an amino acid change at codon 620. Note that another marker (rs6679677) nearly a distance of 100 kb also strongly associates with diabetes, emphasizing that it is difficult to assign the causative locus on the basis of associations alone when extensive linkage disequilibrium exists in a region.
Table 21-7 Measuring Linkage Disequilibrium

Consider a region of a chromosome with two adjacent loci, A and B. At each locus, there are two possible alleles, 1 and 2. There are four possible combinations of alleles, or haplotypes. These are shown below including a color bar representation and designations for the frequency of these haplotypes in the population.

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁B₁</td>
<td>x₁₁</td>
</tr>
<tr>
<td>A₁B₂</td>
<td>x₁₂</td>
</tr>
<tr>
<td>A₂B₁</td>
<td>x₂₁</td>
</tr>
<tr>
<td>A₂B₂</td>
<td>x₂₂</td>
</tr>
</tbody>
</table>

Designate the allele frequencies at locus A as \( p₁ \) and \( p₂ \).
Designate the allele frequencies at locus B as \( q₁ \) and \( q₂ \).

Then
\[
p₁ = x₁₁ + x₁₂ \\
p₂ = x₂₁ + x₂₂ \\
q₁ = x₁₁ + x₂₁ \\
q₂ = x₁₂ + x₂₂
\]

A simple measure of \( D \) is then calculated as \( D = x₁₁ - p₁ * q₁ \). This value directly measures the difference between the observed haplotype frequency (in this case \( x₁₁ \)) from that expected from random association of the alleles at each locus (in this case \( p₁ * q₂ \)). For any allelic combination in a given dataset, the magnitude of \( D \) will be the same, although the sign (+ or −) may change according to the direction of the allelic association on the haplotypes.

A more standardized measure of \( D \) is more useful in practice and is designated \( D' \). \( D' \) is the ratio of the observed \( D \) to the maximal (or minimal) possible value of \( D \) given the observed allele frequencies.

\[
D' = \frac{D}{D_{\text{max}}} \quad \text{when } D \geq 0 \\
D' = \frac{D}{D_{\text{min}}} \quad \text{when } D < 0
\]

\( D' \) can therefore take on values between 0 and 1, and it is a measure of linkage disequilibrium that is normalized for variations in allele frequencies at the two loci.

Another useful measure of LD is the correlation coefficient, \( r^2 \), between alleles at A and B. Like \( D' \), the value of \( r^2 \) can also vary between 0 and 1. However, unlike \( D' \), the value of \( r^2 \) is a more global measure of how alleles at the two loci are associated and is given by:

\[
r = \frac{D}{\sqrt{p₁p₂q₁q₂}}
\]

When \( r^2 = 1 \), there are only two possible haplotypes, and knowing the allele at locus A is completely predictive of the allele present at locus B. In this case, \( D' = 1 \). In the case of \( r^2 < 1 \), there will be four haplotypes in the population. This is illustrated in the colored two-locus phased haplotype displays shown below.

A presentation of phased haplotypes similar to these is shown for the region around the PTPN22 gene in Figure 21-6.

Beyond PTPN22 itself. Indeed, as indicated in the figure, the SNP marker rs6679677 was used by the Welcome Trust Case Control Consortium (WTCCC) to detect the underlying association of RA with PTPN22 even though this marker is 100kb distant from the causative SNP at rs2476601 within the PTPN22 gene.

Three likely explanations for linkage disequilibrium (LD) exist. First, the population may have originated from a mixture of two populations, one of which had a high frequency of a particular haplotype. If this happened recently, there would not have been time (i.e., a sufficient number of generations) to randomize alleles at closely linked loci by recombination at meiosis. Inasmuch as human history is marked by large population migrations,\(^3\) it is probable that population admixture explains many examples of linkage disequilibrium.

A second explanation, related to the first, rests on the observation that certain regions of the genome tend to exhibit relatively low levels of meiotic recombination for reasons related to the underlying genomic structure. Thus genetic variants within these regions tend to stay together on the same haplotype over many generations, even if haplotypes were introduced into a population in the distant past. The importance of this concept for understanding LD is still under active debate in the genetics community. A third explanation for LD posits that the alleles in linkage disequilibrium may be maintained together because of a selective advantage. For example, going back to the
A1-B8-DR3 haplotype mentioned earlier, one could postulate an advantage for immune defense when alleles on this haplotype are maintained and regulated together in the same individual. Although plausible, this hypothesis is difficult to prove for any particular haplotype.

**Alternatives to the Case-Control Method for Detecting Disease Association**

Family-based controls can also be used for doing genetic association studies. Consider the family shown in Figure 21-7. The affected child carries DR4 and DR3, each of which is inherited from one parent. The laws of mendelian inheritance specify that one DR haplotype from each parent is not inherited by any given offspring—in this example, DR2 in the father and DR5 in the mother. These two non-inherited haplotypes can be thought of as forming a genotype for a “control” individual. In this manner, issues of population stratification (see earlier) are eliminated because patients and controls are sampled from the identical (parental) gene pool. Falk and Rubinstein originally proposed this approach to disease association and called it the haplotype RR method. Its validity depends on a number of assumptions including that the genetic marker under study does not influence mating preference or the production of gametes.

An extension of this approach is called the transmission disequilibrium test (TDT). Using the family in Figure 21-7, for a given heterozygous parent (such as the father carrying DR2, 4), there is a probability of 0.5 that any given allele, such as DR4, will be transmitted to the child. If the DR4 allele has no bearing on disease risk, the probability of transmission (T) to an affected child is equal to the probability of nontransmission (NT). This can be stated simply as \( P(T|D) = P(NT|D) \), in which \( D \) indicates the presence of disease in the offspring. However, if the allele being examined is associated with disease risk, then \( P(T|D) > P(NT|D) \). If large numbers of heterogeneous parents with affected offspring are examined, transmission disequilibrium testing can establish an association between disease and the test allele compared with the (noninherited) control alleles. Although the TDT is an elegant solution to getting around the issue of population stratification, it is generally more expensive to collect and genotype trio families and the TDT approach to association and called it the haplotype RR method. Its validity depends on a number of assumptions including that the genetic marker under study does not influence mating preference or the production of gametes.

Against this background of basic information about DNA variation, there has been sweeping technologic change, so it is now rather routine to obtain data on hundreds of thousands or even millions of SNP genotypes on many individuals in a matter of days, and sequencing of the entire human genomes of large numbers of subjects has fast become a reality at sustainable costs. This in turn has provoked the development of sophisticated statistical approaches to large datasets, at the same time highlighting the need to integrate the genetic knowledge with clinical data in ways that are manageable and meaningful to researchers and ultimately to practitioners. Thus although the basic principles of association and linkage disequilibrium discussed earlier in the context of the HLA are still valid, the imminent availability of massive genetic datasets now mandate new, more efficient ways to store and analyze the data.

Another unsettled issue concerns the overall genetic “architecture” of human disease. Until recently, there has been an assumption that common allelic variants are likely to account for a large portion of the genetic risk for autoimmunity in the population. The HLA alleles that associate with rheumatic diseases are examples of this, as are several of the new associations with RA and other autoimmune disorders such as PTNP22. By common variants, we generally mean variants that are present in the population at frequencies 5% or more, and certainly not less than 1%. However, there is no a priori reason to reject the hypothesis that many rare variants actually account for a significant fraction of the genetic burden of disease. The main reason that common variants have been a focus of research is because the current technologies are particularly well suited to investigate them. This is about to change, with the advent of new technologies that permit resequencing on a massive scale; routine sequencing of entire individual human genomes is probably not more than 5 years away. A seminal finding with respect to the importance of rare variants involves regulation of plasma levels of

---

**Figure 21-7** Family structure utilized for determination of haplotype relative risk. The affected child carries the DR3 and DR4 alleles. A fictitious “control” individual can be constructed from the noninherited DR2 and DR5 alleles and used for a relative risk calculation. Such families can also be utilized for transmission disequilibrium testing, as discussed in the text.

---

**GENETICS IN THE ERA OF WHOLE GENOME ANALYSIS**

As pointed out in the introduction, there has been an explosion in new genetic knowledge over the past few years and this is likely to continue for some time. A major foundation was laid by the completion of the Human Genome Project, which provided a draft reference sequence for the entire 3.2 billion base pairs of the human genome. Subsequently, the human HapMap Project has provided a catalog of the common sequence variation across the human genome. There are now more than 30 million SNPs referenced for Homo sapiens in dbSNP, the online SNP reference database. In addition to SNPs, there are tens of thousands of variable numbers of tandem repeats (VNTRs), a common form of which is termed a microsatellite. Microsatellites have been used extensively for linkage analysis over the past decade. Large insertions and deletions are also quite common across the genome, some of which arise de novo, providing yet another source of individual genetic variation that can be associated with human disease. Smaller insertions and deletions are just beginning to be cataloged, and thus the number of these genetic variants will grow substantially as more human genomes are fully sequenced.

As pointed out in the introduction, there has been an explosion in new genetic knowledge over the past few years and this is likely to continue for some time. A major foundation was laid by the completion of the Human Genome Project, which provided a draft reference sequence for the entire 3.2 billion base pairs of the human genome. Subsequently, the human HapMap Project has provided a catalog of the common sequence variation across the human genome. There are now more than 30 million SNPs referenced for Homo sapiens in dbSNP, the online SNP reference database. In addition to SNPs, there are tens of thousands of variable numbers of tandem repeats (VNTRs), a common form of which is termed a microsatellite. Microsatellites have been used extensively for linkage analysis over the past decade. Large insertions and deletions are also quite common across the genome, some of which arise de novo, providing yet another source of individual genetic variation that can be associated with human disease. Smaller insertions and deletions are just beginning to be cataloged, and thus the number of these genetic variants will grow substantially as more human genomes are fully sequenced.

As pointed out in the introduction, there has been an explosion in new genetic knowledge over the past few years and this is likely to continue for some time. A major foundation was laid by the completion of the Human Genome Project, which provided a draft reference sequence for the entire 3.2 billion base pairs of the human genome. Subsequently, the human HapMap Project has provided a catalog of the common sequence variation across the human genome. There are now more than 30 million SNPs referenced for Homo sapiens in dbSNP, the online SNP reference database. In addition to SNPs, there are tens of thousands of variable numbers of tandem repeats (VNTRs), a common form of which is termed a microsatellite. Microsatellites have been used extensively for linkage analysis over the past decade. Large insertions and deletions are also quite common across the genome, some of which arise de novo, providing yet another source of individual genetic variation that can be associated with human disease. Smaller insertions and deletions are just beginning to be cataloged, and thus the number of these genetic variants will grow substantially as more human genomes are fully sequenced.

As pointed out in the introduction, there has been an explosion in new genetic knowledge over the past few years and this is likely to continue for some time. A major foundation was laid by the completion of the Human Genome Project, which provided a draft reference sequence for the entire 3.2 billion base pairs of the human genome. Subsequently, the human HapMap Project has provided a catalog of the common sequence variation across the human genome. There are now more than 30 million SNPs referenced for Homo sapiens in dbSNP, the online SNP reference database. In addition to SNPs, there are tens of thousands of variable numbers of tandem repeats (VNTRs), a common form of which is termed a microsatellite. Microsatellites have been used extensively for linkage analysis over the past decade. Large insertions and deletions are also quite common across the genome, some of which arise de novo, providing yet another source of individual genetic variation that can be associated with human disease. Smaller insertions and deletions are just beginning to be cataloged, and thus the number of these genetic variants will grow substantially as more human genomes are fully sequenced.
lipoproteins,91,94 and examples of rare variants contributing to autoimmune diseases are beginning to emerge.95-98 Thus we are still relatively early in the whole genome era of genetic mapping.

**Estimating the Size of the Genetic Contribution to Rheumatic Diseases**

Epidemiologic studies of familial aggregation and twin concordance for autoimmune diseases are an important background for assessing the genetic contribution to these disorders. The general approach is to compare disease prevalence in groups of individuals with different degrees of genetic relatedness. The most common groups used for such studies are genetically identical individuals (monozygotic [MZ] twins) and individuals who share approximately 50% of their genetic variation in common (dizygotic [DZ] twins and siblings), compared with “unrelated” individuals in the background population. Of course, all humans are related in the sense that we descend from common ancestors, but randomly selected individuals in the population generally differ by little more than 0.1% of their genomes of 3.2 billion base pairs, or 3 to 4 million base pairs.

The use of such family and background population prevalence data to calculate risk ratios remains a primary method of estimating the overall size of the genetic component in complex diseases.99 For example, one can estimate the RR of disease for siblings of affected individuals compared with the general population. This leads to a value called $\lambda_s$, or RR to sibs, which is calculated as follows:

$$\lambda_s = \frac{\text{Disease prevalence in siblings of affected individuals}}{\text{Disease prevalence in general population}}$$

Obtaining a reliable value of $\lambda_s$ depends on having accurate estimates of disease prevalence in the two comparison groups. This is not a trivial matter. A firm diagnosis of autoimmune disease is difficult to make in large population surveys, with errors in both directions possible. Underestimation may occur because of the lack of reporting of disease that is no longer active. Overestimation may result from inadequate distinction between different forms of autoimmune disease. Despite these difficulties, a range of values for $\lambda_s$ has been established for many of the common autoimmune disorders.100 Table 21-8 lists representative examples. With the exception of ankylosing spondylitis, most autoimmune disorders appear to have a $\lambda_s$ in the range of 10 to 20.

Table 21-8 also shows the estimates for $\lambda_{MZ}$. Analogous to the calculation for $\lambda_s$, $\lambda_{MZ}$ is calculated by the following:

$$\lambda_{MZ} = \frac{\text{Disease prevalence in identical (MZ) co-twins of affected individuals}}{\text{Disease prevalence in general population}}$$

The value $\lambda_{MZ}$ can be interpreted as an estimate of the maximal genetic risk for the disease. This assumes that all the increased risk to an MZ co-twin results from the fact that they share the same genetic polymorphisms. This assumption is clearly not entirely correct because at least some environmental sharing probably contributes to the risk. However, it is notable that the estimated value for $\lambda_{MZ}$ is quite high for many autoimmune disorders, and this emphasizes the fact that MZ twin concordance rates must be interpreted in light of the background population prevalence of the disease. Thus a relatively low MZ twin concordance rate does not necessarily imply a low genetic component to the disorder. Of course, the absence of complete concordance for disease among MZ twins indicates the need to include environmental, developmental, and stochastic factors in order to fully understand the role of genetics.

**Screening the Entire Genome for Disease Genes: Approaches Based on Linkage**

There are two basic methods of whole genome analysis, one based on linkage and the other based on association.101 Linkage methods depend on the ability to track polymorphic genetic markers in families and to show that these genetic markers co-segregate with the disease phenotype in families where there are multiple members affected. Thus multiplex families are required for linkage analysis. The details of the statistical methods are complex but are generally based on examining the likelihood of a particular pattern of co-inheritance of marker and disease (linkage), compared with the likelihood that there is no linkage (the null hypothesis). A measure of this likelihood is referred to as the LOD (log of the odds) score, with a LOD score greater than 3 generally interpreted to indicate significant evidence of linkage when markers across the entire genome are examined. Ott102 described linkage methods in detail in 1999. One advantage of linkage analysis is that the entire genome can be effectively interrogated using less than 500 informative microsatellite markers, or around 5000 SNP markers.87 Linkage analysis has been applied with great success to the analysis of rheumatic diseases that exhibit a clear mendelian pattern of inheritance (e.g., dominant or recessive). In 1992, familial Mediterranean fever was mapped to chromosome 16,103 and this led to the identification of the gene for this disease in 1997.104 In addition, an entirely new class of familial periodic fever syndromes has been localized to mutations in the TNF receptor 1 gene on chromosome 12.105,106 Thus for highly penetrant mendelian disorders, classical linkage analysis is a powerful means of identifying the underlying molecular basis of disease. However, one of the drawbacks to classical linkage analysis is that, to be most useful, it should be applied to disorders with a high penetrance and a known genetic model (e.g., dominant vs.

### Table 21-8 Familial Clustering of Selected Autoimmune Diseases, as Measured by the Relative Risk to Siblings ($\lambda_s$) and Monozygotic Twins ($\lambda_{MZ}$)

<table>
<thead>
<tr>
<th>Disease</th>
<th>$\lambda_s$</th>
<th>$\lambda_{MZ}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1 diabetes mellitus</td>
<td>15</td>
<td>60</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>3-10</td>
<td>20-60</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>20</td>
<td>250</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>20</td>
<td>250</td>
</tr>
<tr>
<td>Ankylosing spondylitis</td>
<td>54</td>
<td>500</td>
</tr>
</tbody>
</table>

recessive). An alternative approach based on linkage, broadly called allele sharing, is preferred for the study of autoimmune diseases that have a complex genetic basis.\textsuperscript{101}

The most common approach to allele sharing is the affected sibling pair (ASP) method. This method is based on a simple question: When two sibs are both affected with a disease, do they share alleles at particular genetic markers more frequently than would be expected by chance? Figure 21-8 illustrates this basic approach. In this family two siblings are affected, and the first-born sibling (sib 1) has inherited alleles 1 and 3 at a marker locus, X. By the laws of mendelian inheritance, sibling 2 has a 25% chance of inheriting these same two alleles and has a 25% chance of inheriting neither of these alleles (i.e., sib 2 inherits 2,4 and shares nothing with sib 1 at locus X). By a similar reasoning, there is a 50% chance that these two siblings will share one allele in common. This 25:50:25 distribution of sharing 0, 1, or 2 haplotypes is expected if there is no linkage between the disease and the marker locus. However, if a gene that lies near the marker locus is involved in disease risk, a significant deviation toward increased sharing among affected siblings will be observed. The closer the marker is to the disease locus, the greater the deviation will be from a 25:50:25 distribution. By examining large numbers of affected sibling pairs in this manner, the investigator can develop statistical evidence that this is the case using a standard $\chi^2$ analysis, with the null hypothesis being that there is no increased sharing at the marker locus.

ASP analysis has a number of distinct advantages and disadvantages. Only affected individuals are used, and the problem of falsely assigning a family member as “unaffected” is eliminated. This is a major issue for diseases such as RA because the disease may not express itself until later in life. ASP analysis can be done without committing to a specific model of inheritance (i.e., recessive or dominant). As with linkage in general, the ASP methods suffer from having relatively low power to detect genes that confer only modest risk.\textsuperscript{101} This means that quite large numbers (hundreds or thousands) of multiplex families are required to obtain statistically significant results.\textsuperscript{101}

Linkage analysis based on allele sharing began to be applied to genetically complex rheumatic and autoimmune diseases in the 1990s and has resulted in a few successes, most notably the identification of the NOD2 gene as a major risk factor for Crohn’s disease.\textsuperscript{107,108} Linkage analysis has provided evidence for numerous risk genes involved in lupus.\textsuperscript{109} The presence of a linkage peak (LOD score > 3.5) on chromosome 2q in RA sibling pairs\textsuperscript{10} led to the identification of STAT4 as a risk gene for RA, as well as lupus.\textsuperscript{111} Thus although challenging to carry out, linkage can occasionally be applied successfully to complex diseases that do not have a clear mendelian pattern of segregation.

### Screening the Entire Genome for Disease Genes: Approaches Based on Association

Compared with linkage analysis, association methods have much greater statistical power to detect genetic effects.\textsuperscript{101} However, in contrast to the relatively modest numbers of markers used for linkage studies, hundreds of thousands or even millions of genetic markers are currently utilized to carry out a comprehensive genome-wide study using association methods. This is the basis of the so-called genome-wide association study, or GWAS. For scanning the whole genome, GWAS has generally replaced linkage methods except in special situations in which large multiplex families suggest the presence of a mendelian pattern of inheritance.

The success of whole genome association studies is critically dependent on taking advantage of the underlying haplotype structure of the genome, which in turn reflects the ubiquitous presence of linkage disequilibrium across the genome, as discussed earlier. Thus as Figure 21-6 shows, for the region around the PTPN22 gene on chromosome 1, common variation in large segments of most genetic regions can be interrogated using just a few markers that “tag” the common haplotypes. One can think of these patterns of haplotype variation a kind of bar code of common haplotype across a genomic region, as shown in the central region of Figure 21-6 and described in Table 21-7. As mentioned, the International HapMap Project has established online resources (www.hapmap.org) that allow users to easily explore the haplotype structure of any region of the genome.\textsuperscript{112} The HapMap website also contains informative tutorials on genetic diversity and the use of this resource.

In 2005, the complement regulatory protein Factor H was identified as a significant risk factor for age-related macular degeneration,\textsuperscript{113,114} and this report essentially kicked off the GWAS era. Shortly after, a second major GWAS study led to the identification of the IL-23 receptor as a risk gene for Crohn’s disease.\textsuperscript{115} In the intervening years, GWAS studies have proven to be particularly fruitful when applied to autoimmune disorders. In the aggregate, there are now nearly 200 distinct chromosomal regions that have been identified as containing risk loci for major autoimmune diseases including type 1 diabetes,\textsuperscript{116} RA,\textsuperscript{117} systemic lupus,\textsuperscript{118} and inflammatory bowel disease.\textsuperscript{119} As large datasets are assembled for some of the rarer or less studied autoimmune disorders,\textsuperscript{120,121} GWAS will continue to play a role in genetic analysis.
Overlapping Susceptibility Genes and Pathways for Autoimmune Diseases: Getting to Function

Obviously, from the previous discussion, the genetic factors underlying autoimmunity are quite complex and the full picture is still emerging. Nevertheless, some themes are evident from the current data. First, many of the newly discovered risk genes appear to be involved in multiple different autoimmune disorders. In contrast to HLA, this often seems to involve the same alleles conferring susceptibility across many diseases, as exemplified by PTPN22 (see later). Secondly, many of the risk loci fall into functional categories that suggest common underlying mechanisms. Of course, the assignment of a gene to a particular functional category is subject to bias, and the same gene can be found in multiple categories. Nevertheless, many genes involved in lymphocyte signaling, cytokine pathways, innate immune pathways, and interferon pathways are found associated with autoimmunity. These have been discussed in several recent reviews, and a full discussion of these data is beyond the scope of this chapter. However, several examples that highlight these trends are provided as follows.

In 2005, an association between the intracellular phosphatase, PTPN22, was reported for a number of autoimmune diseases including type 1 diabetes, RA, systemic lupus, and autoimmune thyroid disease. With ORs consistently in the range of 1.5 to 2, this was the first compelling demonstration that a specific common allelic variant outside of HLA can confer risk for multiple different autoimmune phenotypes and has been consistently replicated. In this case, a nonsynonymous change in one of several SH3 binding sites in PTPN22 (a tryptophan substitution for arginine at codon 620) was shown to disrupt the normal association of PTPN22 with a Csk, an intracellular tyrosine kinase. Knockout of PTPN22 in rodents leads to dramatic overactivity of T cells, but the exact functional consequence of the human risk allele remains controversial. However, there is little doubt that PTPN22 is involved in setting thresholds for T cell receptor signaling through Lck, as well as B cell receptor signaling. PTPN22 is also found in many other hematopoietic cells, and its function in these cells is largely unknown. Interestingly, all of the autoimmune diseases associated with PTPN22 have a prominent autoantibody component, and thus a specific role in regulating humoral immunity appears to be a unifying feature. Notably, the PTPN22 risk allele is actually protective for Crohn’s disease and has no role in risk for multiple sclerosis, emphasizing the likely presence of distinct mechanisms of pathogenesis in these disorders. Thus PTPN22 is a prime example of how the discovery of new disease associations with relatively modest effect can redirect hypothesis-driven research into new pathways.

In 2005 researchers showed that interferon regulatory factor 5 (IRF5) was associated with susceptibility to systemic lupus, and they replicated it shortly thereafter. This was a satisfying observation because activation of interferon pathways is clearly central to the pathogenesis of lupus and related disorders. Since these original observations, it is now apparent that multiple genes in interferon pathways are involved in lupus susceptibility. Interferon has emerged as a potential drug target, and there is renewed appreciation and interest in the role of interferon regulation in the immune response generally. The involvement of multiple genes in interferon pathways as risk factors for autoimmune diseases provides important support for continuing biologic studies in this field, as well as the potential for new insights into the details of how this pathway is regulated.

The theme of common variants in genes that share susceptibility alleles across different diseases with common underlying mechanisms begs the question of how these genetic variants affect the normal immune response. Do these genetic variants confer survival advantages? Could one examine the “normal” range of immune response phenotypes to detect a state of risk, without actually defining the complex underlying genetics? The entire issue of what is the normal range of immune function has come under renewed interest and has driven interesting studies of function of autoimmune risk loci in subjects without disease. Indeed, the study of the normal immune system may be the key to understanding how complex arrays of risk alleles actually lead to disease in an individual. Thus in addition to further expansion of gene discovery efforts, advances in the genetics of autoimmunity are going to require more sophisticated ways of phenotyping the normal immune system, as well as the diseases themselves.

FUTURE CHALLENGES OF GENETICS IN RHEUMATIC DISEASES

This chapter has discussed fundamental concepts and analytic approaches that are central to modern human genetics. Although the underlying concepts will not change, it is obvious that genetic technologies are rapidly advancing, with routine analysis of the full sequence of entire human genomes only a few years away. We are already in an era where the available genetic data are exceeding our capacity to analyze them. Therefore the future challenges are likely to be focused mainly on extracting the meaning of genetic data in a way that is useful, on many levels. We need to connect genetic variation to phenotypic variation from the most basic levels of biologic action to systems regulation and finally to disease pathogenesis and heterogeneity. As emphasized in the last section, this implies a major emphasis on more precise phenotypic measurements both in vitro and in vivo including assessment of population diversity. This is going to be an iterative and integrative process; it is only in the setting of relevant (and measurable) phenotypic diversity that the significance of genetic variation can be understood.

Aside from the basic problem of functionally connecting genotype to phenotype, several areas are poised for expansion. One of these is epigenetics. This concerns the varied mechanisms by which the genome can be chemically altered to affect gene function, and this will surely require a separate chapter by the next edition of this textbook. These alterations can be induced by parental effects, developmental events, and environmental influences, and working out the details of these interactions is sure to bring surprises and provide us with a new appreciation for the complexity of gene-environment interactions. Finally, the complexity of the genetic factors involved in rheumatic diseases makes it unlikely that there will be a rapid translation of genetic information into practical action at the level of patient care.
in the near future. Useful algorithms will almost certainly require combining genetic information with other biomarkers, both for diagnosis and risk assessment, as well as treatment decisions. Here again, genotypic and phenotypic information need to be joined to achieve long-term goals.

**Selected References**


Full references for this chapter can be found on www.expertconsult.com.


The term *epigenetic* describes a number of chromatin modifications that stably alter gene expression without changing the sequence of the DNA. In this way, epigenetic mechanisms shape the phenotype of a cell without changing the genotype. In the last few years, the study of epigenetic mechanisms strongly progressed, yielding insight into the mechanisms by which epigenetic chromatin modifications regulate gene expression. It became clear that epigenetic mechanisms are important in cellular development and differentiation, as well as in adaptation to changes in the environment. In particular, Conrad Waddington is given credit for advancing the hypothesis that the environment plays a sculpting role in development and cell fate via epigenetic mechanisms. Epigenetics supports the idea that changes in gene expression that stem from long-term exposure to a certain signal get imprinted, become independent of the activating stimulus, and persist even in its absence. Accordingly, the impact of nutrition, toxins, or infection on gene expression is maintained, leading to fixed changes in the phenotype. This change helps the cell to adapt but might also promote pathologic cellular behavior.

With the acquisition of new knowledge, the original definition of *epigenetic processes* as heritable changes in DNA without alterations in the sequence became a matter of debate. Particularly divisive is the matter of heritability. Even though the classic definition demands that epigenetic modifications persist through mitosis and meiosis (i.e., are inherited to the next generation), transient modifications that indirectly initiate stable changes more and more are put together under the term *epigenetics*. Giving credit to this view, Bird described epigenetic events as “the structural adaption of chromosomal regions so as to register, signal, or perpetuate altered activity states.” In addition to the well-established epigenetic role of DNA methylation, this definition includes a variety of more transient histone modifications such as acetylation, methylation, or phosphorylation that underlie epigenetic effects, and that will be discussed in this chapter along with the influence of SUMOylation, ubiquitination, adenosine diphosphate (ADP) ribosylation, and microRNA.

**EPIGENETIC MECHANISMS**

**DNA Methylation**

DNA methylation is the classic and probably best studied epigenetic modification. Thereby, methyl marks are added to the cytosine ring of cytosine-phosphatidyl-guanine (CpG) dinucleotides. In addition, recently also non-CpG, but CHG or CHH, where H stands for A, C, or T, DNA methylation has been found to occur. This type of DNA methylation seems to be important mainly in pluripotent stem cells and disappears during differentiation. CpG methylation of DNA preferably is found at sites with a high percentage of CpGs, so-called CpG islands. Recently however, differences in the methylation pattern between tissues were found to occur frequently not at CpG islands, but at regions with lower CpG density that lie around 2 kb upstream or downstream of CpG islands, so-called CpG island shores. At both locations, DNA methylation leads to transcriptional repression.

During embryonic development, de novo methylation marks are established by DNA methyltransferase (DNMT) 3a and 3b. During somatic cell replication, methylation marks are stable and are maintained by DNMT1. Gene silencing by DNA methylation is crucial for the formation of heterochromatin. It promotes stable silencing of repetitive sequences such as alu sequences and retrotransposons, genomic imprinting, X chromosome inactivation in female mammals, and tissue-specific gene expression. However, in recent years it has become clear that DNA methylation is more dynamic than was previously assumed. Global genomic demethylation is seen in the paternal pronucleus shortly after fertilization, and gene-specific demethylation has been observed. Promoters of estrogen-responsive genes, for instance, are periodically methylated and demethylated depending on ligand binding of the estrogen receptor α.
Methylation marks are believed to be removed passively (e.g., DNMT1 does not restore methyl marks after replication) or actively. Different mechanisms have been proposed to account for active DNA demethylation; however, as yet definite experimental proof in mammalian cells exists for none of them. In short, conceivable mechanisms for active DNA demethylation include enzymatic removal of the methyl group, excision and substitution of 5-meC by C, oxidative demethylation thereby converting 5-meC to 5-hydroxymethylcytosine, and demethylation by S-adenosylmethionine (SAM) domains, forming a 5-meC radical.

DNA methylation can directly interfere with binding sites of transcription factors, thereby inhibiting induction of gene expression, or transcriptional repressors can bind to methylated DNA, blocking transcription. By their presence, they inhibit the binding of transcription factors and recruit histone deacetylases, which change the histone code and further block gene expression (Figure 22-1).

**Histone Modifications**

Within the nucleus, DNA is packaged around an octamer comprising homodimers of four different histone proteins: H2A, H2B, H3, and H4. This so-called nucleosome is stabilized and kept in place by H1. Each histone protein has an N-terminal tail, which is the main target of a variety of histone modifications. Amino acids in the histone tails can be acetylated, methylated, phosphorylated, ubiquitinated, SUMOylated, ADP ribosylated, and deiminated (Figure 22-2). Not all of these modifications are equally well studied, and detailed knowledge about their regulation and interactions is still missing. The sum of the different histone modifications at a distinct region of the genome is called the histone code. It is generally believed that the histone code can influence transcriptional activity directly by affecting chromatin structure, thereby making it more or less accessible for transcription factors and indirectly leading to the attraction of effector molecules that in turn recruit and stabilize the transcription machinery. For most histone modifications, it is yet impossible to predict whether they enhance or restrict transcriptional activity and how this will influence the biologic output. Also, more and more accumulating evidence indicates that the different histone modifications are strongly interconnected and can enforce or counteract each other. Nevertheless, some modifications can clearly be associated with transcriptionally active or inactive chromatin (Table 22-1).

**Acetylation**

Histone acetylation generally is associated with transcriptional activation. It is believed that histone acetylation directly opens the chromatin structure, allowing easier access to the transcription machinery. Acetylation loosens the interaction of the negatively charged DNA with the positively charged lysine by neutralizing the charge of lysine. In addition, acetylated histone tails are bound by bromodomain effectors. Members of the bromodomain family support protein-protein interactions to assemble a multicomponent complex involved in transcriptional activation. Histone acetylation is catalyzed by histone acetyltransferases (HATs) and is counteracted by histone deacetylases (HDACs). Up to now, two main groups of HATs—type A and type B—have been described, whereby the nuclear type A HATs are those believed to influence transcriptional activity. The family of HDACs comprises four classes. Classes I, II, and IV consist of 11 HDACs—HDAC1 to HDAC11—that are zinc dependent;
class III includes the nicotinamide adenine dinucleotide (NAD)-dependent sirtuins, SIRT1 to SIRT7. HDACs and HATs not only modify histones, they also form complexes with transcription factors, oncoproteins, and tumor suppressors and regulate their activity by changing the acetylation status.14

Methylation

Histone methylation marks are highly complex because histones can be monomethylated, dimethylated, or trimethylated; in addition to lysine, arginine residues can be modified. Depending on the methylated site, histone methylation leads to transcriptional activation or repression. A prominent example of this divergent effect is trimethylation of histone 3 at the fourth lysine (H3K4me3), which is associated with active genes, in contrast to H3K27me3, which is found at silenced genes. It is interesting to note that the location of the methyl mark in relation to the regulated gene plays a role. Thus H3K9 methylation at the promoter region silences transcription, but within the coding region, H3K9 methylation was found to be associated with actively transcribed genes.17 Similar to histone acetylation, methylation of histone tails changes chromatin configuration and binds effector proteins. Methyl marks are recognized by plant homeodomain (PHD) fingers and the Royal-superfamily of chromatin-binding proteins, including chromodomain, tudor, and MBT domains. Histone methylation is catalyzed by histone methyltransferases (HMTs). The various members of this group of enzymes are highly specialized in methylating selected lysines or arginines to a certain degree (monomethylation, dimethylation, trimethylation). Lysine demethylases that counteract the action of HMTs on methylated lysines may contain an amine oxidase or a Jumonji C (JmjC) domain. Up to now, no arginine demethylase has been found that can reverse methylated arginine to arginine. Peptidylarginine deiminase 4 (PAD4) converts methylated arginine to citrulline, and this leads to transcriptional repression.18 Because in this step also the imine group of the arginine is removed, this reaction is termed deimination.

Phosphorylation

Histone phosphorylation is a fast way to induce gene expression, particularly after activation of mitogen-activated protein kinase (MAPK) pathways.19 Even though histone phosphorylation itself seems to be rather transient, reports suggest that histone phosphorylation influences other histone modifications such as acetylation and methylation, thereby indirectly inducing prolonged changes in the histone code.2021 Phosphorylation of histones occurs at serine and threonine residues and is mediated by kinases and removed by phosphatases. Similar to acetylation, phosphorylation opens the chromatin structure by adding a negative charge to the histone but also leads to the attraction of effector proteins. The best studied effector proteins binding phosphorylated histones are 14-3-3 proteins and BRCA1 C-terminus (BRCT) domains.9

SUMOylation

E1 SUMO-activating enzymes, E2 SUMO-conjugating enzymes, and E3 SUMO ligases are the enzymes that add small ubiquitin-related modifiers (SUMOs) to histone lysines.23 SUMOylation can be reversed by isopeptidases. Even though details about gene regulation after histone SUMOylation are still unclear, it is known that SUMOylation mostly represses gene expression in close cooperation with other histone modifications.24 As a counterregulatory mechanism, SUMO marks are increased after histone acetylation and recruit HDACs to remove acetylation.24 Furthermore, SUMOylated histones associate with heterochromatin protein (HP)-1, a methyl-lysine binding protein that interacts with HMTs to repress transcription.25

---

Table 22-1  Epigenetic Modifications

<table>
<thead>
<tr>
<th>Modification</th>
<th>Location</th>
<th>Effect</th>
<th>Mechanism</th>
<th>Enzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylation</td>
<td>Cytosine</td>
<td>Repression</td>
<td>Blocking of transcription factor binding</td>
<td>+DNMT3a/b (de novo), DNMT1 (maintenance)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Recruitment of effectors</td>
<td></td>
</tr>
<tr>
<td>Methylation</td>
<td>Lysine, arginine</td>
<td>Activation/repression</td>
<td>Chromatin structure</td>
<td>+HMT</td>
</tr>
<tr>
<td>Acetylation</td>
<td>Lysine</td>
<td>Activation</td>
<td>Recruitment of effectors</td>
<td>−HDM</td>
</tr>
<tr>
<td>Phosphorylation</td>
<td>Serine, threonine</td>
<td>Activation</td>
<td>Chromatin structure</td>
<td>+HAT</td>
</tr>
<tr>
<td>Ubiquitination</td>
<td>Lysine</td>
<td>Activation/repression</td>
<td>Recruitment of effectors</td>
<td>+HDAC</td>
</tr>
<tr>
<td>SUMOylation</td>
<td>Lysine</td>
<td>Repression</td>
<td>Recruitment of effectors</td>
<td>+PAD4</td>
</tr>
<tr>
<td>ADP ribosylation</td>
<td>Glutamic acid</td>
<td>Activation</td>
<td>?</td>
<td>+Kinas</td>
</tr>
</tbody>
</table>

DNMT, DNA methyltransferase; HAT, histone acetyltransferase; HDAC, histone deacetylase; HDM, histone demethylases; MART, mono-ADP-ribosyltransferases; PAD4, peptidyl arginine deiminase type 4; PARP, poly–adenosine diphosphate (ADP)-ribosyltransferases; SUMO, small ubiquitin-related modifier.
Ubiquitination

Ubiquitin ligases and proteases control histone ubiquitination and deubiquitination, respectively. Ubiquitination of histone 2A and 2B generally activates gene transcription. However, other reports show that histone ubiquitination can repress transcriptional activity. Similar to SUMOylation, ubiquitination is strongly connected to other histone modifications and mainly alters gene expression via recruitment of effector molecules, which subsequently results in histone methylation.

ADP Ribosylation

Histones can be mono-ADP ribosylated by mono-ADP ribosyltransferases (MARTs) or poly-ADP ribosylated by poly-ADP ribosyltransferases (PARPs). Little is known about this histone modification, and neither site specificity nor function has been clearly defined. However, synergy between histone acetylation and ADP-ribosylation has been described.

Epigenetics and Noncoding RNAs

Noncoding RNAs (ncRNAs) are transcribed RNA molecules that do not translate into protein. Long known ncRNAs include transfer RNA (tRNA) and ribosomal RNA (rRNA). More interesting with regard to epigenetic modifications, however, are the more recently discovered long ncRNAs such as the Xist (X inactive specific transcript) transcript and small ncRNAs such as microRNAs (miRs). In female cells, Xist is transcribed from the later inactivated X chromosome (Xi). It specifically binds to Xi and promotes its silencing by histone ubiquitination, methylation, and loss of acetylation. Recently, another long ncRNA named Air has been shown to interact with a specific set of genes and to recruit the chromatin-modifying machinery to these loci. In plants and yeast, multiple examples of interaction between ncRNAs and epigenetic chromatin modifications have been found. Overall these data suggest that recognition of complementary DNA sequences by ncRNAs might guide epigenetic modifications to specific genomic loci.

MicroRNAs are small ncRNAs that are encoded in intergenic regions and overlapping introns or exons of other genes. The primary miR transcript (pri-miR) containing the information of a single miR or an miR cluster is further processed by an enzymatic complex that includes the RNase III enzyme Drosha. After formation of a hairpin, the precursor miR (pre-miR) is exported to the cytoplasm. Pre-miRs are further cleaved by the RNase III enzyme Dicer and are released as short duplexes, on average 22 oligonucleotides long, which constitute the mature miR. From this duplex, two mature miRs can be formed. However, dependent on the thermodynamic stability of each end, in some cases one of the strands is fast degraded; this is usually called the miR star (*) or passenger strand. The other strand, termed the guide strand, is more stable and biologically active. In other cases, both strands are equally stable, and the miR from the 5’ end is referred to as 5p, whereas the miR from the 3’ end is denominated 3p. After incorporation into the RISC (RNA-induced silencing complex), mature miRs bind to the 3’ untranslated region of their target mRNA and modulate their expression by degradation or by inhibition of protein translation. The system is highly promiscuous, so that one miR can modulate expression of multiple proteins, and expression of one protein is influenced by a variety of different miRs. Mainly the miR sequence from base 2 to base 8 or 9 from the 5’ end, the so-called seed region, plays a crucial role in target recognition. In addition, the position of the complementary sequence on the gene and the presence of other miR target sites influence the repressive action of miRs. The expression of around 30% of all genes is believed to be regulated via miRs.

Recently more and more connections between miRs and epigenetics have been found. It is estimated that 50% of the genes coding for miRs are directly controlled by DNA methylation. A comparison between methylation patterns of various tissue types revealed that a large proportion of differentially methylated regions are located outside promoters regions in conserved intergenic elements, where miRs are coded. In cancer, these regions are often found to be affected by genomic changes such as deletions, translocations, and amplifications, leading to disturbances in the miR expression pattern.

As epigenetic mechanisms influence the expression of miRs, miRs regulate the expression of important components of the epigenetic machinery. By targeting chromatin-modifying enzymes, miRs subtly change the epigenome, thereby regulating gene expression. Several miRs have been shown to target DNMT1, thereby promoting loss of DNA methylation. Also, HDACs are targets of miRs, making them potential modulators of histone acetylation.

EPIGENETICS IN DISEASE

In the study of complex diseases, one of the main tasks is to unravel genetic and environmental influences in the search for causative factors and prediction parameters. Twin studies comparing concordance rates of diseases in monozygotic twins can give insight into genetic and environmental influences of a disease. For example, a predominance of genetics is suggested in celiac disease, in which concordance rates between monozygotic twins are greater than 80%. In rheumatoid arthritis, however, heritability accounts for only maximal 60% of disease prevalence, suggesting a strong impact of nongenetic factors on disease development. Changes in epigenetics might explain this “missing heritability.” In particular, the impact of environmental factors on epigenetics has raised considerable attention in the last few years. Differences in the epigenome have been detected in monozygotic twins. They accumulate over time and might be one of the reasons why one twin develops a certain disease while the other stays healthy. It has been proposed that the epigenomes of monozygotic twins diverge more and more because of nonshared environmental influences. Also, the fidelity of the enzymes maintaining epigenetic modifications during mitosis is much lower compared with the DNA polymerases, leading to small changes in epigenetic modifications after each cell cycle.

Alterations in the epigenetic code have been found in a variety of pathologic conditions, ranging from cancer to autoimmune and chronic inflammatory diseases.
psychiatric conditions. However, it is not clear how far these changes are causative of disease, or whether they are bystander effects. In some cases, defects in the epigenetic code have been found to directly cause disease. Particularly, specific imprinting control regions have been found to be sensitive to changes in DNA methylation. About 80 genes in the human genome are currently known to be regulated via genomic imprinting. In these genes, epigenetic silencing of the maternal or paternal allele is of importance for balanced gene expression. Naturally caused by the haploid expression of genomically imprinted genes, the consequences of aberrant imprinting include loss of function, whereas loss of imprinting leads to highly increased levels of the gene product, both of which induce developmental disturbances and disease. For example, Beckwith-Wiedemann syndrome, which is characterized by macrosomia, increased tumor risk, and facial dysmorphia, can be caused by loss of maternal allele-specific DNA methylation at one specific imprinting control region, resulting in increased expression of genomic imprinted genes, or by hypermethylation of the maternal allele at another region, leading to loss of expression. Other examples of diseases associated with dysregulated genomic imprinting are Angelman syndrome and Prader-Willi syndrome.

Most studies elucidating the role of epigenetic changes in disease have focused on cancer. Genome-wide hypomethylation is the most commonly found epigenetic alteration in many tumor types, but changes in histone acetylation and methylation have been demonstrated. DNA hypomethylation can lead to activation of retrotransposons such as endogenous retroviruses, long interspersed nucleotide elements (LINE)-1, and alu repeats. These changes induce chromosome instability and influence the transcriptional activity of nearby genes. In addition, promoter-specific hypomethylation or hypermethylation, altering the expression of tumor suppressor genes and oncogenes, gives tumor cells a growth advantage.

**EPIGENETICS AND RHEUMATIC DISEASES**

The study of epigenetic changes in rheumatic diseases is only at its beginnings but nevertheless has already provided interesting insight into the pathogenesis of these diseases. In particular, dysregulation of the immune response and lack of inflammatory resolution are thought to be associated with epigenetic changes in specific cell types.

**Systemic Lupus Erythematosus**

The importance of epigenetic changes in systemic lupus erythematosus (SLE) has been underlined by the finding that monozygotic twins discordant for SLE show significant differences in DNA methylation patterns. Accordingly, T cells from SLE patients were found to be globally hypomethylated as the result of reduced levels of DNMT1.
Overexpression of miR-21 and miR-148a, both of which directly and indirectly lead to repression of DNMT1, contributes to the loss of methyl marks in lupus T cells. Also, a defect in signaling through the ras-MAPK pathway found in T cells from patients with SLE results in low levels of DNMT1, thereby supporting global DNA hypomethylation. A direct connection between autoreactivity and loss of DNA methylation in T cells is supposed because DNA demethylation of normal T cells by 5-azacytidine promotes loss of tolerance, and in vitro hypomethylated T cells induce a lupus-like disease when transferred to mice. Furthermore, hypomethylated T cells increase macrophage cell death, which, in conjunction with reduced clearing of resulting cell debris, is proposed to lead to the induction of anti-DNA antibodies. This phenomenon could be based on increased expression of perforin through hypomethylation of its promoter in lupus T cells, which promotes killing of monocytes and macrophages. For the autoreactivity of lupus T cells in particular, increased levels of CD70 and CD11b could be responsible. Whereas CD70 is thought to be involved in B cell activation, CD11b is important in promoting cell-cell contact between T cells and other cells of the immune system. CD70 and CD11b promoter regions were found to be hypomethylated, as well as hyperacetylated in lupus T cells.

As mentioned earlier, DNA methylation is crucial for silencing of repetitive sequences in the human genome. In particular, aberrant expression of human endogenous retrovirus (HERV) sequences has long been implicated in the development of autoimmune diseases. HERV-E clone 4-1 transcripts were found to be expressed only in peripheral blood mononuclear cells (PBMCs) from SLE patients, but not in PBMCs from healthy controls. Because in vitro demethylation of DNA by 5-azacytidine induced HERV expression, it can be assumed that global DNA hypomethylation could be responsible for aberrant expression of this retrotransposon. Its presence might also lead to the production of autoantibodies, because expression of HERV-E clone 4-1 gag transcripts correlates with the presence of the anti-nuclear antibodies anti-U1 ribonucleoprotein (RNP) and anti-Sm.

Even though at certain promoter regions, histones in SLE have been found to be hyperacetylated (e.g., CD70, CD11b), global H3 and H4 acetylation measurements showed decreased histone acetylation in T cells of SLE patients. Also decreased levels in H4K9 methylation and changes in the H3K4me3 pattern are found in PBMCs of SLE patients and are connected to increased expression of disease-relevant genes.

**Rheumatoid Arthritis**

The imprinted activated phenotype seen in fibroblast-like synoviocytes (FLSs) in rheumatoid arthritis (RA), characterized by high basal expression of matrix-degrading enzymes, proinflammatory cytokines, and chemokines, is assumed to be maintained by epigenetic changes (Figure 22-4). Genomic DNA is globally hypomethylated in RA FLSs as the result of low levels of DNMT1, which lead to lack of restoration of methyl marks after cell division. A further indication for a hypomethylated state is expression of the LINE-1 retrotransposon in FLSs of RA patients. Reactivation of this retrotransposon in RA induces changes in gene expression (e.g., p38β, galectin-3–binding protein, met proto-oncogene), possibly contributing to the activated phenotype of FLSs in RA. Specific promoter regions also seem to be hypermethylated in RA FLSs, indicating a generally disturbed methylation pattern, as is also seen in cancer cells. The promoter of death domain receptor 3 (DR-3) is hypermethylated in RA FLSs; accordingly, RA FLSs have decreased levels of DR-3, which might contribute to their well-known resistance to apoptosis. Global and promoter-specific hypomethylation was also found in T cells of RA patients. In particular, analysis of the “senescent” CD4+CD28 subset of T cells, which are increased in patients with chronic inflammatory diseases such as RA, revealed that...
demethylation leads to the specific gene expression pattern seen in these cells. Because of replicative stress, these T cells downregulate ERK and JNK signaling pathways; this is associated with loss of methylation via downregulation of DNMT1 and DNMT3a. Also, DNA methylation of interleukin (IL)-6 and ephrinB1 promoters is decreased in RA lymphocytes, leading to increased expression of these genes in RA patients. In FLSs DNA demethylation with 5-azacytidine induced expression of miR-203. Basal levels of miR-203 are increased in RA FLSs and strongly correlate with basal expression of IL-6. Also, miR-155 and miR-146a are highly expressed in RA synovial tissues, FLSs synovial fluid, and PBMCs, whereas expression of miR-124a is decreased in RA FLSs. Because miRs are known to be commonly controlled by DNA methylation, it is feasible to assume that some of the observed changes in miR expression stem from the altered DNA methylation pattern in RA. However, experimental proof is still missing.

Also histone acetylation seems to be affected in RA. Increased and decreased activity and expression of HDACs have been described in RA synovial tissues. In particular, because TNF stimulation could increase levels of HDAC1 and HDAC2 in FLSs, levels of HDACs could be influenced significantly by anti-TNF treatment. However, it is not clear yet whether differences in lifestyle or medical treatment account for differences in HDAC activity in the analyzed patient groups. Nevertheless, in arthritis animal models, HDAC inhibitors show a strong anti-inflammatory effect.

An interesting cooperation between different histone modifications has been shown for the regulation of matrix metalloproteinase (MMP)-1 expression in RA. HERV-E clone 4-1 transcripts were found to be expressed only in peripheral blood mononuclear cells (PBMCs) from SLE patients, but not in PBMCs from healthy controls.

### Future Directions

The field of epigenetics is rapidly evolving, and immense progress has been made in the analysis of epigenetic mechanisms in health and disease. Some of the main tasks to be addressed in the future are decryption of the histone code and elucidation of mechanisms that guide chromatin modifications to specific genes. For the pathogenesis of any given disease, the question remains which epigenetic changes are the causes, and which are the consequences, of a pathologic state. Irrespective of the answer, the therapeutic control of specific epigenetic changes will more and more attract focus. Up to now, HDAC inhibitors and drugs inhibiting DNA methylation are already in use to treat patients with blood cancer. However, mechanisms of action of these drugs are not clearly understood. HDAC inhibitors, even though they increase acetylation of histones and accordingly activate gene transcription, also silence the expression of certain genes, probably as the result of modification of nonhistone targets. HDAC inhibitors and DNA methylation inhibitors are currently in clinical trials for the treatment of a variety of cancers, but could be effective in metabolic, inflammatory, and neurodegenerative diseases. The future will certainly bring new approaches for epigenetic targets such as the modification of specific histone methyl groups. Also, microRNAs have been targeted successfully in primates to treat hepatitis C. The use of so-called antagonists will certainly become a novel option for the treatment of rheumatic diseases in the future.

### References


The references for this chapter can also be found on www.expertconsult.com.
Complement System

JOHN P. ATKINSON

KEY POINTS

- The complement system is activated in the joints in rheumatoid arthritis and in vessel walls in vasculitis.
- Autoantibodies may activate the complement system, which then contributes to damage at the site of autoantigen expression.
- The complement system plays a key role in handling immune complexes, especially those that arise in the circulation. Complement activation is an efficient means to coat antigens so that they adhere to peripheral blood cells and are then phagocytosed or carried to the spleen and liver for disposal. In these organs, the immune complexes are transferred to tissue macrophages for destruction or an immune response.
- Complement measurements are helpful in diagnosing and following patients with systemic lupus erythematosus. Low C4 and C3 in the presence of anti–double-stranded antibodies have nearly a 100% specificity for lupus.
- In rheumatic diseases featuring complement activation, we lack sensitive day-to-day markers to monitor disease activity. Therefore the amount of complement inhibition required to modulate disease is not known. We also lack knowledge whether C3a/C5a, C4b/C3b, or the membrane attack complex is the key mediator of pathology, as well as which pathway and at which step to block.
- Urates and other types of cellular debris activate innate immunity including the complement system.

In this chapter we review the workings of the complement system, with a focus on how the system functions in rheumatic diseases. The involvement of the complement system has long been recognized in two of the most common inflammatory diseases treated by rheumatologists: systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA). What has been learned about the complement system’s role in these two diseases likely applies to related syndromes. The goal of this chapter is to help the rheumatologist appreciate the biologic and pathologic relevance of the complement system in rheumatic diseases. This information can also assist in the interpretation of laboratory tests of complement proteins and in the determination of the significance of complement fragments in pathologic specimens.

HISTORICAL ASPECTS

Today, complement is a collective term designating a group of plasma and membrane proteins that play a key role in innate and adaptive immunity. The complement system is ancient, predating chordates such as the lamprey and hagfish. Complement was discovered in the late nineteenth century by experimental pathologists attempting to understand the protective basis of vaccination and the cause of transfusion reactions. In these studies, it became apparent that the cell-free portion of blood contained a lytic substance for bacteria and for transfused red blood cells (RBCs). This “factor” was heat labile (destroyed by heating at 56°C for 30 minutes), unstable (it lost activity if serum was left on the bench top overnight), and innate or nonspecific (all individuals had this lytic substance). It was contrasted to antibody, which was heat stable, and only animals immunized to the specific bacteria contained this factor. Through mixing experiments, these two factors were found to be required for lysis—a specific recognition piece (antibody) and a nonspecific lytic piece (complement). In retrospect, this was the discovery of the classical complement pathway and led to its characterization as a heat-labile, bactericidal factor that “complemented” the acquired factor raised by immunization. Similarly, in the blood of patients who recovered from transfusion reactions, a heat-stable substance (antibody) was found that recognized the transfused erythrocytes. However, it did not lyse the transfused cells; rather, it cooperated with a second lytic substance (complement) present in fresh serum.

These historical points account for why lysis became linked to the complement system and why the study of antibody and complement dominated the field of immunology for the next half century. Of interest, the first autoimmune disease to be recognized was only a few years later. It was a hemolytic anemia (paroxysmal cold hemoglobinuria) in which the same pair of reactants, antibody and complement, were shown to mediate RBC lysis. Consequently, this double-edged sword aspect of the complement system was recognized early, but not until the 1960s was there such a clear-cut demonstration of its protective versus injurious role. Compared with normal mice, C5-deficient mice were 100-fold more sensitive to death from pneumococcal infection but 100-fold less sensitive to tissue destruction by autoantibodies. Even more to the point, we now know that inherited deficiencies of C1, C4, and C2 strongly predispose to SLE, yet immune complex (IC) deposition and complement activation contribute to tissue damage in SLE.

FUNCTION

A major activity of the complement system is to modify membranes and tag antigens through covalent attachment of its activation fragments (Figure 23-1). For example, several million C3b molecules can bind to a bacterial surface in less than 2 minutes. One goal of these deposited complement proteins is to opsonize the target. C3b (and its degradation fragments) and C4b are ligands for complement
receptors on peripheral blood cells, tissue macrophages, and antigen-presenting cells such as follicular dendritic cells. Complement ligands and receptors are particularly adept at this coupling process, known as immune adherence. On phagocytic cells, this often leads to ingestion of the complement-coated infectious particle, IC, or cellular debris. Additionally, microorganisms (e.g., gram-negative bacteria) and host cells may be lysed by the terminal complement components (C5b-C9). Many microbes, though, possess capsules that make them resistant to lysis (e.g., most gram-positive organisms). In human diseases featuring autoimmune processes, cells and tissues are similarly opsonized and membrane integrity can be compromised. A recent extension of this opsonic function relates to complement’s role in the clearance of cellular debris following necrosis or apoptosis. In particular, one hypothesis to explain autoantibody formation in SLE is the complement system’s failure to properly clear or handle apoptotic cells.5

A second function of the complement system is cellular activation.6 It prepares the local environment for defense against infection (see Figure 23-1). In the proteolytic steps of the early complement activation process, mediators that activate nearby cells are released. These low-molecular-weight fragments (C4a, C3a, C5a) are termed anaphylatoxins because, if released in excessive amounts, they induce shock. C3a and C5a, the two most powerful mediators, bind to their respective receptors at sites of complement activation. This causes histamine release by mast cells and chemotaxis to the area by phagocytic cells. With improved reagents and newer technology, receptors for C3a and C5a have now been shown to be much more widely expressed than initially thought; for example, they are expressed by epithelial cells, T cells, hepatocytes, endothelial cells, neurons, and other cell types. C3a and C5a may engage their receptors on endothelial and neuronal cells to alter blood flow and modulate cellular function.

NOMENCLATURE

There are three activation cascades: the classical pathway (CP), the alternative pathway (AP), and the lectin pathway (LP) (Figure 23-2). They converge to activate C3 and then C5, leading to opsonization and membrane perturbation.

The nine proteins of the CP are designated by the uppercase letter C, followed by a number (Table 23-1). These components act in numeric order, except for C4, which is cleaved in parallel with C2 and before C3. Components of the AP are designated by capital letters (e.g., factor B). Regulatory proteins are designated by a descriptive title (e.g., C4 binding protein) or a letter (e.g., factor H) (Table 23-2). Single components or multimeric complexes that have enzymatic activity are designated by a bar (e.g., C1s).

The loss of hemolytic potential by a component is usually designated by a lowercase prefix (e.g., iC3b). Fragments generated during complement activation are designated by a lowercase letter suffix (e.g., C3a, C3b). Except for the fragments of C2 (which are the opposite), the “a” fragment is liberated into the surrounding milieu, whereas the “b” fragment becomes cell bound and continues the cascade.

ACTIVATION CASCADES

Because many rheumatic diseases feature autoantibodies and therefore ICs, an understanding of the general principles of the activation cascades is essential to characterizing complement’s role in a disease such as SLE5–7,10 (Figure 23-3; see also Figure 23-2).

Classical Pathway

The CP is activated primarily by ICs and a few other substances including C-reactive protein.11 Human IgM and IgG subclasses 1 and 3 efficiently activate the CP. Once C1 is

Figure 23-2 Simplified diagram of the three pathways of complement activation. Deposition of clusters of C3b on a target is the primary goal. As shown by the broken line, the alternative pathway also serves as a feedback loop to amplify C3b deposition (regardless of which cascade deposited the C3b).
bound through the Fc portion of IgG or IgM interacting with the C1q subunit of the C1 complex (Figure 23-4), the C1r enzymatic subcomponent autoactivates by proteolysis and then cleaves C1s to C1s1. C1s cleaves C4 to C4b, releasing the C4a fragment (Figure 23-5). The C4b fragment, having had its thioester bond disrupted, now has the transient ability (a few microseconds) to covalently bind to a hydroxyl or amino group on a nearby target. C2 is also a substrate for C1, and C2 is cleaved by C1 to yield C2a and C2b. C2a interacts with 10% to 20% of the target-bound C4b to form the CP C3 convertase C4bC2a (Figure 23-6). The enzymatic or catalytic domain of this complex is C2a, which cleaves C3 to form C3a and activated C3b (Figure 23-7). Activated C3b, like C4b, has a few microseconds to

---

**Table 23-1 Components of the Complement Activation Cascades**

<table>
<thead>
<tr>
<th>Component</th>
<th>Serum Concentration (µg/mL)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classical Pathway (CP)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1q subcomponent</td>
<td>50</td>
<td>Binds IC to trigger CP</td>
</tr>
<tr>
<td>C1r subcomponent</td>
<td></td>
<td>Binds Fc portion of IgG, IgM</td>
</tr>
<tr>
<td>C1s subcomponent</td>
<td></td>
<td>Protease: cleaves C1s</td>
</tr>
<tr>
<td>C4</td>
<td>200-500</td>
<td>Protease: cleaves C4 and C2</td>
</tr>
<tr>
<td>C2</td>
<td>20</td>
<td>Opsonin* (C4b), C4a†</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Protease: cleaves C3 and C5 as part of the convertases</td>
</tr>
<tr>
<td><strong>Lectin Pathway (LP)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBL</td>
<td>150 (wide range)</td>
<td>Binds sugars to trigger LP</td>
</tr>
<tr>
<td>MASP-1/3</td>
<td>5</td>
<td>Proteases</td>
</tr>
<tr>
<td>MASP-2</td>
<td>5</td>
<td>Protease: cleaves C4 and C2</td>
</tr>
<tr>
<td><strong>Alternative Pathway (AP)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor D</td>
<td>1-2</td>
<td>Pro tease: cleaves factor B</td>
</tr>
<tr>
<td>Factor B</td>
<td>150-250</td>
<td>Pro tease: cleaves C3 and C5 as part of the convertase</td>
</tr>
<tr>
<td>Properdin</td>
<td>25</td>
<td>Stabilizes AP convertases</td>
</tr>
<tr>
<td>Central protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>550-1200</td>
<td>Opsonin* (C3b), C3a; convertases</td>
</tr>
<tr>
<td><strong>Terminal Pathway (TP)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>70</td>
<td>MAC component (C5b), C5a†</td>
</tr>
<tr>
<td>C6</td>
<td>60</td>
<td>MAC component</td>
</tr>
<tr>
<td>C7</td>
<td>60</td>
<td>MAC component</td>
</tr>
<tr>
<td>C8</td>
<td>60</td>
<td>MAC component (pore formation)</td>
</tr>
<tr>
<td>C9</td>
<td>60</td>
<td>MAC component (pore formation)</td>
</tr>
</tbody>
</table>

*C4b forms part of the CP C3 and C5 convertases. It anchors the enzyme complex to the target. Likewise, for the AP C3 and C5 convertases, C3b anchors these enzyme complexes to the target.

†C4a, C3a, and C5a are liberated on cleavage of C4, C3, and C5.

IC, immune complex; Ig, immunoglobulin; MAC, membrane attack complex; MASP, mannose-binding associated serine protease; MBL, mannose-binding lectin.

---

**Table 23-2 Complement Regulatory Proteins**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Tissue Distribution</th>
<th>Function</th>
<th>Disease Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 inhibitor</td>
<td>Plasma</td>
<td>Inactivates C1r, C1s, and MASP; a SERPIN</td>
<td>Hereditary angioedema (secondary to low C3), HUS*</td>
</tr>
<tr>
<td>Factor I</td>
<td>Plasma</td>
<td>Inactivates C1r, C1s, and MASP; a SERPIN; a cofactor protein</td>
<td>Infections (secondary to low C3), HUS*</td>
</tr>
<tr>
<td>Membrane cofactor protein</td>
<td>Most cells</td>
<td>Cofactor for cleavage of C4b and C3b</td>
<td>HUS*</td>
</tr>
<tr>
<td>Decayer-accelerating factor</td>
<td>Most cells</td>
<td>Decays C3 and C5 convertases</td>
<td>PNH</td>
</tr>
<tr>
<td>C4-binding protein</td>
<td>Plasma</td>
<td>Cofactor for cleavage of C4b and C3b</td>
<td>None†</td>
</tr>
<tr>
<td>Factor H</td>
<td>Plasma</td>
<td>Cofactor for cleavage of C3b; decays AP C3 and C5 convertases</td>
<td>Infections (secondary to low C3), HUS, type II MPGN, AMD</td>
</tr>
<tr>
<td>S protein (vitronectin)</td>
<td>Plasma</td>
<td>Blocks fluid-phase MAC</td>
<td>PNH</td>
</tr>
<tr>
<td>CD59</td>
<td>Most cells</td>
<td>Blocks MAC on host cells</td>
<td>None‡</td>
</tr>
<tr>
<td>Anaphylatoxin inactivator</td>
<td>Plasma</td>
<td>Inactivates C3a, C4a, and C5a</td>
<td>Urticaria, angioedema</td>
</tr>
</tbody>
</table>

*Most patients with HUS are heterozygous for function-altering mutations. ‡Insufficient number of cases of complete deficiency described to establish an association.

Complete deficiency has not been reported.

Receptors were named in order of their discovery but are also commonly identified relative to their ligand specificity and CD number (see Tables 23-3 to 23-5). For example, complement receptor type 1 may be referred to in the literature as CR1, the C3b/C4b receptor, the immune adherence receptor, or CD35.

AMD, age-related macular degeneration; AP, alternative pathway; CP, classical pathway; HUS, hemolytic uremic syndrome; MAC, membrane attack complex; MASP, mannose-associated serine protease; MPGN, membranoproliferative glomerulonephritis; PNH, paroxysmal nocturnal hemoglobinuria; SERPIN, serine protease inhibitor.
attach to a target by binding covalently to a hydroxyl group (analogous to the interaction of C4b with a hydroxyl or amino group), where it serves as the major opsonin of the complement system. In addition, some of the newly activated C3b binds covalently to a specific region on the C4b component of the C3 convertase to form the CP C5 convertase, C4bC2aC3b (see Figure 23-6).

CP activation by polyclonal antibodies is efficient on membranes of cells, bacteria, and viruses. In one illustrative study, complement activation by IgG binding to a membrane antigen of a nucleated cell produced about 2.5 million bound C4b molecules, and about 0.5 million C2a molecules subsequently attached to these C4bs to form convertases. In less than 5 minutes, 21 million (representing a 10-million-fold amplification) C4b molecules, and about 0.5 million C2a molecules were deposited. Further, only about 10% to 20% of the C4b and C3b molecules generated bound to the target. The rest were inactivated by hydrolysis in the fluid phase and then degraded by fluid-phase inhibitors. Concomitantly, C4a, C3a, and C5a anaphylatoxins were generated equal to the number of C4b, C3b, and C5b fragments produced. Approximately 1 million membrane attack complexes (MACs) were formed. Interaction of the complement system with soluble antigens is a less efficient process, particularly the formation of the C3 and C5 convertases.

A single IgM molecule can activate the CP. In contrast, two IgG molecules in close proximity (side by side) are required before C1 can be bound. On erythrocytes, this means that several thousand IgGs must bind to an antigen before activation can occur. Consequently, antigenic density and display play a role in determining whether complement is activated. For example, in warm antibody (IgG)-mediated autoimmune hemolytic anemia, complement is usually not fixed (despite the autoantibody’s being of an appropriate IgG subclass) because of low antigenic density.

**Lectin Pathway**

Lectins are carbohydrate-binding proteins (see Figure 23-4). They were initially described as proteins capable of agglutinating RBCs. They are important players in innate immunity and in rheumatic diseases. In particular, mannose-binding lectin (MBL) is a hepatocyte-synthesized plasma protein that preferentially binds to repeating mannosides and certain other repeating oligosaccharides of pathogens. MBL resembles C1q (it belongs to the same family of proteins, known as collectins), consisting of an oligomer with a terminal collagenous domain on one end and a globular domain on the other (see Figure 23-4). The main difference is that the carboxy terminus of MBL possesses a carbohydrate-recognition domain, whereas C1q has an Ig-binding domain.
Like C1q binding to the Fc portion of IgG, MBL engagement with its sugar ligand leads to activation of serine proteases similar to C1r and C1s, called mannose-associated serine proteases (MASPs). MASP-2 cleaves C4 and C2. MBL was discovered during an investigation into the cause of bacteremia in young children. MBL deficiency may be a predisposing factor for SLE and RA and contribute to a more severe disease phenotype.

The LP shares many features with the CP. The differences occur during the first few steps when MBL binds to an activating surface such as bacteria with the appropriate repeating oligosaccharides. Each element of the MBL-MASP complex is structurally and functionally homologous to the corresponding elements in the CP. C1q and MBL have globular ends through which they attach to their respective substrates. The MASP-2 in the complex cleaves C4 to form C4a and C4b (see Figure 23-5) and C2 to form C2a and C2b—identical to C1 cleaving C4 and C2. The C4b2a complex is the same as the C3 convertase generated in the CP (see Figure 23-6). The LP is a key player in innate immunity, where MBL and related lectins such as surfactants and ficolins bind to their cognate residues on microbes to trigger complement activation. At one time, MASPs (and possibly C1) directly activated C3, but through the evolution of C4 and C2, a more efficient C3 activating process was engineered.

Alternative Pathway

The AP was the second complement activation pathway discovered (hence the name), but it is phylogenetically the most ancient. In contrast to the CP and LP, the AP does not require antibody or a lectin to be triggered. It likely plays a larger role in rheumatic diseases than previously appreciated because of its so-called feedback or amplification loop. Thus if C3b is deposited on a target by the CP or LP, the AP can enhance C3b deposition many-fold. The AP is also continuously active at a low level (like an idling car). The plasma inhibitors factor H and factor I and host cell surface regulators keep the system in check. In the absence of these regulators, as exemplified by inherited deficiencies, the system fires to exhaustion. The AP routinely “kicks into gear” in the setting of microbes. If C3b deposits on an activating surface (i.e., one that lacks regulators), the system’s feedback loop allows several million C3b molecules to be deposited in a few minutes. In this pathway, factor B, structurally and functionally homologous to C2, binds to the deposited C3b, and is then cleaved by the active factor D to Bb and Ba. C3bBb is the AP C3 convertase (see Figure 23-6). It is stabilized by properdin (P), increasing the half-life of the enzyme 5- to 10-fold. The catalytic serine protease Bb of the C3bBbP enzyme complex then cleaves C3 to C3a and C3b; thus a feedback loop is set up. Some of the newly generated C3b molecules bind covalently to C3b already deposited on the target to form a C5 convertase (C3bBb). In vivo, the C3 and C5 AP convertases are short-lived owing to spontaneous decay, unless stabilized by properdin. Two new developments in the AP are that (1) MASP 1/3 appears to be required to activate pro-factor D to factor D11 and (2) properdin may bind to certain targets to then serve as a platform to initiate the AP.

Membrane Attack Complex

The terminal pathway begins with the cleavage of C5 by either the CP/LP or the AP C5 convertase. The liberated C5a fragment is a potent cell activator and chemotactic factor. The C5b binds C6, which in turn engages C7 to yield the fluid phase C5b67. This lipophilic complex can now interact with cell membranes. Following membrane insertion, it next binds C8. The C5b678 complex forms an initial pore in the plasma membrane and then recruits multiple (5 to 10) C9 molecules to yield C5b6789n. This complex is an efficient membrane channel former and is responsible for cell lysis. Many organisms and cell types are not easily lysed, however. The MAC also has multiple nonlytic effects on cells. The majority of these lead to activation of signaling pathways, as has been seen in the case of neurons and kidney cells. Some of the organ dysfunction observed in SLE, especially if transient or correctable with treatment, may be secondary to these nonlytic effects of the MAC.

REGULATORS

Physiologic Regulation

Unregulated, the complement system would fire to exhaustion, as illustrated by a complete deficiency of plasma regulatory proteins (factor H or factor I). The complement system has evolved to allow unimpeded activation on a microbe but to limit the process in time and space. The reaction must be finite in time to avoid excessive consumption of components in any one reaction and finite in space to minimize damage to surrounding host tissue. Thus a typical complement reaction on a microbial target occurs within a few minutes, and during this time, self-tissue is protected by plasma and membrane regulators. Many of these inhibitors act at the critical step of convertase formation. The convertase enzyme complexes intrinsically have short half-lives, and many of the inhibitors act to prevent their formation or to dissociate (decay) already formed complexes.
**Fluid-Phase and Membrane Inhibitors**

These regulators function at the initiating step (see Figure 23-3), the formation of the C3 and C5 convertases (see Figure 23-6), and the insertion of the MAC. The plasma inhibitors prevent fluid-phase activation (no target), whereas the membrane inhibitors prevent activation on healthy self-tissue (wrong target). The plasma inhibitors may also bind to damaged cells and extracellular matrices, where they now serve as a membrane regulator.\(^8,9,22,23\)

C1 inhibitor, a serine protease inhibitor (SERPIN), binds to C1r and C1s and to MASPs. The C1 complex is thereby disrupted, leaving C1q bound to antibody in the IC. C1q may then interact with C1q receptors to facilitate clearance of the IC.\(^24,25\) C1 inhibitor also prevents excessive and chronic fluid-phase C1 activation.\(^26\)

The C3 and C5 convertases are regulated by a family of proteins that include the membrane proteins decay-accelerating factor (DAF or CD55) and membrane cofactor protein (MCP or CD46) and the serum inhibitors C4-binding protein and factor H.\(^8-10,19-21\) These proteins act in three ways: prevent convertase assembly, disassemble convertases (decay-accelerating activity; Figure 23-8A), and serve as cofactors for the proteolytic inactivation of C4b and C3b (cofactor activity; Figure 23-8B). The cleavage is mediated by the plasma serine protease factor I. The MAC is also regulated by plasma and cell-anchored protein. CD59 is a widely expressed glycolipid-anchored membrane protein that binds C8 and C9 to prevent their membrane insertion, while the plasma protein vitronectin (S protein) inactivates fluid-phase MAC.

As a result of this regulatory activity, complement attack is focused on foreign surfaces (which usually lack complement regulators) and is held in check on host cells and in body fluids. Interestingly, a number of microorganisms have “captured” complement regulators (e.g., pox viruses) or have evolved proteins (herpesviruses) that inhibit complement activation (virulence factors).\(^27\) However, the activation of the CP by antibody is rapid and efficient, such that the inhibitors generally have modest effects on limiting damage by complement-fixing antibodies. The host inhibitors are particularly efficient at preventing fluid-phase activation, activation on normal self, or amplification via the feedback loop.

**COMPLEMENT RECEPTORS**

The complement system exerts many of its effects through receptors (Tables 23-3 to 23-5). During the activation process, target-bound and released fragments serve as ligands to activate the receptors, which can then trigger various cellular responses.

### Table 23-3  Receptors for C3 and C4

<table>
<thead>
<tr>
<th>Receptor</th>
<th>Primary Ligand</th>
<th>Location</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR1</td>
<td>C3b/C4b</td>
<td>Peripheral blood cells (except for most T cells and platelets), FDCs, B cells, podocytes</td>
<td>Immune adherence &gt; phagocytosis; antigen localization and retention; complement regulation</td>
</tr>
<tr>
<td>CR2</td>
<td>C3dg/C3d</td>
<td>B lymphocytes, FDCs</td>
<td>Co-receptor for signaling through B cell receptor; antigen localization and processing</td>
</tr>
<tr>
<td>CR3/CR4</td>
<td>iC3b</td>
<td>Myeloid lineage</td>
<td>Phagocytosis &gt; adherence</td>
</tr>
</tbody>
</table>

The CD numbers are as follows: CR1, CD35; CR2, CD21; CR3, CD11b/CD18; CR4, CD11c/CD18.  
FDC, follicular dendritic cell.
for complement receptors. For example, the vasomodulatory and chemotactic effects of C3a and C5a are due to interaction with their respective receptors. The opsonic fragments C4b and C3b mediate clearance of ICs and bacteria through adherence to and phagocytosis by CR1. Degradation of C3b by regulators leads to the formation of iC3b and then C3dg, which in turn interact with CR3/CR4 and CR2, respectively.

CR1 plays an important role in IC clearance. On erythrocytes, CR1 binds C3b/C4b-coated ICs (the immune adherence phenomenon) for processing and transport to the liver and spleen.28 In these organs, the ICs are transferred from the erythrocyte to tissue macrophages, allowing the erythrocyte to return to the circulation for another round of clearance. CR1 on granulocytes and monocytes binds and ingests ICs, whereas CR1 on B lymphocytes, tissue macrophages, and follicular-dendritic cells facilitates trapping and processing of IC in lymphoid organs. CR2 is expressed by B lymphocytes and follicular-dendritic cells, where it facilitates antigen trapping and is a coreceptor for activation of the B cell antigen receptor.29 Several new receptors for C3-bearing ICs have been described.30,31 The characterization of C1q receptors remains problematic, but they are likely important in the proper handling of antigens that have undergone complement activation on their surface.32

### COMPLEMENT IN THE INNATE AND ADAPTIVE IMMUNE RESPONSES

#### Innate Immunity

The complement system is the major humoral mediator of innate immunity. Toll receptors and their relatives comprise the major cellular arm of innate immunity. The complement system is activated by at least three mechanisms that are independent of an adaptive immune response (Table 23-6): natural antibodies, lectins, and the AP itself (Figure 23-9). Complement activation is therefore one of the earliest reactions to microbes at sites of infection. The complement response opsonizes organisms for adherence, phagocytosis, and antigen processing while releasing fragments that activate immunocompetent cells and trigger an inflammatory milieu.

#### Adaptive Immunity

As pointed out earlier, complement was discovered because of its role as an effector (lytic) arm of humoral immunity. IgM and IgG subclasses 1 and 3 efficiently activate the CP. More recently, an important role for complement on the afferent side has been “rediscovered.”

Accumulating evidence indicates that complement is an instructor of the adaptive immune response. Nearly 30 years ago, the injection of cobra venom factor was used to destroy an animal’s complement activity. In such an experimental model, a requirement for complement in an animal’s normal immune response was clearly demonstrated. An attenuated IgM response, a lack of class switching from IgM to IgG, and a failure to generate memory B cells were features of a complement-deficient animal. Similarly, in multiple subsequent studies, C4- or C2-deficient guinea pigs and humans were shown to have a defective response to the intravenous administration of a phage antigen. Although they produced IgM antibody, these two species did not class-switch from IgM to IgG, and a failure to generate memory B cells were features of a complement-deficient animal. Similarly, in multiple subsequent studies, C4- or C2-deficient guinea pigs and humans were shown to have a defective response to the intravenous administration of a phage antigen. Although they produced IgM antibody, these two species did not class-switch from IgM to IgG or demonstrate immunologic memory. In other studies, blocking CR2 reduced the antibody response to T cell–dependent antigens and prevented isotype switching. Further, animals with a targeted gene deletion of CR1/CR2 also had a lower IgM response and failed to isotype-switch. If larger doses of antigen or antigen plus adjuvants were administered, the complement-deficient animals responded normally. Finally, coating of antigen with C3d increases its immunogenicity up to 10,000-fold.29,31,34 Consequently, vaccines containing complement-coated antigens are likely to be more immunogenic. Taken together, these data indicate a contribution of the complement system (nature’s adjuvant) to an adaptive immune response.
Table 23-6 Complement System in Immunity

First line of defense as part of the innate immune response (takes place within seconds)
Mediates inflammatory responses
Modifies membranes of microbes
Instructive role to adaptive immunity
Facilitates antigen identification, processing, transportation, and retention
Activates cells to synthesize co-stimulatory molecules and to secrete cytokines and other mediators of an immune response; lowers threshold for B lymphocyte activation
Effector arm of humoral immunity ("complements" antibody)
Recognition of injured, apoptotic, and necrotic cells to enhance cleanup, proper disposal, and wound healing
Recognition of extracellular debris (crystals, pigments, lipids, proteins)∗

∗Examples are urate crystals, gout; lipofuscin pigments, drusen in age-related macular degeneration; oxidized lipids, atherosclerosis; and amyloid, Alzheimer's disease.

CLEARANCE OF NECROTIC AND APOPTOTIC CELLS

The complement system likely plays an important role in facilitating the removal of injured cells and cellular debris. Natural antibodies, lectins, and the AP recognize certain altered surface characteristics of damaged cells and, through opsonization, promote their proper removal. Such a system probably evolved to efficiently remove injured tissue, especially traumatized skin and apoptotic cells. In this process, a second goal is to avoid an immune response. If this system fails, as might happen in Clq or C4 deficiency, the individual is predisposed to developing autoantibodies (the so-called garbage- or waste-disposal hypothesis to explain autoimmunity in SLE). Ischemia-reperfusion injury is a related situation in which complement activation has clearly been shown to mediate damage to viable but at-risk tissues. Altered cellular debris engages the innate immune system. Examples are urate crystals in gout, amyloid proteins in Alzheimer's disease, oxidized lipids in atherosclerosis, and lipofuscin (pigments) in age-related macular degeneration. In these chronic processes, relatively minor changes in complement regulatory activity may accelerate tissue damage. This has been shown most clearly in age-related macular degeneration, where about 50% of the genetic risk appears to be due to a subtle loss of regulatory function in a polymorphic variant of factor H.

Figure 23-9 Activation of the complement system in innate immunity. Shown are three means of complement activation in a nonimmune host. Other means such as a serine protease (e.g., thrombin) directly cleaving C5 have been described. (From Huber-Lang M, Sarma JV, Zetoune FS, et al: Generation of C5a in the absence of C3: a new complement activation pathway, Nat Med 12:682–687, 2006; and Amara U, Rittirsch D, Flierl M, et al: Interaction between the coagulation and complement system, Adv Exp Med Biol 632:71–79, 2008.)

IMMUNE COMPLEX CLEARANCE

The complement system is important for the processing and clearance of ICs. As ICs form, activation of the CP leads to C4b and C3b deposition, which in turn prevents the ICs from precipitating in a vessel wall or at a tissue site (called maintenance of IC solubility). The deposition of C3b on the antibody and antigen reduces the antibody's ability to cross-link and thereby precipitate ICs. Even preformed ICs can be solubilized by exposure to fresh serum. Soluble ICs bearing clusters of C3b become bound to peripheral blood cells, especially erythrocytes (immune adherence). Erythrocytes possess more than 80% of the CR1 in blood. They serve as a “taxi” or “shuttle” to transport ICs to the liver and spleen, where they dissociate from the RBCs. Most such ICs are then destroyed by macrophages in the liver or spleen. This transfer of ICs from the red cells to tissue macrophages may be mediated by simple affinity differences because tissue macrophages possess multiple types of C3 and Fc receptors. Further, there is evidence for a proteolytic cleavage event at tissue sites that occurs near the stalk of CR1 to release ICs. The RBCs return to the circulation, possibly minus a few complement receptors, but ready for another round of immune adherence. This processing system for ICs evolved to prevent ICs from depositing in undesirable locations such as the kidney glomerulus. This clearance process could fail due to (1) a CP component deficiency, up to and including C3; (2) a complement receptor deficiency; (3) synthesis of a non-CP fixing antibody such as IgG4 or IgA; (4) severe hepatic dysfunction; or (5) splenectomy. This concept of IC clearance plays out in many infectious diseases and is especially pertinent to syndromes featuring ICs such as SLE, mixed cryoglobulinemia, serum sickness, and other vasculitic syndromes.

COMPLEMENT MEASUREMENT

Complement can be assessed by antigenic and functional assays (Tables 23-7 and 23-8). In clinical practice, the most common functional measurement is the total hemolytic or whole complement assay (THC or CH50). The assay is based on the ability of the patient’s serum sample to lyse sheep erythrocytes optimally sensitized with a rabbit antisheep red cell antibody. All nine components of the CP (i.e., C1 through C9) are required for a normal THC. A result of 200 units indicates that at a dilution of 1:200, the test serum lysed 50% of the antibody-coated sheep erythrocytes in the reaction mixture. THC is a useful screening tool for detecting a homozygous deficiency of a single component (C1 through C8) because total deficiency of a component produces a result of less than 10 units or an undetectable THC. Deficiency of C9 results in a low but detectable THC. THC can screen for total deficiency and provide an overall assessment of complement activation. It is particularly recommended in the initial evaluation of most SLE patients and in the evaluation of patients with recurrent infections with encapsulated organisms, especially if the neutrophil count and Ig levels are normal.

Commonly used tests, especially to follow a patient’s clinical course, are the antigenic assays for C4 and C3. They are simple, widely available, relatively inexpensive, and accurately carried out by nephelometric (turbidity)-based
**Table 23-7** Assays for Complement Activation in Human Disease

<table>
<thead>
<tr>
<th>Method</th>
<th>Use</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH&lt;sub&gt;50&lt;/sub&gt; or THC</td>
<td>Screen for component deficiency or activation of the classical pathway</td>
<td>Functional assay—requires appropriate sample handling</td>
</tr>
<tr>
<td>AP&lt;sub&gt;50&lt;/sub&gt;</td>
<td>Screen for component deficiency or activation of the alternative pathway</td>
<td>Functional assay—requires appropriate sample handling</td>
</tr>
<tr>
<td>Antigenic (ELISA, immunodiffusion, nephelometry)</td>
<td>Standard method for C3, C4, factor B, C1 inhibitor, MBL, and factor H</td>
<td>Widely available (particularly in case of C4 and C3), easy to perform, reliable, inexpensive samples usually sent to laboratories specializing in complement assays</td>
</tr>
<tr>
<td>Antigenic or hemolytic assay of an individual component</td>
<td>May be elevated in the setting of normal complement levels</td>
<td>Expensive; sample collection important; sent to commercial laboratories specializing in complement assays; more sensitive than assays of static levels</td>
</tr>
<tr>
<td>Activation fragments C3a, C5a, Bb, C1r/C1s, C5b–C9 (neoantigen)</td>
<td>Clinical picture consistent with HAE, but C1 inhibitor levels by antigenic assay are normal or elevated</td>
<td>15% of HAE kindreds have normal or elevated levels of a nonfunctional protein</td>
</tr>
<tr>
<td>C1 inhibitor function</td>
<td>Demonstration of complement fragments on cells and in tissues</td>
<td>C1q and cleavage fragments of C4 and C3 are most commonly analyzed, especially in kidney and skin biopsies; usual fragment detected is C3d</td>
</tr>
<tr>
<td>Immunofluorescence</td>
<td>Demonstration of C3 fragments on erythrocytes in hemolytic anemias</td>
<td>Inherited deficiency or in vitro activation*</td>
</tr>
<tr>
<td>Antiglobulin testing (non-γCoombs)</td>
<td>Demonstration of C3 fragments on erythrocytes in hemolytic anemias</td>
<td>Partial C4 deficiency or fluid-phase activation†</td>
</tr>
</tbody>
</table>

AP<sub>50</sub>, alternative pathway equivalent; CH<sub>50</sub> or THC, total hemolytic assay for classical pathway; ELISA, enzyme-linked immunosorbent assay; HAE, hereditary angioedema; MBL, mannose-binding lectin.

**Table 23-8** Interpretation of Results of Complement Determinations

<table>
<thead>
<tr>
<th>THC (units/mL)</th>
<th>C4 (mg/dL)</th>
<th>C3 (mg/dL)</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>150-250</td>
<td>16-40</td>
<td>100-180</td>
<td>Normal range</td>
</tr>
<tr>
<td>270</td>
<td>45 - 60</td>
<td>200</td>
<td>Acute-phase response (all are high); IL-1, IL-6, and TNF are major mediators to increase hepatic synthesis</td>
</tr>
<tr>
<td>100</td>
<td>10 - 30</td>
<td>80</td>
<td>CP activation (low C4 and C3); AP activation (normal C4 and low C3)</td>
</tr>
<tr>
<td>100</td>
<td>&lt;10 or 0</td>
<td>50</td>
<td>Inherited deficiency or in vitro activation*</td>
</tr>
<tr>
<td>50</td>
<td>&lt;8</td>
<td>100</td>
<td>Partial C4 deficiency or fluid-phase activation†</td>
</tr>
</tbody>
</table>

*In vitro activation is more common than an inherited deficiency state. The lack of activity (THC < 10) in the setting of normal C4 and C3 antigenic levels suggests (1) an improperly handled sample, (2) cold activation (e.g., by cryoglobulins) following collection of the sample, or (3) homozygous component deficiency (most commonly C2 with a lupus presentation or, with a *Neisseria* infection, an AP or membrane attack complex component).†Because THC is detectable, this cannot be a complete deficiency of C4. A partial C4 deficiency, such as of C4A, could give this result. Some types of immune complexes, especially cryoglobulins and a deficiency of the C1 inhibitor (hereditary angioedema), also give this pattern. In these cases, measurement of C2 is often helpful. A low value suggests activation, whereas a normal value suggests an inherited, partial C4 deficiency. Also, C4A and C4B alleles can be assessed by commercial laboratories.

AP, alternative pathway; CP, classical pathway; THC, total hemolytic complement or CH<sub>50</sub> whole complement titer. Modified from Stone JH, Wendum CM, editors: *Primer on the rheumatic diseases*, ed 12, Atlanta, 2001, Arthritis Foundation, p 71.
Table 23-9  Clinical Manifestations of Complement Component Deficiency

<table>
<thead>
<tr>
<th>Deficient Component</th>
<th>Clinical Syndrome*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classical Pathway</strong></td>
<td></td>
</tr>
<tr>
<td>C1q</td>
<td>SLE, infections</td>
</tr>
<tr>
<td>C1r/C1s</td>
<td>SLE, infections</td>
</tr>
<tr>
<td>C4</td>
<td>SLE, infections</td>
</tr>
<tr>
<td>C2</td>
<td>SLE, infections</td>
</tr>
<tr>
<td><strong>Lectin Pathway</strong></td>
<td></td>
</tr>
<tr>
<td>MBL</td>
<td>Infections</td>
</tr>
<tr>
<td>Central</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td></td>
</tr>
<tr>
<td>Membrane attack</td>
<td></td>
</tr>
<tr>
<td>C5, C6, C7, C8, C9</td>
<td>Neisseria infections</td>
</tr>
<tr>
<td><strong>Alternative Pathway</strong></td>
<td></td>
</tr>
<tr>
<td>Properdin, factor D</td>
<td>Neisseria infections</td>
</tr>
</tbody>
</table>

*With early component deficiencies of the classical pathway (C1, C4, or C2), infections are caused by the commonly encountered pyogenic organisms. With a late component (C5-C9) or an alternative pathway component deficiency, Neisseria infections predominate, especially meningococcal infections.

GN, glomerulonephritis; MBL, mannose-binding lectin; SLE, systemic lupus erythematosus.

Table 23-10  Autoimmune Disorders in Complete Complement Component Deficiencies

<table>
<thead>
<tr>
<th>Component</th>
<th>Number of Cases</th>
<th>Dominant Phenotype (%)</th>
<th>Other Phenotype*</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1q</td>
<td>35</td>
<td>SLE (94)</td>
<td>Infections</td>
</tr>
<tr>
<td>C1r/s</td>
<td>11</td>
<td>SLE (55)</td>
<td>Infections</td>
</tr>
<tr>
<td>C4</td>
<td>24</td>
<td>SLE (75)</td>
<td>Infections</td>
</tr>
<tr>
<td>C2</td>
<td>110</td>
<td>SLE (10-20)</td>
<td>Infections\notes{1}</td>
</tr>
<tr>
<td>C3</td>
<td>21</td>
<td>Severe infections (75)</td>
<td>GN, SLE, vasculitis</td>
</tr>
</tbody>
</table>

*Infections may be the presenting manifestation and dominate the clinical course, particularly in C1q, C4, or C2 deficiency.

1. Defective immune complex handling:
   - For foreign antigens
   - For self-antigens (apoptotic and necrotic debris)
2. Defective immune responses:
   - Reduced humoral immune responsiveness
   - Failure to regulate autoreactive B cells

Table 23-11  Classical Pathway Complement Component Deficiency Leading to Autoimmunity

Two leading hypotheses exist; they are not mutually exclusive.
1. Defective immune complex handling:
   - For foreign antigens
   - For self-antigens (apoptotic and necrotic debris)
2. Defective immune responses:
   - Reduced humoral immune responsiveness
   - Failure to regulate autoreactive B cells

More severe in C1q and C4 than in C2 deficiency. Less than 1% of lupus patients have a complete complement component deficiency, with C2 deficiency being the most common.

About one third of SLE patients develop Abs to C1q. These antibodies probably develop after the onset of SLE and represent a secondary phenomenon.\notes{35,46} These patients tend to have renal disease and low complement levels (C4 and C3). Their value in diagnosis and in following patients’ response to therapy are controversial.\notes{47}

The C4 genes are duplicated (all four must be deleted or defective to result in total C4 deficiency), giving rise to C4A (acidic) and C4B (basic) types.\notes{48} C4A deficiency occurs in 10% to 15% of white SLE patients (compared with 1% to 3% of controls). Because C4A preferentially binds amino groups, it reacts more efficiently with some types of ICs than does C4B. This may explain why homozygous C4A deficiency is an independent risk factor for SLE. The association generally occurs across multiple ethnic backgrounds, as well as in the European ancestral autoimmune haplotype (HLA-A1, C7, B8, C4AQ*0, C4B1, DR3, DQ2). The clinical course is similar to that of idiopathic lupus, but there may be less renal disease.\notes{49,51} In fact, one group has argued for caution in the use of immunosuppressive therapy because of the generally favorable course of renal disease in most patients.\notes{49} Assessment of the number of C4A and C4B genes is available through commercial laboratories.

The most common homozygous complement component deficiency is of C2, occurring in 1 in 10,000 to 20,000 white individuals.\notes{35} SLE occurs in about 15% of homozygous C2 null individuals. Notable differences in the setting of C2 deficiency are fewer renal and central nervous system manifestations but more cutaneous (prominent photosensitivity) involvement, earlier age of onset, and higher frequency of anti-Ro antibodies. Heterozygosity for C2 deficiency (1% to 2% of the white population) does not appear to predispose to SLE.

**Alternative Pathway**

C3 deficiency is associated with recurrent pyogenic infections and, less commonly, glomerulonephritis. For the latter, antinuclear antibodies are usually not detected, and other systemic features of SLE are lacking. Deficiency of properdin, factor B (one case), or factor D (a few cases) is strongly associated with meningococcal infections but not autoimmunity.\notes{52}

**Lectin Pathway**

Polymorphisms producing low serum MBL are associated with SLE across several ethnic backgrounds. In one series,
the incidence of SLE was increased twofold in patients with MBL deficiency. Interestingly, infectious complications, particularly severe pneumonias in association with immunosuppressive therapy, were more frequent in MBL-deficient patients. This combination of MBL and C4 deficiency may be a particularly strong predisposing factor for SLE. In RA, MBL deficiency may be a modest risk factor. Of perhaps greater interest, MBL deficiency was associated with earlier age of onset, more severe erosive disease, and increased frequency of infectious complications during treatment. As more studies have been reported, these MBL associations have become controversial.

**Acquired Complement Deficiency States**

**Lupus Paradigm**

Acquired complement deficiency states usually result from accelerated consumption. For example, CP activation is observed in more than 50% of lupus patients. Generally, the more active the disease process, the more likely that complement levels will be low because consumption progressively outstrips the liver’s synthetic capacity. Low complement values (C4, C3, or THC) in SLE are a poor prognostic factor and correlate with antibodies to native DNA, nephritis, and, overall, more severe disease. Complement deposition in tissue, especially in the glomerulus, assists in the diagnosis and classification of lupus nephritis. Antibodies to C1q occur in about 30% of SLE patients and are associated with low complement and renal disease.

Patients with mixed cryoglobulinemia and some with Sjögren’s syndrome, antiphospholipid syndrome, or vasculitis syndrome will follow the lupus paradigm of CP activation leading to low C4 and C3 that correlates with disease severity.

**KEY POINTS: SYSTEMIC LUPUS ERYTHEMATOSUS AND RELATED SYNDROMES**

- C1q, C4, or C2 deficiency is a single-gene defect that causes SLE in humans (rare variant with a major effect).
- Partial deficiency of C4 (namely, C4A) also predisposes to SLE.
- Once autoantibodies are present and ICs are deposited in undesirable tissue sites such as the kidney, complement activation contributes to the inflammatory reaction and tissue damage.
- Complement measurements are helpful in following lupus patients, particularly if C4 and C3 are low at the time of disease activity. If so, the levels usually increase in response to treatment and might become a useful marker of disease activity. However, this point must be established for each patient.
- The complement system is activated in most SLE patients, even though serum levels are not always reduced. Evidence for this statement comes from the assessment of complement turnover studies; the measurement of activation fragments; and, most convincingly, the demonstration of complement fragments in more than 70% of patients on skin biopsy (the old lupus band test). Even in normal skin, C3 fragments are deposited at the dermal-epidermal junction.

Anti-C1q antibodies are present in about 30% of SLE patients and in about 70% of lupus patients with nephritis. They probably arise in response to the high quantities of complement-activating IC. They likely contribute to the hypocomplementemia that is common in such patients, as well as to glomerular damage.

In mouse models of SLE, decreased activity of the early components of the CP accelerate disease development, whereas deficiencies of the late components (C5 to C9) ameliorate disease severity. The literature is not easy to interpret because the models may or may not accurately reflect human disease, and the mouse complement system has many differences from its human counterpart. For instance, mouse complement is a less potent activating system overall, and Fc receptors seem to be responsible for more of the proinflammatory reaction in mice than in humans.

**MODELS OF INFLAMMATORY DISEASE FEATURING COMPLEMENT ACTIVATION**

In vivo models have primarily explored complement activation as a mediator of tissue damage and inflammation in diseases in which complement is an effector arm of humoral immunity. More recently, complement’s role in promoting inflammation and cellular migration, enhancing the afferent limb of the immune response, and handling of debris has received much attention.

**Deficient Animals**

Before gene knockout mice were available, guinea pigs deficient in C4, C2, or C3; rats and rabbits deficient in C6; dogs deficient in C3; and pigs deficient in factor H had been characterized. Further, in the 1960s, many inbred mouse strains were shown to be deficient in C5. Experiments with these species established the double-edged sword aspect of the complement system. Thus the deficient animals were more sensitive to challenges with encapsulated bacteria but were more resistant to antibody-mediated tissue damage. They were also widely used to determine the role of complement of immunopathologic damage—namely the Arthus reaction, serum sickness, and Forssman shock. These models are each discussed as they demonstrate the range of reactions mediated by complement activation.

In the Arthus model, antibodies to a foreign antigen are raised and then the antigen is reinjected into the immune host. If the antigen is injected in a joint, ICs form in the joint space, complement is activated, Fcγ receptors are engaged, and an inflammatory reaction ensues. A transient and nondestructive acute arthritis develops, but if antigen is injected repeatedly, tissue damage can be induced.

In this and other models, emphasis has recently been placed on separating the effects of immune complexes working through Fcγ receptors (FcγRs) on phagocytes and B cells from those induced by activating complement. For example, it had generally been believed that the Arthus reaction in several species was mediated by complement activation. Thus the immigration of neutrophils in this...
process was thought to be caused primarily by the local formation of C5a (i.e., complement dependent). This concept was further investigated when it became possible to generate, by homologous recombination, mice with complement or FcγRII deficiency. In the mouse, FcγRII but not complement was required for the Arthus reaction in the skin. However, complement was critical in IC-mediated peritonitis in mice and IC-mediated lung disease in rats. Interestingly, complement is also of major importance in IC-mediated skin disease in guinea pigs and rats. To summarize, the species, tissue site of activation, genetic background, and undoubtedly additional factors are critical in determining the importance of complement versus FcγRI in disease pathogenesis. In most cases, these two effector systems operate conjointly, both in the response to pathogens and in autoimmunity.

In the usual serum sickness model in the rabbit employing bovine serum albumin (BSA)–anti-BSA, a polyarthritis develops that is mediated by the deposition of IC in joints. In the preantibiotic era, a serum sickness reaction was observed in patients receiving horse serum containing antibodies to treat infectious diseases. The human or rabbit clinical illness lasted 1 to 2 weeks, resolving as the host went into antibody excess and cleared the foreign protein. Of course, if more antigen was injected, an accelerated serum sickness reaction developed. These conditions may be met in SLE and in mixed cryoglobulinemia secondary to chronic hepatitis B or hepatitis C infection. In these conditions, nuclear antigens from cellular turnover and viral antigens synthesized in the liver are continuously present in a host with preformed antibodies.

Forssman antigen is a widely distributed lipopolysaccharide. Species that are Forssman negative as the rabbit develop an immune response to the RBCs of sheep, a Forssman-positive species. Thus, the rabbit responds to the injection of sheep RBCs with high-titer anti-Forssman antibody. On intravenous injection of this rabbit antibody into guinea pigs, a Forssman-positive species, the antibody travels to the lung, where it causes alveolar capillary fluid leakage and hemorrhage that is CP dependent. Phagocytes are not important in this catastrophic reaction that causes death in a few minutes secondary to pulmonary edema. The MAC of complement is essential.

Gene-Targeted Deficiencies

Most components, regulators, and receptors have now been deleted by gene-targeted insertional mutagenesis. C1q, C4, C2, C3, and CR1/CR2 mice demonstrate substantial defects in T cell–dependent immune responses and have the expected defects in IC clearance. Thus for most strains with a defect in the CP or AP, defective clearance of bacteria and viruses was shown. The C1q- and C4-deficient mice (depending on the genetic background) are also predisposed to autoimmunity including glomerulonephritis (lupus-like picture) and delays in the clearance of apoptotic bodies. The C5aR-deficient mice have a decrease in inflammatory cell infiltrates. Although there are substantial differences between the human and mouse complement systems, these mice have been useful in clarifying complement’s role in host defense against infection, the immune response, and autoimmunity.

Antiphospholipid Antibody Syndrome

The antiphospholipid syndrome is often associated with hypocomplementemia in the absence or presence of a positive test for antinuclear antibodies. In a recent review of a large number of such patients, hypocomplementemia was just as frequent as in patients with SLE. Complement activation is also a common event in recurrent fetal loss. In a mouse model of antiphospholipid antibody–induced pregnancy loss, complement-deficient mice were protected from fetal loss. Complement activation by the antiphospholipid antibody is a required intermediary event in the pathogenesis of fetal injury. In particular, complement activation by the autoantibodies is proposed to cause a deregulation of angiogenic factors such as vascular endothelial growth factor. Clinical trials are under way to assess the role of complement activation in patients with the antiphospholipid syndrome.

Complement Activation in Polyarthritis

Human Rheumatoid Arthritis

A long-standing observation in RA and juvenile RA is that the complement system is activated in the synovial fluid and surrounding joint tissue. Thus if the functional activity of joint fluid C4, C2, or C3 is determined, it is lower than expected on the basis of the antigenic level. Such data have established that complement components in RA joint fluid are commonly functionally inactive. In other words, the components have already been engaged in an immune reaction, and activation fragments are largely being measured. Further, essentially every complement activation fragment is elevated in RA joint fluid. The activation profile is primarily that of the CP (low C4 and C2), although there is also substantial evidence for contribution of the AP. Clinical correlations with measures of complement activation in RA include a more severe, erosive, and seropositive disease process. For approximately 3 decades (1950 to 1980), B cells, rheumatoid factors, and complement were postulated to play a predominant role in mediating the synovial inflammation in RA. Then, for the next 2 decades, T cell–mediated immunity was thought to be the predominant system responsible for the synovial reaction in RA. Today, a more prominent role for B cells, autoantibodies, and complement is again thought to exist; their critical role in animal models of RA has been recognized, and antibodies to citrullinated peptides have been identified. Both humoral and cellular immunity are required to produce a persistent synovitis in the RA syndrome.

KEY POINTS: RHEUMATOID ARTHRITIS AND RELATED SYNOVITIDES

Complement activation fragments are present in joint fluid and deposited in synovial tissue.

Autoantibodies (e.g., to citrullinated peptides) can activate the complement system and trigger inflammation.

Rheumatoid factors augment the effects of IgG bound to an antigen by promoting both agglutination of the immune complexes and further complement activation.
Low serum C4 and C3 are common in rheumatoid vasculitis, indicating classical pathway activation by immune complexes; similarly, cryoglobulins and other types of immune complexes activate the classical pathway.

In mouse models of rheumatoid arthritis featuring autoantibodies, complement inhibitors and complement deficiency protect against disease development and/or reduce its severity.

Collagen-Induced Arthritis

Complement activation products are prominent in the joint fluid and synovial tissue of rat and mouse models of collagen-induced arthritis. The disease is mild or absent in mice who are C5-deficient or treated with a function blocking monoclonal antibody to C5. Several complement inhibitors with activity against the convertases and factor B–deficient mice were also protected. These results implicate the AP in generating C5a, C5b to C9 (MAC), or a combination thereof as being responsible for synovitis. One might have predicted that C3 activation products would be major contributors to disease pathogenesis, but that was not the case in these models.

K/BxN Mouse Model

A spontaneous mouse model was generated by crossing the T cell receptor (TCR) transgenic mouse line (known as KRNxC56BL/6) with the nonobese diabetic (NOD) mouse strain. Autoantibodies that arise in the K/BxN mouse recognize the ubiquitously expressed enzyme glucose-6-phosphate isomerase. Important features of this model are (1) a destructive small-joint arthritis that resembles RA develops; (2) the autoantibodies are pathogenic, as the disease can be transferred to other strains by antibody alone; and (3) complement is required for the arthritis to develop. Surprisingly, the CP is not necessary; mice deficient in C1q or C4 are as susceptible as wild-type mice. Mice deficient in C3 or AP component factor B, however, do not develop arthritis. Moreover, C5-deficient mice or mice treated with a monoclonal antibody to C5 are resistant to the disease. C5a receptor–deficient mice are also protected from developing arthritis. These results implicate AP-derived C5a interacting with its receptor as being critically important to disease development.

Another insight from this model is a putative role for the complement regulatory proteins in “allowing” a local inflammatory process. Thus the pathogenic antibody in this model also binds to the kidney glomerulus, but there is no excessive complement deposition and no development of glomerulonephritis. A likely explanation is that both antigen accessibility and a permissive local environment for complement activation are required. Joint cartilage is relatively devoid of complement regulatory proteins. In this model, the AP is engaged but not the CP because the autoantibody is predominantly of the IgG1 subclass. In the mouse, this antibody subclass has poor complement-fixing ability through C1. Nevertheless, antibodies bound to an antigen can protect the cascade from regulation by inhibitors and thereby trigger the AP. Though it remains unclear why the autoantibody arises, this model has provided provocative insights into how an autoantibody- and complement-mediated inflammatory arthritis could occur in humans. Further, it has many parallels to human RA featuring anticitrullinated protein antibodies.

THERAPEUTIC IMPLICATIONS

The study of the complement system has undergone a renaissance. Much of this revival relates to (1) the development of inhibitors for experimental and clinical use; (2) the association of deficiency states with autoimmunity (especially the C1q, C4, C2 lupus association); (3) greater appreciation of the role of the innate immune system and its role in instructing the adaptive immune system; (4) the availability of gene-targeted mice to more precisely assess the role of complement in the normal immune response and in mouse models of autoimmunity; and (5) whole genome screens with single-nucleotide polymorphism analyses that have identified partial deficiencies and polymorphisms in complement regulatory proteins as causing a predisposition to atypical hemolytic uremic syndrome and age-related macular degeneration.

The development of complement inhibitors for clinical use is a long-sought goal. One such compound, a humanized monoclonal antibody to C5, has been used successfully to treat a complement-mediated hemolytic anemia and atypical hemolytic uremic syndrome (aHUS). This monoclonal antibody (eculizumab) is now approved by the FDA to treat paroxysmal nocturnal hemoglobinuria (PNH) and aHUS. Both of these diseases feature illnesses in which there is a decrease in complement regulatory proteins. In PNH the hosts’ RBCs are lysed by the alternative pathway, while in aHUS excessive alternative pathway activation on the glomerular endothelium leads to a hypercoagulative state. Now that it is in the marketplace, it will be tried in most diseases featuring autoantibodies that activate complement. This agent and many others in the pipeline ameliorate inflammatory responses in animal models, where considerable evidence has already been accumulated relative to complement’s participation. Further, a role for complement in clearing apoptotic cells is likely. Characterizing the effects of complement inhibitors in clinical trials of several rheumatic diseases in humans is much anticipated.

References


The references for this chapter can also be found on [www.expertconsult.com](http://www.expertconsult.com).
The addition of oxygen to arachidonic acid and other polyunsaturated fatty acids not bound to membrane phospholipids in nearly all human cell types results in formation of several classes of bioactive products termed eicosanoids. These include prostaglandins (PGs), prostacyclin (PGI), thromboxanes (TXs), leukotrienes (LTs), and lipoxins (LXs). All of these compounds are crucial to the regulation of immunity and inflammation, among other physiologic and pathologic processes. Although eicosanoids are derived from C20 polyunsaturated fatty acids (eicosa = 20), only a small percentage of these polyenoic acids form the eicosanoids: dihomogamma linolenic acid (DGLA), which is 8,11,14-eicosatrienoic acid; arachidonic acid (AA), which is 5,8,11,14-eicosatetraenoic acid; and eicosapentaenoic acid (EPA), which is 5,8,11,14,17-EPA (Figure 24-1).

Two groups of fatty acids are essential to the body: the omega-6 series derived from linoleic acid (18:2 n-6) and the omega-3 series derived from α-linolenic acid (18:3 n-3). The n refers to the number of carbon atoms from the methyl (omega) end of the fatty acid chain to the first double bond (i.e., omega-3 and omega-6 designations). Using this notation, 18 refers to the number of carbon atoms in the fatty acid. The degree of unsaturation (the number of double carbon-carbon bonds) follows the number of carbon atoms. Fatty acids are metabolized by an alternating sequence of desaturation (i.e., removal of two hydrogens) and elongation (i.e., addition of two carbons). Membrane phospholipids are the main storage site for polyunsaturated fatty acids and are particularly rich in eicosanoid precursors, which are located at the sn-2 position (Figure 24-2). Because mammalian cells cannot interconvert n-3 and n-6 fatty acids, the composition of membrane phospholipids is determined by exogenous sources of fatty acids.

**BIOSYNTHESIS OF EICOSANOIDs**

**Phospholipases**

Phospholipase A₂ (PLA₂) in lysosomes or bound to cell membranes catalyzes the breaking of the sn-2 bond, facilitating release of AA or other polyunsaturated fatty acids (see Figure 24-2). The enzyme is crucial to regulation of eicosanoid synthesis because it is in the nonesterified state that the polyunsaturated precursors enter into the cascades, leading to eicosanoid formation. Only a scant amount of oxidation at carbon 15 of AA occurs catalytically when the fatty acid is still covalently bound as part of a phospholipid. More than 25 different PLA₂ isoforms have been characterized and grouped based on primary structure, localization, and Ca²⁺ requirements.

Cytosolic PLA₂ group IV is the major catalyst of arachidonate release that leads to PG and LT production. The structure of cPLA₂ reveals a monomeric cytosolic protein with a molecular size of 85 kD. Potent and selective inhibitors of cPLA₂ have been developed. Lyso-phospholipids “left over” after the action of PLA₂ are direct precursors of platelet-activating factor (PAF), a potent mediator of inflammation, which is generated by acylation (addition of fatty acid) in the open sn-2 position of the lysophospholipid. One member of the PLA₂ superfamily, lipoprotein PLA₂, may serve as a marker for cardiovascular risk.

Four distinct types of PLA₂ activities hydrolyze fatty acids esterified at the sn-2 position: Secretory PLA₂ (sPLA₂) has small disulfide cross-linked proteins that require Ca²⁺ in mM concentrations for optimal activity. Cytoplasmic PLA₂ (cPLA₂) has larger proteins requiring Ca²⁺ in μM concentrations, which are AA-selective and can deacylate diacylphospholipids completely, preventing accumulation of potentially toxic lysophospholipids. Ca²⁺-independent PLA₂ (iPLA₂) exhibits specificity for plasmalogen substrates. Finally, PAF acetylhydrolase (PAF-PLA₂) has a series of isoforms specific for short chains.

Under basal conditions, AA is liberated by iPLA₂ and then reincorporated into cell membranes (reacylated). It is unavailable for appreciable eicosanoid biosynthesis. The acylase enzymes competitively inhibit the cyclooxygenase (COX) isoenzymes. After receptor activation and cell stimulation, intracellular Ca²⁺ levels increase and the Ca²⁺-dependent cPLA₂ liberates arachidonate at a rate that exceeds the rate of reacylation, leading to arachidonate metabolism by COX isoenzymes and by lipoxygenases (LO). Initial formation of PGE₂ seems to be due to preferential coupling among cPLA₂, COX-1, and cytosolic PGH-PGE isomerase. More intense inflammation leads to participation of secreted sPLA₂ and amplification of PGE₂ biosynthesis by inducible COX-2. It is an oversimplification to regard
Peripheral blood monocytes from patients with rheumatoid arthritis (RA) exhibit greater PLA₂ and PLC activity than cells from healthy volunteers. PLA₂ concentrations did not correlate with disease activity, but the greatest increases in PLC enzyme activity were seen in cells from patients with the most severe, persistent, proliferative disease, not in cells from patients with the most active disease at the time the cells were studied.

In accord with the need for balance in regulation of the inflammatory response, sPLA₂, numbering nine members in humans, has both inflammatory and anti-inflammatory actions. Thus, group V sPLA₂ counters the inflammatory activity of group IIA sPLA₂ by regulating cysteinyl leukotriene synthesis and promoting immune complex clearance in a murine model of autoimmune erosive inflammatory arthritis. It is of interest that concentrations of IIA sPLA₂ are far greater than those of V sPLA₂ in synovial fluid from patients with RA. In keeping with the fact that nothing in biology and the pathogenesis of disease is simple is the observation that treatment of RA patients with sPLA(2)-IIA inhibitors has only transient benefit and that sPLA(2)-IIA has
activity (induction of COX-2) independent of its enzyme activity. Hydrolysis of phospholipids by phospholipase D (PLD) produces phosphatidic acid (PA) and the respective polar head groups. The capacity of cells to interconvert PA and DAG through the action of specific cellular phosphatases and kinases (Figure 24-3) suggests that AA release from DAG and a variety of intracellular signaling and protein-trafficking events may be regulated by PLD activity. The targets of phosphorylation within the PLD molecule that are key to its regulation have been mapped. PLD may be activated after or independent of PLC activation.

The tetraenoic precursor (AA) is the most abundant of the three precursor fatty acids in cells of individuals who eat usual Western-style diets. Metabolites of AA constitute the “2” series (dienoic) PGs (two double bonds in the molecule), and the metabolic pathway has acquired the familiar name “arachidonic acid AA cascade.” However, diets enriched in eicosapentaenoic or gamma-linolenic acid, the other eicosanoid precursors, lead to formation of different eicosanoids. Figure 24-4 illustrates the COX and 5-LO lipoxigenase pathways of the cascade.

Cyclooxygenase Pathway

The first step in the biosynthesis of the “prostanoids” (e.g., PGs, TXs, prostacyclin) is catalyzed by the bifunctional PG endoperoxide synthase isozyme (PGHS)-1 (COX-1) and PGHS-2 (COX-2). COX is a homodimeric enzyme that integrates into only a single leaflet of the lipid bilayer of the cell membrane. The COX active site is a long hydrophobic channel. Aspirin and most other nonsteroidal anti-inflammatory drugs (NSAIDs) exclude arachidonate from the upper portion of the channel. To form the characteristic five-carbon ring structure (TXs contain a six-member ring), the precursor fatty acids must have double bonds at carbons 8, 11, and 14 (numbering from the carboxyl group). When a molecule of oxygen is inserted across carbons 9 and 11, ring closure occurs enzymatically across C8 and C12, creating the unstable PG endoperoxide PGG. Subsequent peroxidation yields PGH with formation of the cyclopentane ring. PGH serves as the common precursor for PGs, prostacyclin, and TXs that are formed under the influence of terminal synthases (see Figure 24-4). In addition to the activity of phospholipases, regulation of PG synthesis also occurs at the level of PGHS gene expression. PGHS levels are increased by interleukin (IL)-1, platelet-derived growth factor, and epidermal growth factor, agents that increase PG synthesis.

Cell membranes constitute the source of substrate AA and the site of action of eicosanoid-forming enzymes. PG synthesis can also form at lipid bodies, which are non-membrane-bound, lipid-rich cytoplasmic inclusions that develop in cells associated with inflammation. Lipid bodies isolated from human monocytes express PGHS activity, are reservoirs of arachidonyl phospholipids, and can function as domains of PG synthesis during an inflammatory reaction.

PGHS, the well-known target of NSAIDs, exists in two isoforms. These isoforms are similar in terms of amino acid identities (about 60%), catalytic properties, and substrate specificity, but they differ in their genomic regulation.

Regulation of Cyclooxygenase-1 Expression

The gene for COX-1 (PTGS1) is preferentially expressed constitutively at high levels in selected cells including endothelium, monocytes, platelets, renal collecting tubules, and seminal vesicles. Because the expression level of the enzyme does not vary greatly, it has been difficult to study its transcriptional regulation. The gene has a TATA-less promoter that contains multiple start sites for transcription. It is also known that Sp1/Cis regulatory elements in the promoter bond the Sp1 transcription factor to induce COX-1 gene expression. In addition, COX-1 (COX-2) splice variants may function in tissue-specific normal and pathologic processes and may represent new targets for therapy.
under basal conditions suggests that its major function is to provide eicosanoids for physiologic regulation. This is seen clearly in platelets that do not have nuclei and cannot produce an inducible enzyme on activation. Rather, TXs are produced constitutively so that platelet aggregation can be completed.

**Regulation of Cyclooxygenase-2 Expression**

The regulated formation of eicosanoids implies that cells have an appreciable ability to amplify the rate and amount of eicosanoid synthesis. Several processes contribute to that regulation including silencing of sPLA2 expression by COX-2 and autoinactivation (“suicide inactivation”) of COX-2 and other oxygenases and synthases. In addition, the COX-2 transcript contains at least 12 copies of the AUUUA RNA motif, which makes it unstable and subject to rapid degradation. Factors that regulate COX-2 expression are specific for the physiologic processes involved. Expression of COX-2 in the macula densa in the kidney depends on luminal salt concentrations. Transcriptional activation of the COX-2 gene by mediators of inflammation such as IL-1β and tumor necrosis factor (TNF) is likely regulated by the transcription factors nuclear factor κB (NFκB) and C/EBP. Perhaps the most crucial of the several demonstrated regulatory sequences in the 5’ flanking regions of the COX-2 gene is the ATF/CRE, a site that is activated by the transcriptional activator protein-1 (AP-1) and the cyclic adenosine monophosphate (cAMP) regulatory binding protein. Now that quantitative polymerase chain reaction analysis is routinely used to examine gene expression, it is clear that COX-2 is constitutively expressed under basal conditions at levels too low to be detected by Northern blot analysis. Depending on cell and tissue specificity, several signaling pathways (kinases, Rho, cyclic guanosine monophosphate, Wnt) and transcription factors (NFκB, AP-1, nuclear factor of activated T cells [NFAT]) are involved in COX-2 expression.

COX-1 and COX-2 effect a balance in several physiologic and pathologic situations. Of particular interest are their actions in kidney and stomach. During times of low blood volume, the kidney releases angiotensin and other factors to maintain blood pressure by systemic vasoconstriction. Angiotensin also provokes PG synthesis in the kidney. COX-1, expressed in vessels, glomeruli, and collecting ducts, produces vasodilating PGs, which maintain renal
plasma flow and glomerular filtration during conditions of systemic vasoconstriction. In the antrum of the stomach, COX-1 leads to production of PGs, which increase gastric blood flow and mucus secretion. Inhibition of COX-1 by NSAIDs prevents these protective mechanisms and results in renal ischemia and damage and gastric ulcers (mainly antral) in susceptible individuals. These observations have led to development of NSAIDs that selectively inhibit COX-2 and spare COX-1. AA gains access to the active site of the COX via a hydrophobic channel, and access is blocked by insertion of an acetyl residue on Serine 530 in COX-1 and Serine 516 in COX-2. The irreversibility of the interaction and the unique expression of COX-1 in the anucleate platelet is the reason for the clinical efficacy of low-dose aspirin. Nonacetylated NSAIDs compete with arachidonate for the active site and can interfere with the sustained effects of aspirin. Although the structures of COX isozymes are similar, COX-2 is characterized by a side-pocket extension to the hydrophobic channel, which is where the selective COX-2 inhibitors localize.

The major adverse effects of NSAIDs, gastroduodenal injury and impaired renal function, as well as the induction of myocardial infarction and stroke, are caused by inhibition of COX-1, whereas the analgesic and anti-inflammatory activities of NSAIDs rest in part on their ability to inhibit COX-2. COX-2 also seems to have a regulating role, however, in renal, brain, gastrointestinal, ovarian, and bone function. COX-2 is also expressed in endothelial cells, and its inhibition suppresses prostacyclin synthesis by endothelial cells. COX-2 acts in the initiation and the resolution of inflammation. Its expression increases transiently early in the course of carrageenan-induced pleurisy in rats. Later in the response, COX-2 is expressed at even higher levels, leading to synthesis of PGD$_2$ and its dehydration product 15-deoxy-$\delta$12,14-PGJ$_2$ (15$\delta$PGJ$_2$). Early expression of COX-2 is associated with production of inflammatory PGs, whereas the later peak results in production of PGs that suppress inflammation. That inflammation occurs in COX-2 knockout mice reminds us, as Lewis Thomas stated, “inflammation will take place at any cost.” Nonetheless, the search for more selective and localized COX-2 inhibition goes on. One strategy is achieved by using the mechanism of RNA interference (RNAi). Nonpathogenic Escherichia coli are engineered to invade tumor cells (and perhaps could be designed to invade synovial cells) and to generate anti-COX-2 siRNA molecules (siCOX-2), thereby silencing overexpressed COX-2. The involvement of miRNAs in COX-2 posttranscriptional regulation suggests a possible endogenous silencing mechanism to reduce COX-2 expression.

**Cyclooxygenase-3**

Acetaminophen, similar to NSAIDs, suppresses pain and fever. It is not an anti-inflammatory agent, and its mechanism of action—despite its extensive use—has not been apparent. The finding that acetaminophen inhibits COX activity more in canine brain homogenates than in spleen homogenates gave rise to the concept that variants of the COX enzyme exist that are differentially sensitive to acetaminophen. COX-3 is a splice-variant of COX-1 that retains the intron-1 gene sequence at the mRNA level, which encodes a 30 amino acid sequence inserted into the N-terminal hydrophobic signal peptide of the enzyme protein. COX-3 protein and mRNA transcripts have been identified in human tissues as well as rats, most abundantly in cerebral cortex and heart, although the mechanism for conversion of COX-3 mRNA to active enzyme is not clear. These findings led to a search for other COX variants. A COX-2 variant induced by 0.5 mM diclofenac via stimulation of the nuclear receptor peroxisome proliferator-activated receptor (PPAR$\gamma$) (the receptor for the pro-resolving 15-deoxy-$\delta$12, 14 PGJ$_2$) has been discovered. Whereas induction of COX-2 by LPS results in release of the inflammatory cytokines IL-6 and TNF, diclofenac-induced COX 2 leads to release of the anti-inflammatory cytokines TGF beta and IL-10.

**Prostaglandin Synthases**

Conversion of the endoperoxide intermediate PGH$_2$ to PGs requires the activity of specific terminal synthases: Hemeoxygenase (H-PGDS) catalyzes the isomerization of PGH$_2$ to PGD$_2$, while cyclooxygenase (COX) synthase converts PGF$_2\alpha$ to PGF$_2\alpha$. Cytosolic COX synthase (cPGES) is responsible for constitutive expression of PGE$_2$, and micromental COX synthase (mPGES-1) induces PGE in response to inflammatory stimuli. At least 10 enzymes convert PG precursors into biologically active PGs. Suppression of PG synthase activity might be considered as an alternative strategy that would fall between global blockade by inhibition of COX and blockade of a single eicosanoid receptor. Increased fibrosis may now be added to the well-known adverse effects of NSAIDs including COX-2 selective agents. Thus efforts have been directed at development of drugs that inhibit mPGES-1, rather than COX-2, thereby suppressing PGE$_2$ production while sparing production of prostacyclin. Microsomal PGES-1 is upregulated throughout development of the rat adjuvant-induced arthritis model. Because PGE$_2$ is also important to induction of other eicosanoid synthases, suppression of mPGES-1 might counteract articular cartilage degradation in patients with inflammatory arthritis.

**PRODUCTS OF THE CYCLOOXYGENASE PATHWAY**

**Prostaglandins**

The basic structure of all PGs is a “prostanoid acid” skeleton, a 20-carbon fatty acid with a five-membered ring at C8 through C12 (see Figure 24-4, inset). The term prostaglandins is employed widely but should be used only to describe the oxygenation products that contain the five-membered carbon ring. A family of acidic lipids found first in human seminal fluid, PGs were misnamed because it was thought they were produced in the prostate gland rather than in the seminal vesicles. The alphabetic PG nomenclature (e.g., PGE, PGE$_2$, PGD) is related to the chemical architecture of the cyclopentane ring: PGE and PGF differ only in the presence of a ketone or hydroxyl function at C9 (see Figure 24-4). These compounds are made by a variety of cells (e.g., PGE$_2$ and PGD$_2$ by isomerases, PGE$_{20}$ by a reductase). In
the nomenclature a subscript numeral after the letters indicates the degree of unsaturation in the alkyl and carboxylic acid side chains. The numeral 1 indicates the presence of a double bond at C13-C14 (PGE1), 2 marks the presence of an additional double bond at C5-C6 (PGE2), and 3 denotes a third double bond at C17-C18 (PGE3).

PGs are produced on demand and seem to exert their effects on the cell of origin or nearby structures. They are not stored in cells and are degraded rapidly in vivo by 15-hydroxyprostaglandin dehydrogenase (PGDH) during one passage through the lungs. Abundant experimental evidence supports the view that PGs participate in development of the inflammatory response. Administration to rats of a monoclonal antibody to PGE2 prevents carrageenan-induced pain and inflammation.31 PGs are probably better at potentiating the effects of other mediators of inflammation than they are at inducing inflammation directly. PGE compounds and intermediate hydroperoxides of AA increase pain sensitivity to bradykinin and histamine. The effects of PGE are cumulative, depending on concentration and time. Even small amounts of PGs, if allowed to persist at the site of injury, may in time cause pain.

PGE2 stimulates bone resorption,32 and its 13,14-dihydro derivative is nearly as potent, which is of interest because derivatives of the biologically active PGs are usually assumed to be of no functional significance. Addition of serum to the culture medium stimulates bone resorption, a process that is complement dependent and PG mediated. The mechanism may help explain bone erosion in joints of patients with RA, in which complement is activated, and PGE2 concentrations are high. The observation that PGE2 can stimulate bone formation33 suggests that PGs physiologically participate in coordination of bone formation and resorption. For example, primary idiopathic hypertrophic osteoarthropathy, a familial disorder, is associated with mutations in PDGHR and subsequent impairment of PG degradation. These patients have chronically elevated levels of PGE and exhibit digital clubbing and evidence of both increased bone formation and resorption in their phalanges.34 Many effects of IL-1 and TNF on cells are associated with stimulation of PG production and inflammation. Cartilage explants from osteoarthritis patients express COX-2 (but not COX-1) and release 50 times more PGE2 in culture than cartilage from healthy subjects and 18 times more PGE2 than normal cartilage stimulated with cytokines plus lipopolysaccharide. Explants from osteoarthritis patients release IL-1β, whereas normal cartilage does not express mRNA for pro-IL-1β or release IL-1β spontaneously. It seems that in osteoarthritis—and probably in RA—upregulation of cartilage IL-1β and subsequent production of PGE2 leads to cartilage degradation.35,36

Mast cells, often overlooked as important in inflammatory responses, are seen in large numbers in synovium from patients early in the course of RA.37 PGI2, the major PG formed by mast cells, can also mediate histamine release from mast cells exposed to anti-IgE antibody. PGI2, formed from the dehydration of PGI2, seems to function as a brake on the inflammatory response. It reduces macrophage activation, reduces nitric oxide production from stimulated cells, and induces apoptosis in tumor cell lines. PGI2 is metabolized to 15-deoxy-Δ12,14 PGJ2 and Δ12 PGJ2, which are also biologically active.38 It is not clear whether PGA, formed from loss of water from the cyclopentane ring, has biologic activity. However, in vitro, PGA2 induces apoptosis in HL-60 cells.39

**Prostacyclin**

Prostacyclin, discovered in 1976,40 has been purified, and the cDNA for prostacyclin synthase has been cloned. In addition to a cyclopentane ring, a second ring is formed by an oxygen bridge between carbons 6 and 9. It is generated from PGH2 by a distinct prostacyclin synthase, a 56-kDa member of the cytochrome P-450 superfamily of enzymes found predominantly in endothelial and vascular smooth muscle cells.15 Production of prostacyclin can be stimulated by thrombin or generated by transfer of PGH2 from platelets (the endoperoxide steal), contact with activated leukocytes, or stretching of the arterial wall. It is a powerful vasodilator and inhibits platelet aggregation through activation of adenosine 5′-triphosphate cyclase, which leads to an increase in intracellular cAMP. It is metabolized rapidly (half-life in plasma is less than one circulation time) to the more stable, less biologically active 6-keto-PGF1α. The enzymatic products of its conversion—2,3-dinor-6-keto-PGF1α and 6,15-dinor 2,3-dinor PGF1α—are also chemically stable and have little biologic activity. They are the major metabolites of prostacyclin excreted in urine, in which they can be assayed as indicators of prostacyclin generation.

Prostacyclin generated in the vessel wall has antiplatelet and vasodilator actions, whereas TXA2 generated by platelets from the same precursors induces platelet aggregation and vasoconstriction. These two eicosanoids represent biologically opposite poles of a mechanism for regulating the interaction between platelets and the vessel wall and of formation of hemostatic plugs and intra-arterial thrombi. Given the central role of platelets in inflammatory reactions, an appropriate prostacyclin-TX balance is important to regulation of inflammation. The balance may be altered in patients with antiphospholipid antibody syndrome; in patients treated with cyclosporine; and in patients treated with NSAIDs, especially patients treated with selective COX-2 inhibitors. Although COX-2 inhibitors reduce recurrence of colorectal adenomas, they increase the risk for cardiovascular events such as myocardial infarction and stroke.41

Intravascular infusion of prostacyclin also reduces some of the clinical changes associated with pulmonary embolism. The instability of prostacyclin makes it cumbersome to administer therapeutically. Nonetheless, it has been used with limited success to treat peripheral vascular disease including Raynaud’s phenomenon. New therapeutic approaches for treatment of pulmonary hypertension including patients with sarcoidosis and systemic sclerosis are prostacyclin analogues such as epoprostenol, beraprost, and iloprost, but these are not free of adverse events due to intravenous administration.42,43 Several drugs including angiotensin-converting enzyme inhibitors and statins release endogenous prostacyclin from vascular endothelium in vivo, observations that pave the way for development of new anti-inflammatory and cardioprotective agents.44 In addition to its vasodilator effects, prostacyclin suppresses endothelial cell proliferation. Prostacyclin analogues might prove useful as adjunct treatment for some cancers.
Thromboxanes

The endoperoxide PGH₂ can be converted into TXs after the action of the enzyme TX synthase, a microsomal 60-kD member of the cytochrome P-450 family, which is quite active in the platelet. The gene that encodes the enzyme has been cloned. TXs contain a six-member oxane ring instead of the cyclopentane ring of the PGs. TX synthase converts PGH₂ into equal amounts of TXA₂ and 12L-hydroxy-5,8,10-heptadecatrienoic acid. TXA₂ stimulates platelet activation, contributes to intravascular aggregation of platelets, and contracts arteriolar and bronchiolar smooth muscles. It is hydrolyzed rapidly (half-life is 30 seconds) to the inactive, stable, measurable product, TXB₂; its actions are limited to the microenvironment of its release.

The extraordinary rapidity with which platelets adhere to damaged tissue, aggregate, and release potent biologically active materials suggests that the platelet is well suited to be a cellular trigger for the inflammatory process. Efforts directed at suppression of TX synthesis and platelet aggregation may result in limitation of inflammatory responses, especially in coronary arteries. Inhibition of platelet aggregation may be important to the anti-inflammatory effects of aspirin and other NSAIDs. Long-term administration of low doses of aspirin (40 mg/day—the lowest dose predicted to cause total inhibition of TX formation in serum, according to mathematic modeling)—has inhibitory effects on platelet function ex vivo that are indistinguishable from the effects caused by giving 325 mg/day of aspirin. Aspirin acetylation of COX in platelets occurs in the portal vein where the aspirin concentration is high before it is metabolized in the liver, which explains why the 81-mg dose of aspirin is so effective in prevention of heart attacks and strokes. Platelets lose their ability to aggregate until new platelets are formed in about a week. Nonetheless, the response to aspirin can vary among individuals. It has been suggested that serum TXB₂ levels be monitored to ensure the efficacy of aspirin therapy. Even low-dose aspirin has some depressive effect on prostacyclin synthesis, however. This means recovery of vascular prostacyclin production may be more difficult to attain in elderly patients.

Selective inhibition of TX synthase represents an approach that may be used to suppress TXA₂ synthesis without depressing prostacyclin formation. The endoperoxide stapide seems to function in vivo after administration of a TX synthase inhibitor. Antagonists of the receptors shared by endoperoxide and TXA₂ have been developed, and these agents inhibit platelet aggregation in patients who are recalcitrant to TX synthase inhibition. Novel agents targeting TX receptors and TX synthase have improved treatment of TX receptors and TX synthase inhibition. Novel agents targeting TX receptors and TX synthase have improved treatment of TXA₂ and TXB₂ production may be more difficult to attain in elderly patients.

Lipoxygenase Pathways

In contrast to the COX pathway, in which stable products have three atoms of oxygen covalently attached to AA from 2 moles of molecular oxygen, lipoxygenases insert a single oxygen atom into the molecular structure of AA. Separate lipoxygenases exist in certain cells and have strict structural requirements for their substrates. Three major mammalian lipoxygenases insert their oxygen atoms into the 5, 12, or 15 position of AA, with formation of a new double bond and hydroperoxy group. The hydroperoxy fatty acids (hydroxyperoxy-eicosatetraenoic acid [HPETE]) can be reduced by peroxidases in the cell to yield the corresponding hydroxy fatty acids (hydroxy-eicosatetraenoic acid [HETE]). The exclusive lipoxygenase product of the human platelet is 12-HETE, which on reduction of the hydroperoxy group yields 12-HETE. In contrast, the human neutrophil makes predominantly 5-HETE, but when high concentrations of AA are added, 15-lipoxygenase can be shown. Lipoxygenases that act on AA are found in the cytosol fraction of cells.

The human 5-lipoxygenase (5-LO) gene has been isolated and characterized and produces a 78-kD enzyme. In myeloid cells the 5-lipoxygenase pathway leads to formation of the biologically active LTs (Figure 24-5) that were originally found in leukocytes and that contain three conjugated double bonds (triienes). Cell activation leads to translocation of 5-lipoxygenase from cytosol to the nuclear membrane, where it encounters the 18-kD, 5-lipoxygenase-activating protein (FLAP). AA is also translocated to 5-lipoxygenase-activating protein FLAP for presentation to 5-lipoxygenase (5-LO). In addition, upon cell stimulation, cPLA₂ is activated and also associates with the nuclear membrane, close to FLAP. The ability of macrophages and dendritic cells to respond appropriately during innate immune responses is likely regulated by 5-LO and 12-LO.

The unstable HPETE is the initial metabolite of each lipoxygenase pathway. HPETE is reduced to the more stable HETE or is converted by 5-LO to LTD₄. LTD₄ can be converted to LTE₄ (in neutrophils and macrophages) or conjugated with reduced glutathione to form LTC₄ (in eosinophils, mast cells, endothelial cells, and macrophages). In contrast to lipoxygenase, which is mainly distributed in myeloid cells, LTC₄ hydrolase (5,12-dihydroxy-eicosatetraenoic acid), a zinc-requiring enzyme that converts LTD₄ to LTE₄, is widely distributed. From the cDNA sequence, it was suggested that mRNA for LTD₄ may have a short half-life, which could account for the properties of extremely rapid production and shutdown of LTE₄ and other eicosanoid biosynthesis.

LTA₄ can be exported from the cell of origin and converted in other cells by LTA₄ hydrolase to LTD₄. This variation on the endoperoxide steal—perhaps better called transcellular metabolism—also applies to conversion of LTA₄ to LTC₄ by LTC₄ synthase, a glutathione-S-transferase. Although human endothelial cells do not produce terminal products of the 5-lipoxygenase system, they do generate LTC₄ from LTA₄ provided by neutrophils. LTC₄ and its products, LTD₄ and LTE₄, constitute the biologic mixture previously known as slow-reacting substance of anaphylaxis. LTD₄ and LTE₄ arise from LTC₄ after sequential removal of γ-glutamyl acid and glycine from LTC₄. The enzyme γ-glutamyl transpeptidase is present in many cells as part of a complex enzymatic system involved in glutathione biosynthesis and amino acid transport. In many systems the major sulfidopeptide LT has been reported to be LTD₄, rather than the precursor LTC₄. Removal of glycine from LTD₄ results in LTE₄ with concomitant loss of a significant
amount of biologic activity. The principal route of inactivation of LTB₄ is by omega oxidation.

**Products of the Lipoxygenase Pathways**

The biologic effects of compounds produced in the lipoxygenase pathway indicate their importance in inflammatory diseases. They are the major mediators of inflammation formed by the oxygenation of AA and are implicated as key mediators in several diseases including inflammatory bowel disease, psoriasis, bronchial asthma, and RA.

5-HETE and 5-HPETE stimulate the generation of superoxide in human neutrophils. These compounds also augment intracellular calcium levels, facilitating PKC-dependent activation of a superoxide generating system of neutrophils. LTB₄ increases adherence of leukocytes to endothelial cells, a response that is augmented by exposure of the endothelial cells to TNF. LTB₄ does not seem to have a direct vascular contractile action because it is inactive in the hamster cheek pouch preparation and several other microvasculature systems. In rabbit skin, administration of LTB₄ with a vasodilator PG induces plasma exudation, which suggests that LTB₄ may assist enhanced vascular permeability. Increased venule permeability does occur in response to LTC₄, LTD₄, and LTE₄. LTB₄ is a potent chemotactic factor for neutrophils and is weakly chemotactic for eosinophils. LTB₄ and, to a lesser extent, 5-HETE enhance migration of T lymphocytes in vitro. Synovial cells produce 5-HETE but do not seem to produce significant amounts of LTB₄. However, macrophages that invade the synovium in RA patients generate substantial quantities of 5-lipoxygenation and 15-lipoxygenation products including
LTB₄. In addition to the local signs of inflammation induced by products of the lipoxygenase pathway, these compounds may contribute to the pain, tenderness, and aching common in RA patients. LTB₄ also seems to serve an immunoregulatory function. It stimulates differentiation of competent CD8⁺ T lymphocytes from precursors lacking the CD8 marker. LTB₄ also stimulates interferon-γ and IL-2 production by T cells and biosynthesis of IL-1 by monocytes.⁵⁰

Synovial cell proliferation and endothelial cell proliferation are central to propagation of the rheumatoid joint lesion. LTB₄ and the cysteiny1 LTs act as growth or differentiation factors for several cell types in vitro. These compounds also increase proliferation of fibroblasts when PG synthesis is inhibited,⁵⁷ findings that emphasize the importance of interactions between the COX and lipoxygenase pathways and suggest limitations to NSAID therapy for RA patients.

Strategies for inhibiting production or antagonizing the actions of LTs include development of selective LT receptor antagonists and inhibition of the production of LT by blocking the action of 5-lipoxygenase. Inhibition of enzymes distal in the LT cascade such as LTA₄ hydrolase is also a promising strategy for development of anti-inflammatory drugs. A compound that inhibits binding of 5-lipoxygenase to 5-lipoxygenase-activating protein exhibits anti-inflammatory effects in animal models. Lipoxygenase inhibitors have not been useful for treatment of RA patients. New compounds, mainly resulting from natural products chemistry, seem more promising.⁵⁹ In addition, the existing agents may be aiming at the wrong target. Fibroblasts such as synovial cells do not make much LTB₄, but they do make 12-HETE, which is a growth factor, through a cytochrome P-450 pathway.⁶⁰ Cytochrome P-450 inhibitors may be more to the therapeutic point. Some inhibitors are more effective when activated by exposure to light.⁶¹ These inhibitors might prove useful as topical agents for treatment of inflammatory skin disease.

Lipoxygenase activities do not lead solely to production of mediators of inflammation. DGLA is converted by 15-lipoxygenase into 15-HETE, which is incorporated into DAG and exerts anti-inflammatory effects partly by interfering with PKC-β activity. A lipoxygenase product of linoleic acid, 13-hydroxyoctadecadienoic acid, also suppresses inflammation and cell proliferation by means of a similar mechanism.⁶² EPA is converted by lipoxygenase into 15-hydroxyeicosapentaenoic acid, which also exhibits anti-inflammatory properties.⁶³

### Lipoxins

Another large family of AA metabolites arises from the sequential action of 15-lipoxygenase and 5-lipoxygenase on arachidonic acid.

**Figure 24-6** Lipoxin biosynthesis. The lipoxins result from the sequential action of 15-lipoxygenase and 5-lipoxygenase on arachidonic acid. Introduced. Platelet 12-lipoxygenase can transform neutrophil LTA₄ to lipoxins. The complete stereochemistry and multiple routes of biosynthesis for the biologically active LXA₄ and LXB₄ have been determined.⁶⁴ That macrophages of rainbow trout generate lipoxins rather than LTs or PGs as their major products of AA metabolism indicates that lipoxins have a long evolutionary history. LTs and lipoxins can be generated in parallel in fish. In humans, the process has diverged to a two-cell system. Biosynthesis of eicosanoids by transcellular and cell-cell interactions is recognized as an important way to generate and amplify lipid-derived mediators. Lipoxins can be generated within the vascular lumen during platelet-leukocyte interactions and at mucosal surfaces via leukocyte–epithelial cell interactions. In humans, lipoxins are formed in vivo during multicellular responses such as inflammation, atherosclerosis, and in asthma. These tetraene-containing
products serve as stop signals in that they prevent leukocyte-mediated tissue injury. Acute inflammation is a primitive protective response, and chronic inflammation involves failure of mechanisms designed to stop the acute response. A major problem in joints of patients with RA, as well as of patients with other conditions associated with chronic inflammation and tissue injury, is that inflammation often does not resolve. Lipoxins and aspirin-induced 15-epilipoxins are endogenous components of events governing resolution of inflammation. Aspirin acetylation of COX-2 in endothelial cells suppresses PG synthesis but leads to generation of 15R-HETE from AA, which is transformed to 15-epilipoxin by leukocytes in a transcellular biosynthetic route involving vascular endothelial cells or epithelial cells. These 15-epilipoxins exhibit anti-inflammatory and antiproliferative actions in vitro and in vivo. Stable analogues of LXA4 and of aspirin-triggered lipoxin (ATL) also suppress inflammation in animal models. In addition, acting as antagonists at the receptor CysLT1, ATL analogues counteract the inflammatory actions of cysteiny1 leukotrienes. A stable synthetic analogue of LXA4 suppresses several interferon-induced genes and reduces kidney damage in a murine model of immune-mediated nephritis.

Lipoxins block human polymorphonuclear leukocyte chemotaxis but stimulate monocyte chemotaxis and adherence. Monocytes do not release mediators of inflammation in response to lipoxins, however, and lipoxins are converted rapidly by monocytes to inactive compounds. This selective effect on chemotaxis suggests that lipoxins can play a role in wound healing. LXA4 antagonizes LTD4-induced vasoconstriction in vivo and blocks binding of LTD4 to its receptors on mesangial cells. LXA4 suppresses LTB4-induced plasma leakage and leukocyte migration and blocks LTB4-induced neutrophil inositol trisphosphate generation and calcium mobilization, but not superoxide anion generation. Conversely, LXA4 activates PKC and is more potent in this regard than DAG and AA. LXA4 seems to be specific for the γ subspecies of PKC. These results indicate that lipoxins may regulate the actions of vasoconstrictor LTs and suggest that LXA4 may be an important modulator of intracellular signal transduction.

It was long thought that resolution of inflammation was a passive process. Contrary to this belief, the inflammatory response does not dissipate but is mediated actively by products of COX and LO and acetylation of these enzymes. Discovery that omega-6 and omega-3 polyunsaturated fatty acids are also substrates for enzymatic oxygenation reactions that produce lipid mediators with potent anti-inflammatory and proresolving actions has advanced our understanding of the inflammatory response. In addition to lipoxins, entire spectra of specialized proresolving mediators (SPM) have been uncovered. These include resolvins, protectins, and maresins. Like the lipoxins, resolvins are formed via specific transcellular biosynthetic pathways from both n-6 and n-3 fatty acids. EPA-derived (E series) cellular interaction products formed in resolving exudates have been termed resolvins (RvE1 and E2). The omega 3 docosahexaenoic acid (DHA) can be catalyzed via the 15-LO pathway to dihydroxy products (D series resolvins D1-D5), which also stimulate resolution of inflammation and enhance innate host defense mechanisms. Lipoxins and resolvins act as endogenous receptor agonists at low concentrations (pM to nM) and at specific G protein–coupled membrane spanning receptors to actively downregulate inflammatory events and to stimulate resolution of an inflammatory exudate. A single cell type can also form oxygenated lipid mediators with counterregulatory actions. DHA can also be catalyzed via the 15-LO pathway to a dihydroxy product termed protectin D1 (PD1), which can protect against tissue damage and activate resolution of inflammation. In a second single oxygenation route, a lipid mediator termed maresin (macrophage mediators in resolving inflammation) is formed via the action of human 12-LO. The maresin biosynthetic pathway is activated in macrophages during phagocytosis. Maresin 1 (MaR1) reduces neutrophil migration and increases macrophage phagocytosis of apoptotic cells, thereby exhibiting the hallmarks of SPM. An animal model that has been developed more closely resembles a chronic inflammatory response and allows measurement of the extent and pace of the response including resolution and its sensitivity to anti-inflammatory and proresolving therapies.

**Isoeicosanoids**

Isoeicosanoids, isomers of enzymatically derived eicosanoids, are derived from the free radical–mediated peroxidation auto-oxidation of polyunsaturated fatty acids including omega-3 fatty acids. They include members of the F, D, and E prostanes, isotromboxanes, and isoleukotrienes. Analysis of the isoprostanes (IsoP) indicates that they reflect lipid peroxidation in vivo. Theoretically, 64 F-type isoprostanes may be formed, although few have been characterized. As more isoprostanes are identified, the nomenclature probably will continue to change. One F-type isoprostane, isoprostane F1α, III (formerly 8-isoprostaglandin F1α) has been studied in detail because of its biologic activity in vitro. Iso prostane F1α, III (Figure 24-7) is a potent vasoconstrictor and may function as a mitogen, with actions that are blocked by TX receptor antagonists. Although isoprostanes may act as ligands at TX or PG receptors (8,12-isoprostane F1α, III activates the PGF2α receptor), they may also activate specific isoprostane receptors.

**Isoprostanes**

What all isoprostanes have in common, and what distinguishes them from the PGs, is the fact that the top (α) and the bottom (ω) side chains are always syn—that is, crowded together on the same face of the cyclopentane ring. The minimal requirement for generation of an isoprostane is a polyunsaturated fatty acid with three contiguous methylene-interrupted double bonds, a requirement met by dozens of naturally occurring polyunsaturated fatty acids. Brain tissue contains particularly high levels of fatty acids formed from docosahexaenoic acid (DHA; C22:6n3). By the same peroxidation process undergone by AA, DHA in nerve cell membranes is converted to several neuroprostanes (NeuroP). Of these, F4-NeuroP is a promising biomarker for several neurodegenerative diseases. A similar mechanism in plants
yields phytoprostanes (PhytoP) from α-linolenic acid. In contrast to conventional, enzymatically derived PGs that are formed intracellularly and released immediately, isoprostanes are formed in the cell membrane, cleaved by phospholipases, circulate in plasma, and are excreted in urine. As stable end products of lipid peroxidation, endogenously formed IsoPs are useful markers of oxidative stress and independent risk markers for coronary artery disease. They are increased in several diseases including acute respiratory distress syndrome, in which polymorphonuclear leukocytes generate reactive oxygen species that damage pulmonary epithelium. The immune cells in inflamed tissues are exposed to reactive oxygen intermediates produced by neutrophils and other phagocytic cells. Oxidants are also generated as mediators in intracellular signaling pathways by cytokines such as IL-1β and TNF. Isoprostanes are likely to be proved to be important in inflammatory conditions such as vasculitis and RA. Because isoprostanes are released preformed, their production is not blocked by NSAIDs, which suppress metabolism of free AA. It is possible that inflammation unresponsive to NSAIDs may yield to inhibition of isoprostanes. In keeping with the notion of eicosanoids as regulators, however, some isoprostanes suppress release of mediators of inflammation from macrophages, and a PhytoP (E1-PhytoP) signals through PPARγ and NFκB to suppress IL-12 production by activated dendritic cells and reduce cytokine production by Th1 and Th2 cells.  

Endocannabinoids

Groups of naturally occurring members of the eicosanoid superfamily that can activate cannabinoid receptors and are derivatives of long-chain fatty acids have been referred to as *endocannabinoids*. They are not stored in cells. Rather, they are synthesized rapidly in the manner of PGs and LTs from lipid precursors and are released from—among other cells—cells of the immune system during immune/inflammatory responses. They can then activate cannabinoid receptors on the same or adjacent cells and are metabolized rapidly by the specific serine hydrolase, fatty acid amide hydrolase (FAAH), or by monoglyceride lipase or N-acetylthanolamine. Thus inhibition of FAAH offers a strategy for treatment of chronic inflammatory pain. One of the most important endocannabinoids is anandamide (from the Sanskrit word for “bliss”), the amide conjugate of AA and ethanolamine (arachidonoyl ethanolamide) (Figure 24-8). Anandamide is not stored in cells. Rather, it is synthesized rapidly in response to stimuli in the manner of PGs and LTs. Anandamides and other endocannabinoids such as 2-arachidonoylglycerol and virodamine are involved in a wide range of regulatory functions including pain perception and modulation of immune responses, actions mediated via cannabinoid 1 (CB1) and CB2 receptor subtypes resulting in activation of G proteins of the G(i/o) family. Of particular relevance to joint tissue injury is the capacity of anandamide to suppress TNF-induced NFκB activation by direct inhibition of I-κ-B kinase, an action that is independent of CB1 or CB2 activation.

Acid congeners of anandamides are lipoamino acids (elmiric acids) that exist as endogenous substances, regulate tissue levels of anandamide, and exhibit anti-inflammatory effects and the capacity to assist resolution of inflammation. One such compound, N-arachidonoylglycine (NAGly), is found in many tissues at higher concentrations than anandamide. It has analgesic actions similar to those reported for anandamide but does not exhibit psychotropic action. The polyunsaturated amides dihomogammalinolenoyl (20:3n6) and adrenoyl (22:4 n6) ethanolamide are found in mammalian brain. It is likely that n-3 fatty acid ethanolamides also exist in mammalian tissues. A library of elmiric acids (n-3 and n-6) has been synthesized. Several of these compounds exhibit anti-inflammatory activity, likely by increasing production of PGI2 and also seem to modulate immune responses. Discovery of naturally occurring and synthetic analogues and metabolites of anandamide suggest that a family of such biologically active substances exists. The polyunsaturated amides dihomogammalinolenoyl (20:3 n-6) and adrenoyl (22:4 n-6) ethanolamides have been found in mammalian brain. It is likely that n-3 fatty acid ethanolamides also exist in mammalian tissues. That anandamide can enhance its own synthesis in macrophages suggests the presence of a rapid response to counter excessive inflammatory or immune responses. Anandamide is

**Figure 24-7** Isoprostane IsoP structures. Jagged lines indicate that stereochemistry is uncertain.

**Figure 24-8** Chemical structure of anandamide (arachidonoyl ethanolamide).
converted by COX-2 (but not by COX-1) into PGE2 or PGF2α. Ethanolamide directly, without going through free AA.25 These novel PGs ("prostamides") are pharmacologically active.26 Because anandamide is a substrate for COX-2, inhibitors of COX-2 may reduce anandamide metabolism with a subsequent increase in concentration of the anandamide. A combination of anandamide with ibuprofen produces synergistic analgesia in rats.26 In addition, anandamide is metabolized by fatty acid amide hydrolase. Compounds that inhibit fatty acid amide hydrolase and maintain anandamide levels may be useful anti-inflammatory agents. A naturally occurring lipoxygenc acid, N-arachidonoylglycine (NAGly), is also a substrate for COX-2, giving rise to amino acid conjugates of the PGs.31

**EICOSANOID RECEPTORS**

For many years, it was thought that the lipophilic eicosanoids—in contrast to the peptide molecules for which receptors were characterized routinely—simply “diffused” into cell membranes or were carried in by a binding protein. The isolation and cloning of eicosanoid receptors changed that thinking.2

**Prostaglandin Receptors**

PGs exert most of their actions through G protein–coupled receptors (GPCRs). They also bind to peroxisome proliferator-activated receptors (PPARs) in vitro, but it is not clear that PPARs mediate PG effects in vivo. Receptors for the COX products are designated P receptors, depending on the prostanoid that has most affinity for them. These receptors include the PGD receptors (DP); four subtypes of the PGE receptor (EP1 through EP4); the PGF receptor (FP); the PGJ2 receptor (IP); and the TX receptor (TP). The IP, DP, EP2, and EP4 receptors mediate increases in cellular cAMP, whereas TP, FP, and EP1 receptors induce calcium mobilization. EP2, EP4, and IP regulate macrophage cytokine production in a similar manner. As might be expected, signaling through these receptors is more complicated. Studies designed to understand PGE2 signaling suggest involvement of PI3 kinase, mitogen-activated protein kinase (MAPK), and Wnt pathways in EP2R/EP4R regulation of cell growth, migration, and apoptosis.34 PGE2 is the most abundant of the major eicosanoids in the joint space of patients with RA. It is therefore of interest that thinking.

Further evidence of the importance of eicosanoids in immune cell function and regulation of immune responses derives from experiments with mice deficient in PG receptors, which indicate that PGE2 assists and amplifies IL-12-mediated Th helper 1 (Th1) cell differentiation and IL-23-mediated Th17 cell expansion. These PGE2 actions and those of PGD2 and PGI2—also discovered using...
receptor-deficient mice—contribute to development of immune diseases. Mice lacking the PGF2α receptor are protected against bleomycin-induced pulmonary fibrosis, a finding with implications for treatment of systemic sclerosis.

The implications for therapy derived from this new knowledge are clear and exciting, but prostanooid analogues with selective binding properties need to be developed. Some progress has been made, and it seems that deletion of P receptors, with the exception of EP4, is not associated with serious problems of fetal development or physiologic function in animals.

**Leukotriene Receptors**

Less is known about the surface receptors for LTs. Surface receptors for LTB4, denoted BLT1 and BLT2 (LTB4 R-1 and LTB4 R-2), and for the cysteinyl LTs also exert their actions through transmembrane spanning GPCRs. High-affinity LTB4 receptors transduce chemotaxis and adhesion responses, whereas low-affinity receptors are responsible for secretion of granule contents and superoxide generation. The BLT1 receptor has been cloned and characterized as a 43-kD GPCR with seven transmembrane spanning domains, is expressed in inflammatory cells, and has a high degree of specificity for LTB4, with a Kd of 0.15 to 1 nm. BLT2 is homologous to BLT1 but has lower affinity for LTB4 (Kd 23 nm) and a different specificity and binding profile for BLT antagonists. BLT2 knockout mice express normal levels of BLT1 but are protected from development of disease in the K/BxN model of inflammatory arthritis. The COX-1–derived ligand 12(S)-hydroxyeicosatetraenoic acid (12-HHT), produced during thromboxane synthesis, is an endogenous high-affinity ligand for BLT2 (now called BLT2/HHTR), another example of a connection between the LO and COX pathways. LTB4 in particular among the LTs is involved in development of atherosclerosis, which has emerged as a major concern for patients with inflammatory arthritis and lupus. The receptor for cysteinyl LTs includes two subtypes—Lys LTα and Cys LTα—that have been identified pharmacologically, although their molecular structures are unknown. The cysteinyl LTs are recognized by at least two GPCRs denoted CysLT1 and CysLT2. CysLT1 contains 336 amino acid residues, and the gene encoding the receptor is located on the X chromosome. CysLT2 contains 345 amino acids with 40% sequence identity to CysLT1. CysLT1 mediates calcium mobilization and inhibition of adenylate cyclase, whereas CysLT2 mediates calcium mobilization and increased cAMP concentrations. The preferred ligands for CysLT1 are LTD4 > LTC4 > LTE4. CysLT2 binds LTC4 and LTD4 equally, whereas LTE4 exhibits low affinity for the receptor. Both receptors have wide tissue and cellular distribution including a presence in cells which participate in immune responses. Most actions of the cysteinyl LTs are mediated by CysLT1. More than a dozen chemically distinct, specific, and selective antagonist drugs that block the binding of LT to CysLT1 have been identified. Clinical use of these compounds has mainly been in the treatment of asthma. A 5-lipoxigenase-specific inhibitor reduced whole-blood LTB4 production but did not suppress synovitis in a 4-week trial of RA patients. A challenge in designing “antireceptor therapy” is the genetic variation in GPCRs that can be associated with disease. Adding to the complexity is the fact that variants may result in altered predisposition to disease, rather than manifestation of the disease. Each variant provides an opportunity to understand receptor function such as recycling or desensitization, enhancing the potential for development of therapy.

**Lipoxin Receptors**

Lipoxins can act at their own specific receptors for LXA4 and LXB4, and LXA4 can interact with a subtype of LTD4 receptors. Lipoxins can also act at intracellular targets within their cell of origin or after uptake by another cell. The cDNA for the seven-transmembrane-spanning, G protein–coupled LXA4 receptor named ALX/FPR2 (Kd = 0.7 nM) has been cloned and characterized. Its signaling involves a novel polyisoprenyl-phosphate pathway that regulates phospholipase D. Lipoxin actions are cell type specific. The monocyte and neutrophil LXA4 receptors are identical at the cDNA level, but they evoke different responses and the LXA4 receptor on endothelial cells seems to be a structurally distinct form. LXA4 also binds to the human orphan receptor GPR32, a member of the chemoattractant receptor family. Like LXA4, 15-epi-LXA4 is an anti-inflammatory SPM that binds and activates ALX/FPR2 (Kd = 2 nM). Lipoxin B4 (LXB4) and aspirin-triggered 15-epi LXB4 also have anti-inflammatory actions by oral administration and topical application. Stereoselective actions of these compounds indicate they have their own yet-to-be-identified receptors. The resinol RvE1 binds to the GPCR CMKL1 (Kd = 11 nM). The functional importance of this interaction is demonstrated by downregulation of IL-12 in murine dendritic cells. RvE1 also exerts partial agonism at the LTB4 receptor BLT1 (Ki = 70 nM). Displacement of LTB4 constitutes a mechanism whereby RvE1 dampens the inflammatory actions of LTB4. Experimental evidence has been obtained for the existence of specific receptors for D-series resolvins. The LXA4 receptor ALX/FPR2 also binds RvD1. Thus two counterregulatory lipid mediators are ligands for the same receptor. The mechanism of that process is not clear.

**Nuclear Receptors**

Nuclear receptors are a superfamily of ligand-regulated transcription factors that interact with other transcription factors and with co-regulators that either enhance (co-activators) or inhibit (co-repressors) transcription. The major nuclear receptors involved in regulation of inflammation are the glucocorticoid receptors (GRs), peroxisome proliferator activated receptors, liver X receptors (LXRs), and the orphan receptor nuclear receptor related protein (Nurr1). Other members of the nuclear receptor family that contribute to regulation of inflammation include estrogen receptors, vitamin D receptor, and retinoic acid receptors. The clinical efficacy of glucocorticoids is well known, but knowledge of their mechanisms of action has been slow to emerge. The ability of GR to repress inflammatory responses is due in part to interference with other signal-dependent transcription factors and disruption of activator/co-activator complexes. PPARs are members of the nuclear receptor family of transcription factors, a large and diverse group of
proteins that mediate ligand-dependent transcriptional activation and repression. PPARs were first cloned as nuclear receptors that mediate the effects on gene transcription of synthetic compounds called peroxisome proliferators. Several mechanisms account for the anti-inflammatory action of PPARγ including inhibition of NFκB; inhibition of transcription of genes encoding for chemokines, IL-1β, IL-12, and MMP-9; and promotion of expression of anti-inflammatory mediators including IL-10 and LXR. LXRs regulate immune synapse formation in dendritic cells and reduce T cell proliferation. Polyunsaturated fatty acids, including γ-linolenic acid (GLA), also act via PPARγ by stimulating phosphorylation and translocation to the nucleus. Interest in PPARs increased dramatically when they were shown to be activated by medically relevant compounds including NSAIDs and PGD2 and its metabolite 15-deoxy-Δ12,14-PGJ2. Most information available on a potential role of PPARs in inflammation relates to PPARγ. Upregulation of PPARγ reduces expression of several mediators of inflammation, raising the possibility that PPARγ ligands may be therapeutic for diseases characterized by inflammation. However, the metabolite 15-deoxy-Δ12,14-PGJ2, also can exhibit anti-inflammatory activity in a PPARγ-independent manner. PPARα is expressed mainly in tissues that have a high fatty acid catabolism including liver and the immune system. LTB4 is an activator and natural ligand of PPARα. Activation of PPARα results in induction of genes involved in fatty acid oxidation pathways that degrade fatty acids and derivatives including LTB4. Thus a feedback mechanism that controls inflammation is established. Mechanisms for the anti-inflammatory actions of PPARβ/δ include induction of an anti-inflammatory co-repressor protein, inhibition of NFκB, and induction of anti-inflammatory mediators. Experiments with PPAR knockout mice indicate that PPARα suppresses LTB4-induced inflammation. PPARs are being considered as therapeutic targets for a wide range of immune-mediated diseases characterized by chronic inflammation. Overexpression of Nurr receptors reduces inflammatory cytokine expression, and mutations in the Nurr 1 gene are associated with diminished counterregulation of inflammation.

Several kinases that facilitate co-repressor turnover have been identified. These kinases represent important pharmacologic targets because their inhibition should block gene expression of inflammatory mediators while bypassing the clinically significant adverse events associated with direct targeting of the nuclear receptors. A small molecule inhibitor of c-Jun terminal kinase is effective treatment in animal models of arthritis.

**PLATELET-ACTIVATING FACTOR**

Platelet-activating factor (PAF, 1-0-alkyl-2-acetyl-sn-glycero-3-phosphocholine) is a potent mediator of inflammation that causes neutrophil activation, increased vascular permeability, vasodilation, and bronchoconstriction in addition to platelet activation. PAF is formed by a smaller number of cell types than the eicosanoids, mainly leukocytes, platelets, and endothelial cells. Because of the extensive distribution of these cells, however, the actions of PAF can manifest in virtually every organ system. In contrast to the two long-chain acyl groups present in phosphatidylcholine, PAF contains a long-chain alkyl group joined to the glycerol backbone in an ether linkage at position 1 and an acetyl group at position 2 (Figure 24-9). PAF represents a family of phospholipids (PAF-like lipids: PAF-LL) because the alkyl group at position 1 can vary in length from 12 to 18 carbons. PAF, similar to the eicosanoids, is not stored in cells. Rather, it is synthesized when cells are stimulated, at which time the composition of the alkyl group may change. The immediate effects of PAF are mediated through a cell surface GPCR, PAFR. PAFR is coupled to G1, Gq, and G12/13. Activation of PAFR results in inhibition of cyclic AMP, mobilization of calcium, and activation of mitogen-activated protein kinases, whereas long-term responses depend on intracellular—probably nuclear—receptor activation.

Figure 24-9 Chemical structure of platelet-activating factor.

Despite the potent inflammatory effects of PAF, its inhibition in animal models does not lead to marked suppression of inflammatory responses. The synthesis of PAF is tightly regulated, and a family of intracellular and extracellular phospholipases A2 (PAF acetylhydrolases [PAF AHs]) degrade PAF and PAF-LL, thereby regulating their half-life and their engagement with the PAFR. In addition, receptor desensitization controls binding of PAF to its receptor. Plasma PAF Ah derives from macrophages, dendritic cells, and platelets during an inflammatory response and can serve as a circulating marker of inflammation and of atherosclerosis. Mast cells also produce PAF during immediate hypersensitivity reactions, and PAF Ah reduces mortality in murine models of anaphylaxis. Plasma PAF acetylhydrolase, an enzyme that hydrolyzes PAF, may be a particularly important terminator of PAF-induced tissue injury and may find a place among strategies designed to suppress inflammation. Strategies to interrupt cellular activation mediated by dysregulation of PAF signaling include competitive blockade of the PAFR with receptor antagonists and use of recombinant PAF Ah to hydrolyze PAF and PAF-LL upstream from receptor occupancy. In addition, PAF production is suppressed by adenosine, which may account for some of the therapeutic efficacy of methotrexate. Given the involvement of PAF in immediate hypersensitivity reactions and inflammation, further search for PAF antagonists is warranted.

**EICOSANOIDS AS REGULATORS OF INFLAMMATION AND IMMUNE RESPONSES**

The role of PGs in the inflammatory process is not as well defined as previously supposed because in addition to their well-known actions as mediators of inflammation, the stable
PGs PGE and PGI₂ have anti-inflammatory, inflammatory, and immunomodulatory actions. As noted, PGJ, lipoxins, and an array of eicosanoids seem to act as brakes to protect against runaway inflammatory responses. Even LTB₄ is capable of modulating inflammation and immune responses. The observations that PGE₁ inhibits platelet aggregation and that it suppresses acute and chronic inflammation and joint tissue injury in animal models led to the notion that COX products of AA metabolism might have anti-inflammatory activity. As it became more clear that NSAIDs have anti-inflammatory effects other than interference with COX production and subsequent PG inhibition, consideration was given to the potential protective effects of PGs.

PGE₁ has remained an orphan among the eicosanoids, mainly because of a long-held notion that not enough of it is made by human cells to be of use and that its biologic effects are no different from the effects of PGE₂ and PGI₂. Contrary to popular belief, PGE₁ is found in physiologically important amounts in humans. Lost in the vast literature on the “AA cascade” are the early observations of Bygdeman and Samuelsson, who found (using bioassay) the concentration of PGE₁ in human seminal plasma (16 µg/mL) to be higher than PGE₂ (13 µg/mL), PGE₃ (3 µg/mL), PGF₂α (2 µg/mL), and PGE₄ (12 µg/mL). Karim and colleagues found PGE₁ to be the sole PGE in the human thymus. PG immunoassays usually do not distinguish between PGE₁ and PGE₂. To identify PGE₁, it must first be separated from PGE₂ by thin-layer or high-performance liquid chromatography. When such methods have been used, PGE₁ has been identified consistently in platelets, leukocytes, macrophages, vas deferens, oviducts, uterus, heart, and skin. Evidence from in vitro and in vivo experiments indicates that PGs, notably PGE compounds, can suppress diverse effector systems of inflammation. PGE can enhance and diminish cellular and humoral immune responses, observations that reinforce a view of these compounds as regulators of cell function. These actions of eicosanoids depend on the stimulus to inflammation, the predominant eicosanoid produced at a particular time in the host response, and the profile of eicosanoid-receptor expression.

It is now clear that the 2 series prostaglandins (E₂, D₂, I₂) also regulate T cell function and immune responses. PGE₂ reduces production of several inflammatory cytokines including TNF, IFN-γ, and IL-12 and reduces IFN-α production by plasmacytoid dendritic cells (PDCs) from patients with systemic lupus erythematosus (SLE). PGE₂-treated PDCs from SLE patients also induce CD4⁺ T cell proliferation and skew cytokine production toward a Th2 profile. Another example of an endogenous link between the COX and LO pathways is provided by the observation that PGE₁ preserves resolution of inflammation in the murine collagen–induced arthritis model by increasing production of the proresolving lipoxin A₄.

MODULATION OF EICOSANOID SYNTHESIS BY ADMINISTRATION OF PRECURSOR FATTY ACIDS

The relationship between essential fatty acids and PGs was discovered simultaneously and independently by van Dorp and colleagues and Bergstrom and colleagues. Both groups reported that AA was converted to PGE₂ and shortly thereafter they showed that PGE₁ is formed from DGLA. Attempts to modulate eicosanoid production have been directed at providing fatty acids other than AA as substrates for oxygenation enzymes in an effort to generate a unique eicosanoid profile with immunosuppressive and anti-inflammatory effects. The fatty acids themselves, by virtue of their incorporation into signal-transduction elements, also have effects that are independent of eicosanoid effects on cells involved in inflammation and immune responses.

Experiments directed at suppression of TX synthesis, enhancement of prostacyclin production, and inhibition of platelet aggregation have been done in an effort to limit inflammatory responses. EPA is not found in appreciable amounts in cells from individuals who eat a Western-style diet. Fish oil lipids, rich in EPA (20:5 n-3), inhibit formation of COX products (e.g., TXA₂, PGE₂) derived from AA, and the newly formed TXA₁ has much less ability than TXA₂ to constrict vessels and aggregate platelets. Production of PGI₂ (prostacyclin) by endothelial cells is not reduced appreciably by increased EPA content, and the physiologic activity of newly synthesized PGI₂ is added to that of PGI₁. Administration of fish oil to humans leads to reduced production of LTB₄ by means of 5-LO in stimulated neutrophils and monocytes and induces EPA-derived LTB₃, which is far less biologically active than LTB₄. Fish oil also reduces production of IL-1β, TNF, and PAF by activated blood monocytes. Meta-analysis of randomized controlled trials of administration of fish oil to patients with RA indicate reduction in tender joint counts and duration of morning stiffness and decreased use of NSAIDs. Because NSAIDs confer an increased risk for cardiovascular disease and there is increased mortality from cardiovascular disease in patients with RA, an added benefit of fish oil for RA patients may be reduced risk of cardiovascular disease directly and by virtue of less use of NSAIDs. Fish oil supplements in the treatment of RA for 6 to 12 months result in significant reductions in number of tender joints and time of morning stiffness compared with the same measures done at baseline. Fish oil treatment allowed patients to reduce or stop NSAID treatment. As noted earlier, after acetylation by aspirin, COX-2 acquires the capacity to oxygenate EPA and DHA, leading to formation of novel resolvins and protectins, compounds that assist resolution of inflammation.

The other “alternative” eicosanoid precursor fatty acid, DGLA (20:3 n-6), can also be increased by administration of certain plant seed oils, notably oils extracted from the seeds of *Oenothera biennis* (evening primrose) and *Borago officinalis* (borage), which contain relatively large amounts of GLA. GLA is converted to DGLA, the immediate precursor of PGE₁, an eicosanoid with known anti-inflammatory and immunoregulating properties. Administration of GLA to volunteers and RA patients results in increased production of PGE₁ and reduced production of the inflammatory eicosanoids PGE₂, LTB₄, and LTC₄ by stimulated peripheral blood monocytes. In addition to competing with AA for oxidative enzymes, DGLA cannot be converted to inflammatory LTs. Rather, it is converted by means of 15-LO to a 15-hydroxy-DGLA,
which has the capacity to inhibit 5-lipoxygenase and 12-LO activities. DGLA should have anti-inflammatory actions because of its capacity to reduce synthesis of oxygenation products of AA through the COX and the lipoxygenase pathways.\textsuperscript{117,121}

In addition to their roles as precursors of eicosanoids, essential fatty acids are important for the maintenance of cell membrane structure and function and protect the gastric mucosa from NSAID-induced injury. DGLA can also modulate immune responses in an eicosanoid-independent manner. DGLA suppresses IL-2 production by human peripheral blood monocytes in vitro, suppresses proliferation of IL-2-dependent human peripheral blood and synovial tissue T lymphocytes, and reduces expression of activation markers on T lymphocytes directly in a manner that is independent of its conversion to eicosanoids. Oral administration of oils enriched in GLA but not administration of oils enriched in linoleic acid (the parent n-6 fatty acid) or α-linolenic acid (the parent n-3 fatty acid) reduce proliferation of human lymphocytes activated through the T cell receptor complex.\textsuperscript{121}

Addition to peripheral blood mononuclear cells in vitro or administration of GLA in vivo reduces secretion of IL-1β and TNF from stimulated human cells. GLA also reduces autoinduction of IL-1β in human monocytes, preserving the protective effects of IL-1β, while suppressing excessive production of the cytokine.\textsuperscript{126} IL-1β and TNF are important polypeptide mediators of inflammation and joint tissue injury in patients with RA, and both cytokines are targets for biologic agents that have proven to be major advances in treatment of patients with RA. GLA suppresses acute and chronic inflammation including arthritis in several animal models, and randomized, double-blind, placebo-controlled trials of GLA in patients with RA and active synovitis indicate that GLA treatment results in statistically significant and clinically relevant reduction in signs and symptoms of disease activity compared with baseline and placebo. GLA also reduces the need for NSAID and corticosteroid therapy.\textsuperscript{126,127}

EPA suppresses conversion of DGLA to AA, and a combination of EPA-enriched and GLA-enriched oils exhibits synergy in its capacity to reduce synthesis of synovitis in animal model.\textsuperscript{128} In addition, administration of black currant seed oil, which contains the n-3 fatty acid α-linolenic acid (which is converted to EPA) and the n-6 GLA, suppresses active synovitis in patients with RA.\textsuperscript{129} These observations that particular marine and botanical oils reduce signs and symptoms of joint inflammation and have a positive impact on the profile of serum lipids suggest that both GLA- and EPA/DHA-enriched oils, or the isolated fatty acids themselves, or a combination of these might be useful treatment for diseases characterized by chronic inflammation and tissue injury. A combination of GLA and EPA may be useful therapy for RA patients. Continued study of the eicosanoids and their precursor fatty acids should delineate mechanisms by which these lipids influence the function of cells that participate in immune responses and inflammatory reactions.

References

29. von Euler US: On the specific vasodilating and plain muscle stimulating substances from accessory genital glands in man and certain animals (prostaglandin and vesiglandin), J Physiol (Lond) 88:213, 1936.


The references for this chapter can also be found on www.expertconsult.com.
Cell Recruitment and Angiogenesis

ZOLTAN SZEKANECZ • ALISA E. KOCH

ENDOTHELIAL PATHOPHYSIOLOGY IN INFLAMMATION

Endothelial cells are active players in inflammation. The vascular endothelium undergoes vasodilatation and increased permeability (leakage) during synovitis. Increased endothelial permeability results from several mechanisms including endothelial contraction and retraction, leukocyte- or antiendothelial antibody (AECAs)-mediated vascular injury and regeneration. The endothelium secretes several inflammatory mediators resulting in vasodilatation and leakage including prostacyclin (PGI₂), nitric oxide (NO), platelet-activating factor (PAF), and others. In turn, endothelial cells respond to histamine, serotonin, complement factors (C3a, C5a), bradykinin, leukotrienes, PAF, and AECAs that are released in inflammatory sites. Cytoskeletal reorganization leading to endothelial retraction may be regulated by proinflammatory cytokines such as interleukin (IL)-1, tumor necrosis factor (TNF), or interferon-γ (IFN-γ). Among other soluble mediators, AECAs that have been detected in several inflammatory rheumatic conditions have been correlated with clinical activity and vascular damage.

High-endothelial venules (HEVs) are usually detected in lymphoid tissues, and they are major sites of leukocyte extravasation during the homing process. Such HEVs have been described at least in some synovial tissues with lymphoid neogenesis.

INTERCELLULAR ADHESION MOLECULES

Process of Leukocyte Extravasation in Inflammation

Adhesion of peripheral blood leukocytes to endothelium leads to the process of leukocyte transendothelial migration into inflammatory sites such as the arthritic synovium. High endothelial venules (HEVs) primarily found in lymphoid organs are also present at sites of lymphoid neogenesis in the synovium. Lymphocytes recirculate through HEVs during homing, and thus inflammatory leukocyte recruitment may be considered as “pathologic homing” (see Figure 25-1).

During leukocyte adhesion and transendothelial migration, an early, weak adhesion termed rolling occurs first. This step involves selectins and their ligands and leads to leukocyte activation. Activation-dependent, firm adhesion involves mostly integrin-dependent interactions, as well as the secretion of numerous chemokines. Chemokines preferentially attract endothelium-bound leukocytes (see Figure 25-1).
**Adhesion Receptors and Ligands**

CAMs have been classified into integrin, selectin, immunoglobulin, and cadherin superfamilies. (Table 25-1). E-, P-, and L-selectin contain a lectin-like extracellular N-terminal domain, an epidermal growth factor (EGF)-like motif, and two to nine moieties related to complement regulatory proteins. E- and P-selectin are expressed by ECs, whereas L-selectin is mostly expressed by leukocytes. During leukocyte transendothelial migration, selectins mediate the initial tethering and rolling of leukocytes. E-selectin is a marker for cytokine-induced EC activation. E-selectin ligand-1 (ESL-1) and P-selectin ligand-1 (PSGL-1) contain sialylated glycan motifs such as sialyl Lewis-X (sLeX). P-selectin is constitutively present on the membrane of EC Weibel-Palade bodies. P-selectin is involved in the early phases of leukocyte-EC adhesion. L-selectin serves as a lymphocyte homing receptor, where it mediates the physiologic recirculation of naïve lymphocytes through specialized HEV. However, L-selectin has also been implicated in inflammatory leukocyte recruitment. All three selectins are expressed in the arthritic synovium (see Table 25-1). Integrins are αβ heterodimers. Each of the common β chains is associated with one or more α subunits. Cell adhesion to the extracellular matrix (ECM) is mostly mediated by β3 and β1, whereas intercellular adhesion is assisted through β1 and β2 integrins. β3 and β1 integrins are expressed on ECs, whereas β2 integrins are leukocyte CAMs. Integrin-mediated adhesion and migration have been associated with arthritis. The αα-αα, αα-αβ, and β1-β1 integrin subunits have all been detected in the inflamed synovium. The immunoglobulin superfamily of CAMs is a group of transmembrane glycoproteins containing one or more immunoglobulin-like motifs of 60 to 100 amino acids. Vascular cell adhesion molecule-1 (VCAM-1) is

**Table 25-1 Relevant Members of Adhesion Molecule Superfamilies**

<table>
<thead>
<tr>
<th>Adhesion Receptors</th>
<th>Ligands</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selectins</strong></td>
<td></td>
</tr>
<tr>
<td>L-selectin (CD62L, LAM-1)</td>
<td>Sialylated carbohydrates, GlyCAM-1</td>
</tr>
<tr>
<td>E-selectin (CD62E, ELAM-1)</td>
<td>Sialy-Lewis-X</td>
</tr>
<tr>
<td>P-selectin (CD62P, PAGDEM)</td>
<td>Sialy-Lewis-X, other carbohydrates</td>
</tr>
<tr>
<td><strong>Integrins</strong></td>
<td></td>
</tr>
<tr>
<td>αβ (VLA-1)</td>
<td>Laminin, collagen</td>
</tr>
<tr>
<td>αβ (VLA-2)</td>
<td>Laminin, collagen</td>
</tr>
<tr>
<td>αβ (VLA-3)</td>
<td>Laminin, collagen, fibronectin</td>
</tr>
<tr>
<td>αβ (VLA-4)</td>
<td>Fibronectin, VCAM-1</td>
</tr>
<tr>
<td>αβ (VLA-5)</td>
<td>Fibronectin</td>
</tr>
<tr>
<td>αβ (VLA-6)</td>
<td>Laminin</td>
</tr>
<tr>
<td>αβ (LFA-1, CD11a/CD18)</td>
<td>ICAM-1, ICAM-2, ICAM-3, JAM-A</td>
</tr>
<tr>
<td>αβ (Mac-1, CD11b/CD18)</td>
<td>ICAM-1, ICAM-2, ICAM-3, JAM-A</td>
</tr>
<tr>
<td>αβ</td>
<td>iC3b, fibrinogen</td>
</tr>
<tr>
<td>αβ</td>
<td>E-cadherin</td>
</tr>
<tr>
<td>αβ</td>
<td>Fibrinectin, VCAM-1, MadCAM-1</td>
</tr>
<tr>
<td><strong>Immunoglobulin Superfamily</strong></td>
<td></td>
</tr>
<tr>
<td>ICAM-1 (CD54)</td>
<td>LFA-1, Mac-1</td>
</tr>
<tr>
<td>ICAM-2</td>
<td>LFA-1</td>
</tr>
<tr>
<td>ICAM-3</td>
<td>LFA-1</td>
</tr>
<tr>
<td>VCAM-1</td>
<td>αβ, αβ</td>
</tr>
<tr>
<td>MadCAM-1</td>
<td>αβ, L-selectin</td>
</tr>
<tr>
<td>CD2</td>
<td>LFA-3</td>
</tr>
<tr>
<td>PECAM-1 (CD31)</td>
<td>PECAM-1, αβ</td>
</tr>
<tr>
<td><strong>Cadherins</strong></td>
<td></td>
</tr>
<tr>
<td>E-cadherin (cadherin-1)</td>
<td>E-cadherin</td>
</tr>
<tr>
<td>N-cadherin (cadherin-2)</td>
<td>N-cadherin</td>
</tr>
<tr>
<td>Cadherin-11</td>
<td>Cadherin-11</td>
</tr>
</tbody>
</table>

CHUNK 1: Chemokines are small proteins that exert chemotactic activity toward leukocytes. They include CD44, vascular adhesion proteins (VAP-1 and VAP-2), endoglin, E- and N-cadherin, cadherin-11, and junctional adhesion molecules (JAMs). CD44 is a receptor for hyaluronate and is expressed on activated ECs in inflammatory sites. VAP-1 has originally been isolated from synovial ECs. The expression of VAP-1 is increased in RA. Endoglin (CD105) is a receptor for transforming growth factor (TGF)-β1. Endoglin is involved in EC adhesion and is expressed in synovitis. Among other CAMs, cadherins are primarily involved in embryogenesis; however, synovial fibroblast cadherin-11 has been implicated in arthritis as well (see Table 25-1).

CHUNK 2: Chemokines are small proteins that exert chemotactic activity toward leukocytes. They include CD44, vascular adhesion proteins (VAP-1 and VAP-2), endoglin, E- and N-cadherin, cadherin-11, and junctional adhesion molecules (JAMs). CD44 is a receptor for hyaluronate and is expressed on activated ECs in inflammatory sites. VAP-1 has originally been isolated from synovial ECs. The expression of VAP-1 is increased in RA. Endoglin (CD105) is a receptor for transforming growth factor (TGF)-β1. Endoglin is involved in EC adhesion and is expressed in synovitis. Among other CAMs, cadherins are primarily involved in embryogenesis; however, synovial fibroblast cadherin-11 has been implicated in arthritis as well (see Table 25-1).

CHUNK 3: Chemokines are small proteins that exert chemotactic activity toward leukocytes. They include CD44, vascular adhesion proteins (VAP-1 and VAP-2), endoglin, E- and N-cadherin, cadherin-11, and junctional adhesion molecules (JAMs). CD44 is a receptor for hyaluronate and is expressed on activated ECs in inflammatory sites. VAP-1 has originally been isolated from synovial ECs. The expression of VAP-1 is increased in RA. Endoglin (CD105) is a receptor for transforming growth factor (TGF)-β1. Endoglin is involved in EC adhesion and is expressed in synovitis. Among other CAMs, cadherins are primarily involved in embryogenesis; however, synovial fibroblast cadherin-11 has been implicated in arthritis as well (see Table 25-1).

CHUNK 4: Chemokines are small proteins that exert chemotactic activity toward leukocytes. They include CD44, vascular adhesion proteins (VAP-1 and VAP-2), endoglin, E- and N-cadherin, cadherin-11, and junctional adhesion molecules (JAMs). CD44 is a receptor for hyaluronate and is expressed on activated ECs in inflammatory sites. VAP-1 has originally been isolated from synovial ECs. The expression of VAP-1 is increased in RA. Endoglin (CD105) is a receptor for transforming growth factor (TGF)-β1. Endoglin is involved in EC adhesion and is expressed in synovitis. Among other CAMs, cadherins are primarily involved in embryogenesis; however, synovial fibroblast cadherin-11 has been implicated in arthritis as well (see Table 25-1).

CHUNK 5: Chemokines are small proteins that exert chemotactic activity toward leukocytes. They include CD44, vascular adhesion proteins (VAP-1 and VAP-2), endoglin, E- and N-cadherin, cadherin-11, and junctional adhesion molecules (JAMs). CD44 is a receptor for hyaluronate and is expressed on activated ECs in inflammatory sites. VAP-1 has originally been isolated from synovial ECs. The expression of VAP-1 is increased in RA. Endoglin (CD105) is a receptor for transforming growth factor (TGF)-β1. Endoglin is involved in EC adhesion and is expressed in synovitis. Among other CAMs, cadherins are primarily involved in embryogenesis; however, synovial fibroblast cadherin-11 has been implicated in arthritis as well (see Table 25-1).

CHUNK 6: Chemokines are small proteins that exert chemotactic activity toward leukocytes. They include CD44, vascular adhesion proteins (VAP-1 and VAP-2), endoglin, E- and N-cadherin, cadherin-11, and junctional adhesion molecules (JAMs). CD44 is a receptor for hyaluronate and is expressed on activated ECs in inflammatory sites. VAP-1 has originally been isolated from synovial ECs. The expression of VAP-1 is increased in RA. Endoglin (CD105) is a receptor for transforming growth factor (TGF)-β1. Endoglin is involved in EC adhesion and is expressed in synovitis. Among other CAMs, cadherins are primarily involved in embryogenesis; however, synovial fibroblast cadherin-11 has been implicated in arthritis as well (see Table 25-1).

CHUNK 7: Chemokines are small proteins that exert chemotactic activity toward leukocytes. They include CD44, vascular adhesion proteins (VAP-1 and VAP-2), endoglin, E- and N-cadherin, cadherin-11, and junctional adhesion molecules (JAMs). CD44 is a receptor for hyaluronate and is expressed on activated ECs in inflammatory sites. VAP-1 has originally been isolated from synovial ECs. The expression of VAP-1 is increased in RA. Endoglin (CD105) is a receptor for transforming growth factor (TGF)-β1. Endoglin is involved in EC adhesion and is expressed in synovitis. Among other CAMs, cadherins are primarily involved in embryogenesis; however, synovial fibroblast cadherin-11 has been implicated in arthritis as well (see Table 25-1).

CHUNK 8: Chemokines are small proteins that exert chemotactic activity toward leukocytes. They include CD44, vascular adhesion proteins (VAP-1 and VAP-2), endoglin, E- and N-cadherin, cadherin-11, and junctional adhesion molecules (JAMs). CD44 is a receptor for hyaluronate and is expressed on activated ECs in inflammatory sites. VAP-1 has originally been isolated from synovial ECs. The expression of VAP-1 is increased in RA. Endoglin (CD105) is a receptor for transforming growth factor (TGF)-β1. Endoglin is involved in EC adhesion and is expressed in synovitis. Among other CAMs, cadherins are primarily involved in embryogenesis; however, synovial fibroblast cadherin-11 has been implicated in arthritis as well (see Table 25-1).

CHUNK 9: Chemokines are small proteins that exert chemotactic activity toward leukocytes. They include CD44, vascular adhesion proteins (VAP-1 and VAP-2), endoglin, E- and N-cadherin, cadherin-11, and junctional adhesion molecules (JAMs). CD44 is a receptor for hyaluronate and is expressed on activated ECs in inflammatory sites. VAP-1 has originally been isolated from synovial ECs. The expression of VAP-1 is increased in RA. Endoglin (CD105) is a receptor for transforming growth factor (TGF)-β1. Endoglin is involved in EC adhesion and is expressed in synovitis. Among other CAMs, cadherins are primarily involved in embryogenesis; however, synovial fibroblast cadherin-11 has been implicated in arthritis as well (see Table 25-1).

CHUNK 10: Chemokines are small proteins that exert chemotactic activity toward leukocytes. They include CD44, vascular adhesion proteins (VAP-1 and VAP-2), endoglin, E- and N-cadherin, cadherin-11, and junctional adhesion molecules (JAMs). CD44 is a receptor for hyaluronate and is expressed on activated ECs in inflammatory sites. VAP-1 has originally been isolated from synovial ECs. The expression of VAP-1 is increased in RA. Endoglin (CD105) is a receptor for transforming growth factor (TGF)-β1. Endoglin is involved in EC adhesion and is expressed in synovitis. Among other CAMs, cadherins are primarily involved in embryogenesis; however, synovial fibroblast cadherin-11 has been implicated in arthritis as well (see Table 25-1).
asthma, a known Th2 type disease, is rather associated with CCR3, CCR4, and CCR8 tissue expression.\textsuperscript{12,64,65}

**Inflammatory and Homeostatic Chemokines: Is It a Justified Classification?**

Chemokines have recently been functionally classified into these subgroups.\textsuperscript{10,56} As many functions of these chemokines overlap, this classification may not be really justified. Numerous CXC, CC, and CX3C chemokines implicated in the pathogenesis of arthritis are termed inflammatory chemokines. These chemokines recruit mostly effector cells including monocytes, neutrophils, and T cells into tissues.\textsuperscript{9,10,56,66}

As included in Table 25-2, there is a great body of evidence supporting the role in RA of CXCL1 (growth-regulated oncogene α; GROα); CXCL4 (platelet factor 4; PF4); CXCL5 (epithelial-neutrophil activating protein 78; ENA-78); CXCL6 (granulocyte chemotactic protein 2; GCP-2); CXCL7 (connective tissue activating protein III; CTAP-III); CXCL8 (interleukin 8; IL-8); CXCL9 (monokine induced by interferon-γ; Mig); CXCL10 (interferon-γ-inducible 10-kD protein; IP-10); CXCL12 (stromal cell-derived factor 1; SDF-1); CXCL13 (B cell–activating chemokine 1; BCA-1); and CXCL16. Among CC chemokines, CCL2 (monocyte chemoattractant protein 1; MCP-1); CCL3 (macrophage inflammatory protein 1α; MIP-1α); CCL5 (Regulated upon Activation, Normal T cell Expressed and Secreted; RANTES); CCL19 (Epstein-Barr virus–induced gene 1 ligand chemokine; ELC); CCL20 (MIP-3α); and CCL21 (secondary lymphoid tissue chemokine; SLC) have been implicated in leukocyte recruitment underlying inflammatory synovitis. Finally, CXCL1 (fractalkine) and XCL1 (lymphotactin) are also considered as inflammatory chemokines.\textsuperscript{7,9,56} Accordingly, CXCR1-CXCR6, CCR1-CCR6, and XCR1 and CXCR1 are involved in the pathogenesis of RA.\textsuperscript{10,56,67}

Homeostatic chemokines are constitutively produced in microenvironments of lymphoid or nonlymphoid tissues such as in the skin or mucosa. These chemokines promote lymphocyte homing into these tissues, a process associated with the physiologic function of the adaptive immune system. Lymphocyte recirculation is involved in antigen sampling and immune surveillance.\textsuperscript{10,68,69} Among CXC chemokines, CXCL12 (SDF-1), CXCL13 (BCA-1), and CXCL16, as well as their respective receptors, CXCR4, CXCR5, and CXCR6, exert such effects.\textsuperscript{10,68,69} Among CC chemokines, CCL17 (TARC), CCL19 (ELC), CCL21 (SLC), CCL22 (MDC), CCL25 (TECK), CCL27 (CTACK), CCL28 (MEC), as well as their receptors, CCR4, CCR7, CCR9, and CCR10, are involved in the homeostasis of lymphoid tissues.\textsuperscript{50,68,69} As described earlier, among others, CXCL12, CXCL13, CXCL16, CCL19, and CCL21 have also been implicated in arthritis-associated inflammatory cell recruitment and synovial lymphoid neogenesis.\textsuperscript{7,10,56,67-71} The synovium is, in many ways, similar to mucosa-associated lymphoid tissues (MALT), which may explain the dual role of some chemokines in physiologic lymphocyte homing and inflammation.\textsuperscript{10,68,69,71}

### ANGIOGENESIS AND VASCULOGENESIS IN INFLAMMATION

**Angiogenesis and Vasculogenesis**

Angiogenesis is the formation of new capillaries from preexisting blood vessels, whereas vasculogenesis is the outgrowth of vessels from endothelial progenitor cells (EPCs).\textsuperscript{7,13-15,72-77} Angiogenesis may increase the total endothelial surface and thus may enable leukocyte extravasation into inflammatory sites.\textsuperscript{14,15} The perpetuation of angiogenesis has been associated with inflammatory diseases such as RA or psoriasis and in malignancies.\textsuperscript{14,15,77} The outcome of such “angiogenic diseases” is dependent on the balance or imbalance between angiogenic mediators and angiostatic factors.\textsuperscript{7} Several cytokines, growth factors, chemokines, certain CAMs, and other mediators can modulate neovascularization in inflammation\textsuperscript{14,15,73,77} (Table 25-3; see Figure 25-1).

**Angiogenic Factors**

The hypoxia-vascular endothelial growth factor (VEGF)-angiopoietin system seems to be of outstanding importance in arthritis-associated angiogenesis.\textsuperscript{73,70,80} VEGF is induced by hypoxia and hypoxia-inducible factors 1 and 2 (HIF-1, HIF-2) in RA.\textsuperscript{11,73,80-85} Significant hypoxia is characteristic of RA joints.\textsuperscript{73,86} Recently, the stimulatory effect of hypoxia on the angiogenic drive of RA synovial fibroblasts has been demonstrated.\textsuperscript{84,87} Hypoxia-inducible HIF-1 and HIF-2 are

<p>| Table 25-3 | Angiogenic and Angiostatic Factors in Rheumatoid Arthritis* |</p>
<table>
<thead>
<tr>
<th>Mediators</th>
<th>Inhibitors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chemokines</strong></td>
<td>CXCL1, CXCL5, CXCL7, CXCL8, CXCL12, CCL2, CCL21, CCL23, CX3CL1</td>
</tr>
<tr>
<td><strong>Matrix molecules</strong></td>
<td>Type I collagen, fibronectin, laminin, heparin, heparan sulfate</td>
</tr>
<tr>
<td><strong>Cell adhesion molecules</strong></td>
<td>β1 and β3 integrins, E-selectin, P-selectin, CD34, VCAM-1, endoglin, PECAM-1, VE-cadherin, Leu8, H, MUC-18</td>
</tr>
<tr>
<td><strong>Growth factors</strong></td>
<td>VEGF, bFGF, aFGF, PDGF, EGF, IGF-I, HIF-1, TGF-β†</td>
</tr>
<tr>
<td><strong>Cytokines</strong></td>
<td>TNF, IL-6†, IL-15, IL-18</td>
</tr>
<tr>
<td><strong>Proteases</strong></td>
<td>MMPs, plasminogen activators</td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td>Angiogenin, substance P, prolatin</td>
</tr>
</tbody>
</table>

*See text for abbreviations. See Table 25-2 for traditional chemokine designations.†Mediators with both proangiogenic and antiangiogenic effects.
strongly expressed in the RA synovium. However, hypoxia may also act via HIF-independent regulatory pathways including the peroxisome proliferator-activated receptors (PPARs). The angiopoietin-1 (Ang1)/Tie2 complex interacts with VEGF during the stabilization of newly formed blood vessels. In contrast, Ang2, an antagonist of Ang1, inhibits vessel maturation. Ang1 and Tie2 have been detected in the RA synovium.

Apart from VEGF, other growth factors involving fibroblast growth factors (FGF-1 and FGF-2), transforming growth factor β (TGF-β), connective tissue growth factor (CTGF), and platelet-derived growth factor (PDGF) have been implicated in synovial angiogenesis. Recently, the role of placenta growth factor (PIGF) in RA and inflammatory angiogenesis has been postulated. PIGF, like VEGF, binds to the flt-1/VEGF-R1 receptor. There is abundant expression of PIGF in the synovial tissue of arthritic mice.

Proinflammatory cytokines may exert direct angiogenic activity or may act indirectly via VEGF-dependent pathways. Primarily TNF, IL-1, IL-6, IL-15, IL-17, IL-18, oncostatin M, granulocyte (G-CSF), and granulocyte-macrophage colony-stimulating factors (GM-CSFs) have been implicated in synovial angiogenesis. Monocyte migration inhibitory factor (MIF) has been implicated in angiogenesis, as well as atherosclerosis, a major cause of death in RA patients.

Among numerous other angiogenic mediators not mentioned earlier, serum amyloid A (SAA), an important acute-phase reactant, has also been implicated in arthritis and angiogenesis. Interaction of SAA with formyl peptide receptor-like 1 (FPRL1) induces endothelial cell proliferation, migration and angiogenesis, and synovitis.

Further angiogenic factors include, without mentioning further details, endothelin 1 (ET-1), members of the cyclooxygenase-2 (COX-2)-prostaglandin E2 network, angiogenin, angiopoietin, pleiotrophin, platelet-activating factor (PAF), substance P, erythropoietin, adenosine, histamine, prolactin, thrombin, and sphingosine-1-phosphate (SIP).

The involvement of chemokines, chemokine receptors, and CAMs in angiogenesis is discussed later in context with the complex regulation of leukocyte recruitment.

### Vasculogenesis in Inflammatory Conditions

EPCs are hematopoietic stem cells expressing, among other antigens, CD34, CD133, type 2 VEGF receptor (VEGFR-2 or Flk-1), and the CXCR4 chemokine receptor. During vasculogenesis, EPCs differentiate into mature endothelial cells. Vasculogenesis is involved in tissue development, vascular repair, atherosclerosis, and inflammation.

Several groups have described defective vasculogenesis related to impaired EPC numbers and functions in RA and scleroderma. Impaired vasculogenesis has been associated with increased cardiovascular morbidity and mortality in these disease states. Effective control of inflammation using corticosteroids and anti-TNF agents may stimulate EPCs and thus may restore defective vasculogenesis. In addition, the induction of vasculogenesis may be beneficial for patients with cardiovascular disease and the stimulation of EPCs and vasculogenesis may also suppress premature atherosclerosis in RA.

### INTERACTIONS AMONG ADHESION RECEPTORS, CHEMOKINES, AND ANGIogenesis: THE “REAL” BERMUDA TRIANGLE IN THE REGULATION OF INFLAMMATORY SYNOVITIS

#### Chemokines and Adhesion Receptors

The molecular mechanisms and signaling pathways of chemokine-induced CAM expression have been described. Briefly, an atypical protein kinase C, PKC-ζ, has been identified. Treatment of cells with chemokines induces PKC-ζ kinase activity through its interaction with PI3K. This leads to increased cell surface integrin expression via further signaling steps.

Another example for chemokine-CAM interactions is the regulation of β3 integrin expression by CCL2 (MCP-1) through the Ets-1 transcription factor, and the ERK-1/2 cascade. This pathway is involved in both inflammation and angiogenesis. The CCL21-CCR7 interaction results in the stimulation of LFA-1- and ICAM-1-dependent adhesion. Stimulation of CXCR1- and CXCR2-dependent pathways in the antigen-induced arthritis (AgLA) model resulted in increased neutrophil adhesion to endothelium. Thus various chemokines and chemokine receptors are involved in driving leukocyte transendothelial migration.

#### Chemokines and Chemokine Receptors in Angiogenesis and Vasculogenesis

Numerous CXC and CC chemokines, fractalkine, and their receptors have been implicated in angiogenesis underlying RA. The angiogenic nature of most CXC chemokines has been associated with the glutaryl-leucyl-arginyl (ELR) amino acid motif within their structure. ELR-containing CXC chemokines that mediate angiogenesis include CXCL1 (GROα), CXCL5 (ENA-78), CXCL7 (CTAP-III), and CXCL8 (IL-8). In contrast, the ELR-lacking CXCL4 (PF4), CXCL9 (Mig), and CXCL10 (IP-10) inhibit angiogenesis. Interestingly, some authors suggest that the effect of VEGF on endothelial cells may be, in part, mediated by CXCL10. It seems that CXCL12 (SDF-1) may play a significant role in RA-associated angiogenesis despite the fact that this chemokine lacks the ELR motif. CXCL12 is also a major mediator of lymphoid neogenesis in the RA synovium. Hypoxia stimulates CXCL12 production by RA synovial fibroblasts. CXCL12 has even been implicated in vasculogenesis. Virtually all EPCs express CXCR4 and migrate in response to SDF-1/CXCL12.

Much less evidence is available regarding the role of CC chemokines in angiogenesis. CCL2 (MCP-1) may induce endothelial cell chemotaxis in vitro and angiogenesis in vivo. As described earlier, CCL2-induced angiogenesis may occur via β1 integrins. CCL23 has been implicated in the migration of vascular endothelial cells and angiogenesis-associated matrix metalloproteinase (MMP) production. In contrast, CCL21 (SLC) may exert strong angiostatic and antitumor effects. CXCL1 (fractalkine) is involved in both angiogenesis and atherosclerosis underlying inflammatory rheumatic diseases.
Adhesion Receptors, Ligands, and Proteases in Angiogenesis and Vasculogenesis

Both CAM receptors and their extracellular matrix (ECM) macromolecule ligands mediate adhesive interactions during inflammatory neovascularization. Among ECM components, type I and other minor collagens, fibronectin, heparin, laminin, tenasin, vitronectin, and fibrinogen promote angiogenesis. Vasculogenesis is stimulated by the laminin matrix Matrigel. Thrombospondin-1 (TSP-1) is an angiostatic ECM component naturally produced within the RA synovium.

Among endothelial CAMs, soluble E-selectin; soluble P-selectin; the L-selectin ligand CD34; soluble VCAM-1; some endothelial β1, β3, and β4 integrins; PECAM-1 (CD31); endoglin (CD105); and some cadherins have been implicated in angiogenesis. The αvβ3 integrin and the ITGAV gene play critical roles in inflammatory angiogenesis. The integrin has become a major target for specific therapy. Other angiogenic factors such as chemokines may act via integrin-dependent pathways. Recently, the role of the ITGAV allele in angiogenesis has been further analyzed in four Caucasian sample sets. The genetic association could not be confirmed in New Zealand and Oxford (UK) sample sets, suggesting that the link between ITGAV gene polymorphism and RA may be limited.

Focal adhesion kinases (FAKs) are involved in αvβ3 integrin signaling underlying synovial inflammation and angiogenesis. Among glycoconjugates with adhesive properties, blood group antigens Lewis-y and Lewis-H promote neovascularization. JAMs (JAM-A, JAM-B, and JAM-C) have also been implicated in the adhesive processes underlying RA, as well as in synovial angiogenesis. Vasculogenesis also involves numerous integrins including αvβ3 and E-selectin. MMPs promote angiogenesis by synovial matrix degradation. The role of hypoxia in MMP production is described earlier. Some ADAM and ADAMTS proteases have also been implicated in inflammatory neovascularization.

Inhibition of Cell Adhesion Receptors and Leukocyte-Endothelial Adhesion

Traditional DMARDs including sulfasalazine, methotrexate (MTX), and leflunomide suppressed serum and synovial fluid soluble ICAM-1, VCAM-1, and E-selectin levels in both early and established RA, as well as juvenile idiopathic arthritis (JIA). MTX and leflunomide also decrease synovial tissue CAM expression in RA. Statins, currently used for the treatment of dyslipidemia, may also modify endothelial function and CAM expression.

Infliximab therapy reduced the serum levels of soluble ICAM-1, ICAM-3, VCAM-1, and E- and P-selectin in RA and JIA. Adalimumab therapy resulted in the attenuation of neutrophil chemotaxis in RA. Two anti-integrin strategies, the anti-LFA-1 antibody efalizumab and the LFA-3-Ig fusion protein alefacept, have been registered for the treatment of psoriasis. Alefacept yielded to a moderate effect in psoriatic arthritis. Efalizumab was withdrawn from the market in 2009 due to severe side effects. Other anti-LFA-1 antibodies have still been in preclinical arthritis studies. Natalizumab (anti-α4 integrin) has been tried in multiple sclerosis and Crohn’s disease. Vitaxin, a humanized antibody to the αvβ3 integrin, was administered to patients with ulcerative colitis and JIA. The combination of MTX and leflunomide also decreased CCR2 expression on monocytes isolated from RA patients. The combination of MTX and leflunomide inhibited the production of CCL2 (MCP-1), CCL17 (TARC), and CCL22 (MDC) in RA. Leflunomide itself suppressed CCL2 (MCP-1) and CCL5 (RANTES) levels in RA patients. Regarding biologics, most anti-TNF agents exert inhibitory effects on chemokine production. For example, infliximab reduced synovial expression of CXCL8 (IL-8) and CCL2 (MCP-1) in RA patients, which was associated with diminished inflammatory cell ingress into the synovium. Treatment of RA patients with infliximab or etanercept resulted in the sustained retention of CXCR3+

Targeting Cell Adhesion, Chemokines, and Angiogenesis: Possible Therapeutic Approaches in Inflammatory Arthritides

Leukocyte recruitment inhibition may be a result of non-specific anti-inflammatory therapeutic strategies. Numerous traditional and biologic disease-modifying drugs (DMARDs) and immunosuppressive agents may, in addition to other effects, suppress leukocyte recruitment, chemokine production, and angiogenesis. Inhibition of cell adhesion and migration, angiogenesis, chemokines, and chemokine receptors using specific antibodies or purified ligands has provided an important perspective on the molecular pathogenesis of RA. In addition, some of these strategies may be included in the future therapy of arthritis.*

*References 3, 4, 7, 14, 15, 72, 76, 134, 135.
T cells in the circulation indicating a clearance of these cells from the synovium. In recent studies, infliximab or etanercept also suppressed the release of CXCL1 (GROα), CXCL8 (IL-8), CXCL10 (IP-10), CXCL16, CCL2 (MCP-1), CCL5 (RANTES), CCL20 (MIP-3α), and CX3CL1 (fractalkine). Infliximab also reduced chemokine production in response to Mycobacteria in RA patients, which may have relevance for increased incidence of tuberculosis during anti-TNF therapy. Among newer biologics, tocilizumab also acts by suppressing CAM and chemokine production. Depletion of B cells by rituximab also interferes with the CXCL8 (IL-8) network. Interestingly, antioxidants such as N-acetyl-L cysteine and 2-oxothiazolidine-4-carboxylate inhibited mRNA expression of CXCL8 (IL-8) and CCL2 (MCP-1) by cytokine-pretreated human synovial fibroblasts.

Regarding direct chemokine receptor inhibition, most human trials using small molecule inhibitors of chemokine receptors failed. For example, MLN3897, an oral CCR1 antagonist, in combination with MTX had no significant clinical efficacy in RA. Similarly, SCH351125 and AZD5672, oral CCR5 inhibitors, did not give any clinical benefit in RA. Yet numerous CCR1, CCR2, and CCR5 antagonists are in clinical development in arthritis, as well as other inflammatory diseases.

A limited number of human antichemokine studies have been done. There has been one trial using an anti-CXCL8 (IL-8) antibody in RA, but results of this trial were not published and the further development of this compound was terminated. Antibodies or peptide inhibitors against CXCL4 (PF4), CXCL5 (ENA-78), CXCL8 (IL-8), CXCL10 (IP-10), CXCL16, CCL2 (MCP-1), and CX3CL1 (fractalkine) have been tried successfully in animal models of arthritis, but none of these agents has yet reached human development. Our group assessed a neutralizing polyclonal anti-CXCL5 antibody administered intravenously to rats using the AIA model. The antibody injected before the onset of arthritis attenuated the severity of the disease. This antibody also prevented the ingress of IL-1-expressing leukocytes into the synovium. Regarding chemokine receptor targeting in rodents, CXCR2, CXCR4, CCR1, and CCR5 antagonists have been tried in these animal models. Bicyclam, also known as AMD3100, a highly selective antagonist of CXCL12 (SDF-1)/CXCR4, inhibited inflammation and angiogenesis. The failure of numerous oral CCR1 and CCR5 antagonists in human trials is discussed earlier.

Some studies have addressed the use of combined chemokine blockade. For example, a combination of CCL2 (MCP-1) and CXCL1 (GROα) inhibition resulted in more pronounced arthritis suppression than CCL2 blockade alone in a murine AIA model. Certainly, there may be increased toxicity using combined strategies. Hence chemokine or chemokine receptor blockade using antibodies or other inhibitors may be promising for future therapies.

**Angiogenesis Targeting: Use of Angiostatic Compounds**

Angiogenesis can be inhibited by either blocking the action of angiogenic mediators or by using angiostatic compounds (Table 25-3, Figure 25-2). Focusing on leukocyte recruitment, anti-CAM antichemokine strategies are discussed earlier. A number of currently used antirheumatic agents such as dexamethasone, chloroquine, sulfasalazine, MTX, azathioprine, cyclophosphamide, lefunomide, thalidomide, minocycline, anti-TNF agents, and possibly cyclosporine A nonspecifically suppress angiogenesis.

VEGF inhibitors have been tried in arthritis and cancer studies. One can inhibit VEGF-mediated
neovascularization by using monoclonal antibodies to VEGF or VEGF receptors (VEGFR), soluble VEGFR constructs, small molecule VEGF and VEGFR inhibitors, or inhibitors of VEGF and VEGFR signaling.\(^\text{12,26,190}\) VEGF or VEGFR inhibition has been included in arthritis trials.\(^\text{192,195}\)

The VEGF tyrosine kinase inhibitor vatalanib (PTK787) and an anti-VEGFR1 antibody exerted significant angiostatic and antiarthritic effects in animal models of arthritis.\(^\text{194,195}\) A soluble VEGFR1 chimeric protein dose-dependently inhibited synovial endothelial proliferation.\(^\text{193}\) Soluble Fas ligand (sFasL, CD178) is a functional inhibitor of the 165 amino acid form of VEGF (VEGF165). sFasL inhibits angiogenesis in arthritis.\(^\text{196}\)

The involvement of PPARs in hypoxia-induced VEGF production is discussed earlier. PPARγ ligands rosiglitazone and pioglitazone inhibited VEGF-induced angiogenesis.\(^\text{197}\) Moreover, pioglitazone also improved joint and skin symptoms in psoriatic arthritis.\(^\text{198}\) Hypoxia-induced and VEGF-mediated neovascularization may also be targeted by the inhibition of HIFs.\(^\text{85,87}\) For example, YC-1, a superoxide-sensitive stimulator of soluble guanylyl cyclase that also inhibits HIF-1α, has been developed for the treatment of hypertension but may potentially be used to suppress inflammatory angiogenesis.\(^\text{199}\) The synthetic benzophenone analogue, BP-1, a HIF-1α inhibitor, ameliorated AIA in rats.\(^\text{200}\) In a recent trial, HIF-1 signaling was targeted in inflammatory bowel disease.\(^\text{201}\) RA synovial fibroblasts enhanced myeloid cell recruitment and angiogenesis in synovial tissues engrafted into immunodeficient mice. In this model, targeting HIF-1α expression by either siRNA or by the small molecule inhibitor chetomin significantly reduced these processes.\(^\text{202}\)

The role of the Ang1/Tie2 system in arthritis is described earlier.\(^\text{80,90}\) A soluble Tie2 receptor transcript delivered via an adenoviral vector to mice attenuated the incidence and severity of collagen-induced arthritis (CIA).\(^\text{203}\)

Regarding the targeting of other angiogenic growth factors, imatinib mesylate is a specific inhibitor of PDGF receptor activation. This compound inhibited pannus formation and the development of arthritis in the murine CIA model.\(^\text{204,205}\) The PPARγ agonists rosiglitazone and pioglitazone suppressed not only VEGF- but also bFGF-mediated neovascularization.\(^\text{197}\) An anti-flt-1 hexapeptide, GNQWF1, abrogated PI GF-induced angiogenesis, cytokine production, and the development of CIA in mice.\(^\text{92}\)

Matrix metalloproteinases (MMPs) and plasminogen activators are involved in matrix degradation underlying leukocyte recruitment and angiogenesis. Numerous protease inhibitors including tissue inhibitors of metalloproteinases (TIMPs) and plasminogen activator inhibitors (PAIs) that antagonize the effects of proteases have been tried in angiogenesis models.\(^\text{206,207}\)

Anticytokine therapy currently used to treat arthritides may influence angiogenesis, as well as cell adhesion and chemokines. For example, TNF blockade by infliximab reduced VEGF expression and vascularity within the RA synovium.\(^\text{208}\) Infliximab also reduced synovial Ang1 and Tie2 expression.\(^\text{209}\) The anti-IL-6 receptor antibody tocilizumab also decreased serum levels of VEGF in RA.\(^\text{210}\)

IL-4 and IL-13 are anti-inflammatory and angiostatic cytokines in RA.\(^\text{211,212}\) IL-4 suppresses VEGF release by RA synovial fibroblasts.\(^\text{211}\) IL-4 and IL-13 gene transfer inhibited synovial inflammation and angiogenesis in rats.\(^\text{212,213}\)

Antibiotic derivatives including minocyline, fumagillin analogues, deoxyspergualin, roxithromycin, and clarithromycin also inhibit the release of VEGF and other angiogenic mediators and thus neovascularization.\(^\text{214-216}\) Synthetic fumagillin derivatives TNP-470 and PPI-2458 inhibit VEGF, as well as methionine aminopeptidase-2, an enzyme involved in angiogenesis.\(^\text{217}\) In human RA trials minocycline, roxithromycin, and clarithromycin exerted moderate but significant clinical effects.\(^\text{216,218}\)

Among other angiogenic mediators, endothelin-1 antagonists currently used in the treatment of primary and scleroderma-associated pulmonary hypertension may also exert antiangiogenic effects.\(^\text{75}\) Angiostatin, a fragment of plasminogen; endostatin, a fragment of type XIII collagen; and their derivatives block α5β3 integrin-dependent angiogenesis.\(^\text{219,222}\) Angiostatin, endostatin, and kallistatin gave promising results in cancer therapy trials and preclinical arthritis studies.\(^\text{76,219,222-225}\) Type IV collagen derivatives including arresten, canstatin, and tumstatin also inhibit neovascularization.\(^\text{226}\) 2-Methoxyestradiol (2-ME) is a natural metabolite of estrogen with low affinity for estrogen receptors. 2-ME inhibits angiogenesis by disrupting microtubules and by suppressing HIF-1α activity.\(^\text{227}\) In recent preclinical studies, 2-ME suppressed arthritis in animal models.\(^\text{226,229}\) The microtubule destabilizer paclitaxel (Taxol), used in cancer therapy, inhibits HIF-1α expression and activity and thus indirectly blocks angiogenesis. Taxol has been found effective and safe in a phase I RA clinical trial.\(^\text{206}\) Thrombospondin-1 (TSP-1) is a proinflammatory but angiostatic ECM component that binds integrins.\(^\text{230,231}\) In one study, a TSP-1-derived peptide suppressed synovial inflammation and angiogenesis in a rat arthritis model.\(^\text{116}\) The role of sphingosine-1-phosphate (S1P) in angiogenesis is discussed earlier.\(^\text{202}\) Chemical lead 2 (CL2, Edg-1) is a non-S1P analogue. CL2 inhibited tube formation in endothelial cultures, suppressed VEGF-induced angiogenesis, and attenuated arthritis in the CIA model.\(^\text{232}\)

Recently, an emerging number of compounds in traditional Chinese and Korean medicine have been implicated in angiogenesis research. These compounds may also have angiostatic effects in arthritis. For example, scopolin, a coumarin derivative found in the stems of *Erycibe obtusifolia* Benth was injected intraperitoneally in the rat AIA model. Scopolin reduced paw swelling and arthritis scores and reversed body weight loss in the rats. This antiarthritic effect was accompanied by the suppression of synovial tissue VEGF, bFGF expression, and reduced angiogenesis.\(^\text{233}\) Celastrol, an active ingredient of *Tripterygium wilfordii,* also known as “Thunder God Vine,” has been widely used to treat RA in China. In a recent study, celastrol exerted antiangiogenic effects in various angiogenesis assays. Celastrol also inhibited endothelial cell migration and VEGFR1 and VEGFR2 expression.\(^\text{234}\) The flavonol-rich fraction of *Rhus verniciflua* Stokes (RVHxR) and its active ingredient, fisetin, inhibited the proliferation and cytokine production of RA synovial fibroblasts, as well as VEGF release. RVHxR and fisetin acted via the inhibition of ERK and JNK phosphorylation.\(^\text{235}\)

The restoration of impaired vasculogenesis is also crucial in arthritis and scleroderma patients.\(^\text{15,16,18,105}\) As discussed
earlier, corticosteroids and anti-TNF agents may stimulate EPCs and thus normalize vasculogenesis in RA.\textsuperscript{101,102} The CXCL12 (SDF-1) inhibitor bicyclam (AMD3100) mobilized EPCs in mice.\textsuperscript{103} As the number of EPCs correlate with the improvement of DAS28 in RA,\textsuperscript{103} the suppression of systemic inflammation and disease activity by any means would improve vasculogenesis.

**SUMMARY**

In this chapter we have discussed the putative role of leukocyte-endothelial adhesion, CAMs, chemokines, chemokine receptors, angiogenesis, and vasculogenesis in leukocyte recruitment underlying inflammation. The presence of various CAM pairs and interacting chemokines may account for the diversity and specificity of leukocyte recruitment. A number of soluble and cell-bound factors may stimulate or inhibit angiogenesis. The outcome of inflammatory and other “angiogenic diseases” such as various forms of arthritis depends on the imbalance between angiogenic and angiostatic mediators. Some CAMs and chemokines, as well as growth factors, proteases, antibiotics, and other agents are also involved in neovascularization. Impaired EPC function and vasculogenesis have been associated with active arthritis. There have been several attempts to therapeutically interfere with the cellular and molecular mechanisms described earlier. Specific targeting of leukocyte adhesion, CAMs, chemokines, chemokine receptors, and/or angiogenesis, as well as the restoration of defective vasculogenesis, may be useful for the future management of arthritis and other inflammatory diseases.

**Selected References**


Full references for this chapter can be found on www.expertconsult.com.
References


133. Halloran MM, Carley WW, Polverini PJ, et al: Ley/H: an endothelial-


201. Hirota SA, Beck PL, MacDonald JA: Targeting hypoxia-inducible factor-1 (HIF-1) signaling in therapeutics: implications for the treatment of inflammatory bowel disease, Recent Pat Inflamm Allergy Drug Discov 3(1):1–16, 2009.


Cytokines

Cytokines are peptides that have a fundamental role in communication within the immune system and in allowing the immune system and host tissue cells to exchange information.

Cytokines act via binding to a receptor that in turn sends a signal to the recipient cell that leads to a change in function or phenotype. Such signal cascades are complex and integrate a variety of environmental factors.

Cytokines exist in broad families that are structurally related but exhibit diverse function (e.g., tumor necrosis factor [TNF]/TNF receptor superfamily, IL-1 superfamily, IL-6 superfamily).

Cytokine targeting has proven effective in many rheumatic diseases, particularly therapeutics that inhibit TNF and IL-6—many more cytokines are currently under investigation as therapeutic targets or as therapeutic agents.

Immune function depends on the biologic activities of numerous small glycoprotein messengers termed cytokines. Originally discovered and defined on the basis of their functional activities, cytokines are now designated primarily by structure. Typically, cytokines exhibit broad functional activities that mediate not only effector and regulatory immune function but also in wider effects across a range of tissues and biologic systems. As such, cytokines play a role in not only host defense but also a variety of normal physiologic and metabolic processes. By this means they integrate de facto host defense and host metabolic function. The Human Genome Project has assisted the discovery of numerous cytokines, posing considerable challenges resolving their respective and synergistic functions in complex tissues in health and disease. Such understanding is, however, essential with the increasing application of cytokine-targeted therapies in the clinic. This chapter reviews general features of cytokine biology and the cellular and molecular networks within which cytokines operate; the focus is on the effector functions of cytokines that are important in chronic inflammation and in rheumatic diseases.

CLASSIFICATION OF CYTOKINES

In the absence of a unified classification system, cytokines are variously identified by numeric order of discovery (currently interleukin [IL]-1 through IL-37); by a given functional activity (e.g., tumor necrosis factor [TNF], granulocyte colony-stimulating factor); by kinetic or functional role in inflammatory responses (early or late, innate or adaptive, proinflammatory or anti-inflammatory) by primary cell of origin (monokine = monocyte derivation; lymphokine = lymphocyte derivation); and, more recently, by structural homologies shared with related molecules. Superfamilies of cytokines share sequence similarity and exhibit homology and some promiscuity in their reciprocal receptor systems (Figure 26-1). They do not exhibit functional similarity. Cytokine superfamilies also contain important regulatory cell membrane receptor-ligand pairs, reflecting evolutionary pressures that use common structural motifs in diverse immune functions in higher mammals. The TNF/TNF receptor superfamily contains immunoregulatory cytokines including TNF, lymphotoxins, and cellular ligands such as CD40L, which mediates B cell and T cell activation, and FasL (CD95), which promotes apoptosis. Similarly, the IL-1/IL-1 receptor superfamily contains cytokines including IL-1β, IL-1α, IL-receptor antagonist, IL-18, IL-33, IL-36 (α,β,γ), IL-36 receptor antagonist, and IL-37 (α,β,γ), which mediate physiologic and host-defense function, but this family also includes the Toll-like receptors, a series of mammalian pattern-recognition molecules with a crucial role in recognition of microbial species early in innate responses.

ASSESSING CYTOKINE FUNCTION IN VITRO AND IN VIVO

Although originally identified by bioactivity and quantified by bioassay, most cytokines are now identified via homologous receptor binding or sequence homology in gene databases. They are quantified in biologic solutions by enzyme-linked immunosorbent assay, multiplex technology, or meso platform techniques, the latter allowing many (25 to 360) cytokines to be measured in single, small sample volumes (~20 µL). Function is thereafter assessed by identification of the cellular source of cytokine, determination of native stimuli, characterization of receptor distribution, and determination of function in target cells. Experimental in vivo models use the addition of neutralizing cytokine-specific antibodies or soluble receptors (often as fragment crystallizable fusion or pegylated proteins to enhance half-life and modulate functional interaction with leukocytes) to modulate cytokine function. Genetically modified knockout and knockin mice (cytokine or receptor modified by embryonic stem cell technology) or transgenic mice (tissue/cell lineage–specific overexpression) have proven particularly useful. Conditional gene-targeting approaches (e.g., using the Cre system) facilitate circumvention of embryonic lethal deficiencies or allow kinetic evaluation of the relative contribution of a cytokine throughout a response. Moreover, recent multiphoton microscopic techniques have
allowed the additional evaluation of cytokine contributions in three-dimensional tissue orientation and in real time in vivo. Cytokine function is normally assessed in vitro in primary or transformed cell lines stimulated in the presence or absence of recombinant cytokine or specific anticytokine antibody or soluble receptor. Gene knock-down approaches using siRNA or antisense oligonucleotides are also increasingly employed.

This general approach has been crucial in rheumatic disease research. Studies in which cytokine addition and neutralization occur in synovial tissue explants or disaggregated cell populations, chondrocyte explants, bone culture models, skin, and renal tissue explants and cell lines have been informative. Ex vivo methodologies now include intracellular fluorescence activated cell sorter methods, confocal and laser scanning microscopy, and quantitative histologic evaluation using automated image analysis. Such modalities, particularly when employed in human therapeutic cytokine neutralization studies in which inflammatory tissues are obtained throughout therapeutic interventions, advance the understanding of basic and pathogenetic cytokine function. Analysis of synovial biopsy specimens obtained before and after infliximab, adalimumab, abatacept, rituximab, IL-1Ra, IL-10, and interferon (IFN)-β administration in rheumatoid arthritis provides the strongest evidence for the success of this approach.  

**CYTOKINE RECEPTORS**

Cytokine receptors exist in structurally related superfamilies and comprise high-affinity molecular signaling complexes that assist cytokine-mediated communication (see Figure 26-1). Such complexes often include heterodimeric or heterotrimeric structures that use unique, cytokine-specific recognition receptors together with common receptor chains shared across a cytokine superfamily. Examples include the use of the common γ chain receptor by IL-2, IL-4, IL-7, IL-9, IL-15, IL-21, and glycoprotein 130 (gp130) by members of the IL-6 family. Alternatively, distinct receptors may use shared signaling domains. Homologous death domains are found in many TNF-receptor family members. Similarly, the IL-1 signaling domain is common to not only IL-1R but also other IL-1R superfamily members including IL-18R, IL-33R, and the Toll-like receptors. Signaling pathways dependent on these are discussed in detail elsewhere. It has been recognized more recently that unrelated cytokine receptor systems exhibit close cross-communication on the cell membrane, allowing a cell to integrate a variety of external stimuli to optimize signaling pathways and the cellular response in real time in a changing environment. Although best elucidated in the epidermal growth factor receptor system, this also has been identified for members of the common γ chain signaling family.

Cytokine receptors can operate via several mechanisms. Membrane receptors, with intracellular signaling domains intact, can transmit signals to the target cell nucleus after soluble cytokine binding and promote effector function (Figure 26-2). Membrane receptors may bind cell membrane cytokines assisting cross-talk between adjacent cells. Membrane-bound and soluble cytokines may promote distinct receptor function. Useful exemplars exist relevant to the rheumatic diseases. Thus TNF binds TNF-R1 and TNF-R2 with similar affinity, but it has a slower rate of dissociation fromTNF-R1. Soluble TNF may dissociate rapidly from TNF-R1 to bind TNF-R1, promoting preferential signaling by the latter (ligand passing) . In contrast, during cell-cell contact, stable TNF/TNF-R1 and TNF/TNF-R2 complexes form, allowing for differential signaling contribution by TNF-R1 and TNF-R2.
complex are derived from adjacent cells. IL-15/IL-15Rα complexed on one cell may bind IL-15Rβ/γ on another. Receptors also exist in soluble form, derived either from alternative mRNA processing to generate receptor-lacking transmembrane or intracellular domains or from enzymatic cleavage of receptor from the cell surface (e.g., sTNF-R, sIL-1R1). Soluble receptors may act to antagonize cytokine function, regulating responses. Soluble receptors also may preform complexes with cytokine to promote subsequent ligand-receptor assembly on the target cell membrane and enhance function. Soluble receptors can deliver cytokine to the cell membrane via ligand passing. IL-6 provides a particularly important example given its core role in a range of rheumatic disorders. IL-6 binds to a heterodimeric receptor (IL-6R and gp130) and provokes cell activation thereby via conventional signal pathways that involve STAT3. Thus IL-6 may activate a cell expressing the combination of IL-6R and gp130 by conventional (cis) signaling. In addition, however, circulating soluble IL-6R may form functional gp130/IL-6R effector complexes on any cell expressing membrane gp130 and by this means confer on circulating IL-6 the ability to exert broad functional effects (trans signaling). Finally, it is now recognized that some cytokines with the capacity to be retained in the membrane may themselves function as signaling molecules (reverse signaling).

The precise intracellular signal pathways whereby cytokines mediate their effects have now been elucidated and reveal a multiplicity of potential therapeutic targets (Figure 26-3). This is exemplified in the role of the Janus kinases (JAK) and their downstream transcriptional factors (e.g., STATs) in mediating signals serving those cytokines that bind to common gamma chain, or to gp130 receptor. Similarly the mitogen-activated protein (MAP) kinases mediate signals that integrate cell responses to stress, cytokines, and proliferation factors. Spleen tyrosine kinase (Syk) is an additional kinase that mediates the effector function of immunoglobulin Fc receptor and the B cell receptor and interacts with many downstream cytokine-mediated signal pathways. Finally, the NFκB pathway is central to the function of the TNF receptor and to those pathways that mediate signals of Toll-like receptors and IL-1 superfamily members. There is considerable interest in targeting upstream members of the MAP kinase family, JAKs, and Syk, with the latter two demonstrating efficacy in clinical trials focused on rheumatoid arthritis. One key question will be whether the broader cytokine-regulating effects of signal transduction inhibitors will have an acceptable safety profile compared with biologics that target individual cytokines or receptors.

**REGULATION OF CYTOKINE EXPRESSION**

Cytokines are synthesized in the Golgi apparatus and may traffic through the endoplasmic reticulum to be released as soluble mediators, or they may remain membrane bound, or
PART 3 | EFFECTOR MECHANISMS IN AUTOIMMUNITY AND INFLAMMATION

Figure 26-3  Cytokine signaling pathways. Numerous signaling mechanisms participate in transmitting information from the cell surface to the cytoplasm or the nucleus. Several representative examples are shown. In each case, a cytoplasmic amplification system permits modulating and integrating environmental stresses to orchestrate the appropriate cellular response. ERK, extracellular receptor kinase; IKK, inhibitor of κB kinase; JAK, Janus kinase; JNK, Jun N-terminal kinase; MAPK, mitogen-activated protein kinase; NFκB, nuclear factor κB; STAT, signal transducer and activator of transcription; SYK, spleen tyrosine kinase.

they may be processed into cytosolic forms that can traffic intracellularly, even returning to the nucleus, where they can act as transcriptional regulators. Cytokines mediate autocrine function either through release or membrane expression and immediate receptor ligation on the source cell or intracellularly within the source cell. Alternatively, cytokines operate in a paracrine manner, allowing cellular communication beyond that assisted by local cell-cell contact. The distance and kinetics for effective function may be limited; however, by numerous factors including physicochemical considerations of the peptide structure itself, extracellular matrix binding (e.g., to heparan sulfate), enzymatic degradation (e.g., serine protease degradation of IL-18), or the presence of soluble receptors (e.g., TNF/soluble TNF-RI and TNF-RII, IL-2/soluble IL-2RIα) or novel cytokine-binding proteins (e.g., IL-18/IL-18 binding protein) in the inflammatory milieu.

Numerous factors promote cytokine expression in vivo (Figure 26-4) including cell-cell contact, immune complexes/autoantibodies, local complement activation, microbial species and their soluble products (particularly via TLR and nucleotide oligomerization domain [NOD]-like receptors [NLRs]), reactive oxygen and nitrogen intermediates, trauma, shear stress, ischemia, radiation, ultraviolet light, extracellular matrix components, DNA (mammalian or microbial), heat shock proteins, electrolytes (e.g., K+ via P2X7 receptors), and cytokines themselves operating in autocrine loops. Commonly used in vitro stimuli include many of these and chemical entities including phorbol esters, calcium ionophores, lectins (e.g., phytohemagglutinin), and receptor-specific antibodies such as anti-CD3 and anti-CD28 for T cell activation or anti-immunoglobulin and anti-CD40 for B cells.

Cytokine regulation within the cell can be usefully considered at several levels (see Figure 26-2). Transcriptional regulation depends on the recruitment of discrete transcription factors to the cytokine promoter region. Transcription factor binding allows for numerous signal pathways to regulate cytokine expression across a range of stimuli. Several transcription factors (e.g., nuclear factor κB [NFκB], activator protein-1 [AP-1], nuclear factor of activated T cell) are

Figure 26-4  Overview of cytokine regulatory function. Numerous and diverse stimuli (1) promote cytokine expression arising either from novel gene expression (2) or from activation of preformed cytokine (3). Cytokine proteins are thereafter expressed in the cytosol, on the cell membrane, or in soluble form in the extracellular environment (4). Cytokines bind to reciprocal receptors that reside either on the membrane of a target cell or in the soluble phase (5). Membrane receptors, on cytokine ligation, signal to the recipient cell nucleus (6) and drive novel gene expression to promote effector function. Each phase of cytokine function offers rich therapeutic potential. IC, immune complexes.
crucial in cytokine production. Inhibition of NFκB activity using either chemical inhibitors or adenoviral delivery of regulatory peptides leads to amelioration of inflammatory synovitis in vivo and in vitro. Sequence polymorphism within cytokine promoters offers potential for differential cytokine expression between individuals that could confer selective advantage against infection but could also increase susceptibility to, or progression of, autoimmunity. This is best exemplified in the TNF and IL-1 promoters. Single nucleotide polymorphisms in the TNF promoter region (e.g., −308) are associated with altered TNF release on leukocyte stimulation in vitro. Similarly, homozygotes for the A2 allele at +3954 in the IL-1β gene produce more IL-1β with lipopolysaccharide stimulation. Polymorphisms also exist in the IL-1Ra gene, rendering the functional significance of individual single nucleotide polymorphisms on IL-1 protein release difficult to interpret. In general, the net effect of haplotypes may be more important at the functional level or only play a role in the context of networks where multiple minor polymorphisms can synergize, particularly when their relevance to disease entities is considered.

Posttranscriptional regulation is important in determining longevity of cytokine expression. This regulation may operate by promoting translational initiation, mRNA stability, and polyadenylation. AU-rich elements (AREs) within the 5′ or 3′ untranslated regions (UTRs) of cytokine mRNA are crucial for stability; 3′ UTR AREs down-regulate TNF expression such that transgenic knockin mice that lack TNF AREs develop spontaneous inflammatory arthritis and bowel disease. Regulatory proteins bind AREs to mediate such effects. HuR and AUFI exert opposing effects, stabilizing and destabilizing ARE-containing transcripts. TIA-1 and TIAR have been identified as RNA recognition motif family members that function as translational silencers. Macrophages from TIA-1-deficient macrophages produce excess TNF, whereas TIA-1-deficient lymphocytes exhibit normal TNF release, suggesting distinctions in mRNA regulation in discrete cell types. Alternatively, cytokines may generate stable mRNA a priori to assist subsequent rapid response in tissues. IL-15 mRNA 52 UTR contains 12 AUG triplets that significantly reduce the efficiency of IL-15 translation. Deletion of this sequence permits IL-15 secretion. IL-15 mRNA can produce a 48 amino acid signal peptide that allows IL-15 release and a shorter 21 amino acid signal peptide that targets intracellular distribution. IL-15 forms thus generated exhibit discrete functions.

Posttranslational regulation also modulates cytokine expression via several mechanisms. Patterns of glycosylation are important for cytokine function and may regulate intracellular trafficking. Modified leader sequences can alter intracellular trafficking of cytokines. Some cytokines are translated without functional leader sequences. Their secretion depends on nonconventional secretory pathways that are poorly understood. IL-1β employs, among other pathways, a purine receptor–dependent pathway (PXR) for cellular release. Enzymatic activation of cytokines is common, whereby nonfunctional promolecules are cleaved to generate functional subunits. Examples include the cleavage by caspase 1 of pro-IL-1β to generate active IL-1β and, similarly, of pro-IL-18 to generate an active 18-kD species. This is an organized process both sequentially (in time) and by orientation within the cell (in space). IL-1 processing occurs in a protein complex within the cytosol termed the inflammasome. The latter has recently attracted considerable interest as a therapeutic target in conditions such as crystal-induced arthritis and in diseases that arise from mutations in certain inflammasome genes (e.g., cryopyrin) (see Chapters 18, 94, and 97).

Alternative processing pathways for cytokines include the serine proteases, proteinase 3 and elastase, and adamaspin family members. Enzyme cleavage pathways operate within and outside cells, providing for extracellular cytokine activation. Similarly, cell membrane enzymes serve to cleave membrane-expressed cytokine. Members of the adamaspin family regulate TNF release; TNF-converting enzyme cleaves and mediates the release of TNF and its receptors. Extensive molecular machinery exists to regulate tightly not only the production and stability of cytokine mRNA but also its translation and cellular expression and distribution. At each level, opportunities exist for intervention and therapeutic cytokine modulation.

**EFFECTOR FUNCTION OF CYTOKINES**

Cytokines possess pleiotropic and potent effector function in acute and chronic inflammatory responses. The identity, receptor specificity, and key effects of cytokines understood to have particular importance in pathogenesis of human autoimmunity and chronic inflammation are summarized in Tables 26-1 through 26-8.

**Cytokines in Acute Inflammation**

Cytokines operate at every stage in the crucial early events that promote acute inflammation. Cells that make up the innate immune response including neutrophils, natural killer cells, macrophages, mast cells, and eosinophils all produce and respond to cytokines generated within seconds of tissue insult. Cytokines prime leukocytes for response to microbial and chemical stimuli, upregulate adhesion molecule expression on migrating leukocytes and endothelial cells, and amplify the release of reactive oxygen intermediates, nitric oxide, vasoactive amines, and neuropeptides, as well as the activation of kinins and arachidonic acid derivatives, prostaglandins, and leukotrienes, which regulate cytokine release. Similarly, cytokines regulate the expression of complement processing and membrane defense molecules, scavenger receptors, NLRs, and TLRs. Cytokines, particularly IL-1, TNF, and IL-6, are crucial in driving the acute-phase response. Tables 26-1 through 26-8 provide descriptions of the function of cytokines expressed within the acute inflammatory response.

**Cytokines in Chronic Inflammation**

Cytokines critically modulate the cellular interactions that characterize chronic inflammation. Studies using real-time image analytic techniques such as two-photon microscopy and confocal scanning suggest continuous cellular motility during inflammation. Inflammatory lesions might properly be considered fluid states in which individual cells under
### Table 26-1 Interleukin-1 Superfamily Cytokines with Roles in Rheumatic Disease

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>Size (kD)*</th>
<th>Receptors</th>
<th>Major Cell Sources</th>
<th>Key Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL-1β</td>
<td>35 (pro)</td>
<td>IL-1RI</td>
<td>Monocytes; B cells; fibroblasts; chondrocytes; keratinocytes</td>
<td>Fibroblast cytokine, chemokine, MMP, iNOS, PG release ↑</td>
</tr>
<tr>
<td></td>
<td>17 (active)</td>
<td>IL-1RAcP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IL-1RII (decoy)</td>
<td></td>
<td>Monocyte cytokine, ROI, PG ↑</td>
</tr>
<tr>
<td>IL-1α</td>
<td>35 (pro)†</td>
<td>IL-1RI</td>
<td>Monocytes; B cells; PMNs; epithelial cells; keratinocytes</td>
<td>similar to IL-1β</td>
</tr>
<tr>
<td></td>
<td>17 (active)</td>
<td>IL-1RAcP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IL-1RII (decoy)</td>
<td></td>
<td>Osteoclast activation</td>
</tr>
<tr>
<td>IL-1Ra</td>
<td>22</td>
<td>IL-1RI</td>
<td>Monocytes</td>
<td>Antagonize effects of IL-1β and IL-1α</td>
</tr>
<tr>
<td>IL-18</td>
<td>23 (pro)</td>
<td>IL-18R</td>
<td>Monocytes; PMNs; dendritic cells; platelets; endothelial cells</td>
<td>T cell effector polarization (Th1 with IL-12/Th2 with IL-4)</td>
</tr>
<tr>
<td></td>
<td>18 (active)</td>
<td>IL-18Rβα</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL-33</td>
<td>30 (pro)</td>
<td>ST2L</td>
<td>Epithelial cells; monocytes; smooth muscle cells; keratinocytes</td>
<td>Promote Th2 cell activation, mast cell activation, and cytokine production</td>
</tr>
<tr>
<td></td>
<td>18 (active)</td>
<td>IL-1RAcP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Pro forms cleaved to active moieties by proteases including caspase-1, calpain, elastase, and cathepsin G.†Pro-IL-1α retains bioactivity before cleavage.

GAG, glycosaminoglycan; iNOS, inducible nitric oxide synthase; MMP, matrix metalloproteinase; NK, natural killer; PG, peptidoglycan; PMN, polymorphonuclear neutrophil; ROI, reactive oxygen intermediates.

### Table 26-2 Tumor Necrosis Factor Superfamily Cytokines* with Potential Role in Rheumatic Disease

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>Size (kD)</th>
<th>Receptors</th>
<th>Major Cell Sources</th>
<th>Selected Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNF</td>
<td>26 (pro)</td>
<td>TNF-RI (p55)</td>
<td>Monocytes; T, B, NK cells; PMNs; eosinophils; mast cells; fibroblasts; keratinocytes; glial cells; osteoblasts; smooth muscle</td>
<td>Monocyte activation, cytokine, and PG ↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TNF-RII (p75)</td>
<td></td>
<td>PMN priming, apoptosis, oxidative burst ↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Endothelial cell adhesion molecule, cytokine release ↑; fibroblast proliferation and collagen synthesis ↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MMP and cytokine ↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T cell apoptosis; clonal (auto)regulation; TCR dysfunction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adipocyte FFA release ↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Endocrine effects—ACTH, prolactin ↑; TSH, FSH, GH ↓</td>
</tr>
<tr>
<td>LTα</td>
<td>22-26</td>
<td>TNF-RI</td>
<td>T cells; monocytes; fibroblasts; astrocytes; myeloma; endothelial cells; epithelial cells</td>
<td>Peripheral lymphoid development</td>
</tr>
<tr>
<td>RANK ligand</td>
<td>35</td>
<td>RANK</td>
<td>Stromal cells; osteoblasts; T cells</td>
<td>Otherwise similar bioactivities to TNF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TNF-RII</td>
<td></td>
<td>Stimulates bone resorption via osteoclast maturation and activation</td>
</tr>
<tr>
<td>OPG</td>
<td>55</td>
<td>RANKL TACI BCMA</td>
<td>Stromal cells, osteoblasts Monocytes; T cells; DCs</td>
<td>Modulation of T cell-DC interaction</td>
</tr>
<tr>
<td>BlyS†</td>
<td>18-32</td>
<td>TACI BCMA</td>
<td></td>
<td>Soluble decoy receptor for RANKL</td>
</tr>
<tr>
<td>APRIL</td>
<td>—</td>
<td>TACI BCMA</td>
<td>Monocytes; T cells; tumor cells</td>
<td>T cell co-stimulation</td>
</tr>
</tbody>
</table>

*Additional members of importance include TRAIL, TWEAK, CD70, FasL, and CD40L. At least 18 members of the family are now described.†Also called BAFF.

ACTH, adrenocorticotropic hormone; APRIL, a proliferation inducing ligand; BAFF, B cell activating factor belonging to the TNF family; BCMA, B cell maturation protein; BlyS, B lymphocyte stimulator protein; DC, dendritic cell; FFA, free fatty acid; GAG, glycosaminoglycan; LT, lymphoxygenin; MMP, matrix metalloproteinase; NK, natural killer; OPG, osteoprotegerin; PG, peptidoglycan; PMN, polymorphonuclear neutrophil; RANK, receptor activator of NFκB ligand; TACI, transmembrane activator and calcium modulator and cyclophilin ligand; TNF, tumor necrosis factor; TRAIL, TNF-related apoptosis-inducing ligand; TWEAK, TNF-like weak inducer of apoptosis.
### Table 26-3  Cytokines Associated Predominantly with Effector Function for T Cells*

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>Size (kD)</th>
<th>Receptors</th>
<th>Major Cell Sources</th>
<th>Key Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type II Interferon</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| IFN-γ         | 20-25     | IFNγR     | Th/c1 cells; NK cells; γδT cells; B cells; macrophage/DCs                         | Macrophage activation, DC APC function $\uparrow$
|               |           |           |                                                                                    | Endothelial adhesion molecule $\uparrow$
|               |           |           |                                                                                    | MHC class II expression $\uparrow$
|               |           |           |                                                                                    | T cell growth $\downarrow$; opposes Th2 responses
|               |           |           |                                                                                    | Bone resorption $\downarrow$; fibroblast collagen synthesis |
| **4α-Helix Family** |           |           |                                                                                   |                                                                                |
| IL-2          | 15        | IL-2Rα    | Th/c cells; NK cells                                                               | T cell division; maturation; cytokine release; cytotoxicity
|               |           | IL-2/15Rβ $\gamma$-chain |                                                                                   | NK cell cytokine release; cytokotoxicity; monocyte activation
| IL-4          | 20        | IL-4Rα/$\gamma$-chain IL-4Rα/IL-13R1 | Th/c cells (Th2); NK cells                                               | Lymphocyte apoptosis $\downarrow$
|               |           |           |                                                                                    | Th2 differentiation, maturation, apoptosis $\downarrow$
|               |           |           |                                                                                    | B cell maturation; isotype switch (IgE)
|               |           |           |                                                                                    | Eosinophil migration, apoptosis $\downarrow$
|               |           |           |                                                                                    | Endothelial activation; adhesion molecule expression
| IL-5          | 25 monomer | IL-5Rα    | Th/c2 cells; NK cells; mast cells; epithelial cells                               | B cell differentiation; immunoglobulin production (IgA)
|               | 50 homodimer | IL-5Rβ |                                                                                   | Eosinophil differentiation and activation |
| **IL-17 Family†** |           |           |                                                                                   |                                                                                |
| IL-17A/F      | 20-30     | IL-17R    | T cells (Th17); fibroblasts                                                          | Chemokine release, fibroblast cytokine release, MMP release $\uparrow$
|               |           |           |                                                                                    | Osteoclastogenesis; hematopoiesis
|               |           |           |                                                                                    | Chondrocyte GAG synthesis $\downarrow$
|               |           |           |                                                                                    | Leukocyte cytokine production $\uparrow$
| IL-25 (IL-17E)| 20-30     | IL-17R    | Th2 cells                                                                          | Th2 cytokine release; B cell IgA and IgE synthesis; eosinophilia; epithelial cell hyperplasia |

*Additional T cell–derived cytokines of potential interest include IL-13 from Th2 and NK2 cells.
IL-17 family also contains IL-17B and IL-17C, the distinct functions of which are currently unclear.
APC, antigen presenting cell; DC, dendritic cell; GAG, glycosaminoglycan; IFN, interferon; MHC, major histocompatibility complex; MMP, matrix metalloproteinase; NK, natural killer; Th/c, T helper/cytotoxic.

### Table 26-4  Cytokines Described Initially with Primary Role in Regulation of T Cells*

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>Size</th>
<th>Receptors</th>
<th>Major Cell Sources</th>
<th>Key Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL-12</td>
<td>IL-12/23p40 IL-12p35</td>
<td>IL-12Rα IL-12Rβ1 IL-12Rβ2 IL-15Rα IL-2/15Rβ $\gamma$-chain</td>
<td>Macrophages; DCs</td>
<td>Th1 cell proliferation, maturation T cell cytotoxicity B cell activation T cell chemokinesis, activation, memory maintenance NK cell maturation, activation, cytotoxicity Macrophage activation, suppression (dose dependent) PMN activation, adhesion molecule, oxidative burst Fibroblast activation B cell differentiation and isotype switching B cell activation Th17 cell expansion and activation; IL-17 release</td>
</tr>
<tr>
<td>IL-15</td>
<td>15 kD</td>
<td>IL-15Rα</td>
<td>Monocytes; fibroblast; mast cells; B cells; PMNs; DCs</td>
<td></td>
</tr>
<tr>
<td>IL-21</td>
<td>15 kD</td>
<td>IL-21R $\gamma$-chain IL-23R</td>
<td>Activated T cells; others (?)</td>
<td>Macrophages; DCs</td>
</tr>
<tr>
<td>IL-23</td>
<td>IL-12/23p40 IL-23p19</td>
<td>IL-12Rβ1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 26-5  IL-10 Superfamily Cytokines*

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>Receptors</th>
<th>Cellular Sources</th>
<th>Key Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL-10</td>
<td>IL-10R1</td>
<td>Monocytes; T cells; B cells; DCs; epithelial cells; keratinocytes</td>
<td>Macrophage cytokine release, iNOS, ROI ↓; soluble receptor ↑</td>
</tr>
<tr>
<td></td>
<td>IL-10R2</td>
<td></td>
<td>T cell cytokine release, MHC expression ↓; anergy induction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Treg cell maturation; effector function (?)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DC activation, cytokine release ↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fibroblast MMP, collagen release ↓; no effect on TIMP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B cell isotype switching enhanced</td>
</tr>
<tr>
<td>IL-19</td>
<td>IL-20R1/IL-20R2</td>
<td>Monocytes; others (?)</td>
<td>Monocyte cytokine and ROI release; monocyte apoptosis</td>
</tr>
<tr>
<td>IL-20</td>
<td>IL-22R/IL-20R2</td>
<td>Keratinocytes; others (?)</td>
<td>Autocrine keratinocyte growth regulation</td>
</tr>
<tr>
<td>IL-22</td>
<td>IL-22R/IL-10R2</td>
<td>Th17 cells; CD8 T cells; γδ T cells; NK cells</td>
<td>Acute-phase response, keratinocyte activation proliferation ↑</td>
</tr>
<tr>
<td>IL-24</td>
<td>IL-22R/IL-20R2</td>
<td>Monocytes; T cells</td>
<td>Tumor apoptosis; Th1 cytokine release by PBMC</td>
</tr>
<tr>
<td></td>
<td>IL-20R1/IL-20R2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Additional members include IL-26, IL-28, and IL-28A. Many functions of IL-10 superfamily are as yet poorly understood, but they likely reside beyond the immune system.

DC, dendritic cell; iNOS, inducible nitric oxide synthase; MMP, matrix metalloproteinase; NK, natural killer; PBMC, peripheral blood mononuclear cells; ROI, reactive oxygen intermediates; TIMP, tissue inhibitor of metalloproteinase.

### Table 26-6  IL-6 Superfamily Cytokines*

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>Size (kD)</th>
<th>Receptors</th>
<th>Major Cell Sources</th>
<th>Key Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL-6</td>
<td>21-28</td>
<td>IL-6R† gp130</td>
<td>Monocytes; fibroblasts; B cells; T cells</td>
<td>B cell proliferation; immunoglobulin production</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hematopoiesis, thrombopoiesis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T cell proliferation, differentiation, cytotoxicity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hepatic acute-phase response</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hypothalamic-pituitary-adrenal axis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Variable effects on cytokine release by monocytes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Megakaryocyte differentiation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fibroblast, TIMP, and cytokine release</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acute-phase reactants, fibroblast protease inhibitors ↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Monocyte TNF release ↓; IL-1 effector function ↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hypothalamic-pituitary axis ↑; corticosteroid release</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Modulatory effect on osteoblast (?)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Proinflammatory effects in some models (?)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acute-phase reactants ↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hematopoiesis, thrombopoiesis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Role in neural development, neural effector function, implantation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bone metabolism; extracellular matrix regulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Leukocyte adhesion molecule expression</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eosinophil priming</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mixed proinflammatory versus anti-inflammatory effects in models</td>
</tr>
</tbody>
</table>

*Additional members of potential importance include IL-11, cardiotorpin-1, and ciliary neurotrophic factor. Note overlapping effects within family.

†Membrane or soluble form can dimerize gp130 to promote signaling, which promotes signal transduction.

LIFR, leukemia inhibitory factor receptor; OMR, oncostatin M receptor; TIMP, tissue inhibitor of metalloproteinase; TNF, tumor necrosis factor.
### Table 26-7  Growth Factors Relevant to Rheumatic Diseases

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>Receptors</th>
<th>Cellular Sources</th>
<th>Key Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGF-β*</td>
<td>Type I TGFβR</td>
<td>Broad—including fibroblasts, monocytes, T cells, platelets</td>
<td>Wound repair, matrix maintenance, and fibrosis</td>
</tr>
<tr>
<td>Isoforms 1-3†</td>
<td>Type II TGFβR</td>
<td>Others</td>
<td>Initial activation then suppression of inflammatory responses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— T cell (Treg and Th17) and NK cell proliferation and effector function ↓</td>
<td>Early-phase leukocyte chemotaxant, gelatinase, and integrin expression ↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— Early macrophage activation then suppression, reduced iNOS expression</td>
<td></td>
</tr>
<tr>
<td>BMP family (BMP2-15)</td>
<td>BMPRI</td>
<td>Varied (e.g., epithelial and mesenchymal embryonic tissues); bone-derived cell lineages</td>
<td>Regulate critical chemotaxis, mitosis, and differentiation processes during chondrogenesis and osteogenesis, tissue morphogenesis (e.g., heart, skin, eye)</td>
</tr>
<tr>
<td>PDGF</td>
<td>PDGFRαβ</td>
<td>Platelets; macrophages; endothelial cells; fibroblasts; glial cells; astrocytes; myoblasts; smooth muscle cells</td>
<td>Local paracrine or autocrine growth factor for variety of lineages</td>
</tr>
<tr>
<td>FGF family</td>
<td>PDGFRβ</td>
<td>Widespread</td>
<td>Wound healing</td>
</tr>
</tbody>
</table>
|                   | FGFR (various) |                                    | Growth and differentiation of mesenchymal, epithelial, and neuroectodermal cells | *Members of TGF-β superfamily include BMP, growth and differentiation factor, inhibinA, inhibinB, müllerian inhibitory substance, glial-derived neurotrophic factor, and macrophage inhibitory cytokine.  
†Bound to latency-associated peptide to form small latency complex and to latent TGF-β binding protein to form large latent complex; activated by proteolytic and nonproteolytic pathways.  
BMP, bone morphogenetic protein; FGF, fibroblast growth factor; iNOS, inducible nitric oxide synthase; NK, natural killer; PDGF, platelet-derived growth factor; TGF, transforming growth factor.

### Table 26-8  Miscellaneous Cytokines with Potential Roles in Rheumatic Diseases

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>Size (kD)</th>
<th>Receptors</th>
<th>Cellular Sources</th>
<th>Key Functions</th>
</tr>
</thead>
</table>
| MIF               | 12        | Unclear     | Macrophages; activated T cells; fibroblasts (synoviocytes) | Macrophage cytokine release, phagocytosis, NO release ↑  
|                   |           |             |                                                       | T cell activation; DTH                                                       |
|                   |           |             |                                                       | Fibroblast proliferation; COX expression; PLA; expression Intrinsic oxidoeductase activity (“cytozyme”) |
| HMGB1             | 30        | RAGE, dsDNA | Widespread expression; necrotic cells; macrophages; pituicytes | DNA-binding transcription factor  
|                   |           | Others (?)  |                                                       | Necrosis-induced inflammation  
|                   |           |             |                                                       | Macrocytokine activation—delayed proinflammatory cytokine  
|                   |           |             |                                                       | Smooth muscle chemotaxis  
|                   |           |             |                                                       | Disrupts epithelial barrier function  
|                   |           |             |                                                       | Bacterial (direct)  
| GM-CSF            | 14-35     | GM-CSFRαβ  | T cells; macrophages; endothelial cells; fibroblasts   | Granulocyte and monocyte maturation; hemopoietic effects  
|                   |           |             |                                                       | Leukocyte PG release; DC maturation  
|                   |           |             |                                                       | Pulmonary surfactant turnover  
|                   |           |             |                                                       | Granulocyte maturation; promotes PMN function |
| G-CSF             | 19        | G-CSFR      | Monocytes; PMNs; endothelial cells; fibroblasts; various tumor cells; stromal cells |                                             |
| M-CSF             | 28-44     | M-CSFR      | Monocytes; fibroblasts; endothelial cells               | Monocyte activation, maturation                                                |
| IL-32α-δ          | Unknown   | Unknown     | Monocytes; T cells; NK cells; epithelial cells         | Promotes proinflammatory cytokine release from variety of cells               |
| Type I interferons | Various   | IFNαβR      | Widespread                                             | Antiviral response  
| IFNα/β family     |           |             |                                                       | Broad immunomodulatory effects (promotes MHC expression)  
|                   |           |             |                                                       | Macrocytokine activation; lymphocyte activation and survival  
|                   |           |             |                                                       | Antiproliferative, cytoskeletal alteration, differentiation ↑ |

COX, cyclooxygenase; DC, dendritic cell; DTH, delayed-type hypersensitivity; G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor; HMGB, high mobility group box chromosomal protein; IFN, interferon; M-CSF, macrophage colony-stimulating factor; MIF, macrophage inhibitory factor; NO, nitric oxide; PG, prostaglandin; PLA, phospholipase A; PMN, polymorphonuclear neutrophil; RAGE, receptor for advanced glycation end products.
cytokine control transiently contribute to organized functional subunits—such as the ectopic germinal center, synovial lining layer, or renal interstitial nephritis—but remain competent to migrate thereafter under the influence of chemotactic gradients on the extracellular matrix. Cytokines also may promote cell death (apoptosis) either by withdrawal (e.g., IL-2, IL-7, IL-15) or by binding cytokine receptors containing death domains (e.g., TNF-R1). Cytokines contribute at every stage of inflammatory lesion development in a dynamic equilibrium, rather than in a static, linear manner. Chronic inflammation in rheumatic disease usually contains cytokine activities reminiscent of innate and acquired immune responses. For convenience, cytokines can be considered by their effect on cell subsets and cellular interactions (Figure 26-2 describes a notional positioning for cytokine activity in a developing and chronic lesion). Investigation of cytokine-regulated pathways in several rheumatic diseases has identified numerous common pathways.

**T Cell Effector Function in Chronic Inflammation**

T cells depend on cytokine function at every developmental stage from bone marrow stem cell maturation, through thymic education, to functional determination and maturation after primary or secondary antigen exposure (see detailed overview in Chapter 13). The latter is of prime importance because re-education of phenotypic T cell responses may be achieved through alteration of the ambient cytokine milieu. T cell receptor–peptide–major histocompatibility complex (MHC) interactions during T cell–dendritic cell interaction rely on co-stimulatory molecule and local cytokine expression to determine functional outcome (see Tables 26-3 and 26-4). IL-12, in the presence of IL-18, promotes type 1 phenotypic development, characterized ultimately by IFN-γ-producing T helper type 1 (Th1) effector cells. IFN-γ drives macrophage priming and activation and adhesion molecule expression and promotes granuloma formation and microbial killing. IFN-γ has a complex role in tissue destruction, however, with contradictory data obtained in inflammation models in IFN-γ-deficient and IFN-γ receptor–deficient mice. IFN-γ may retard tissue destruction, perhaps by suppressing osteoclast activation.

A novel T cell subset has been defined that secretes IL-17A predominantly (Th17 effector cells), together with IL-22. Th17 cells are generated in the presence of IL-6 and transforming growth factor (TGF)-β, expanded by IL-1β and IL-23, and antagonized by IL-25 (IL-17E) and paradoxically by IFN-γ. IL-17A provides a direct and rapid route to tissue damage via means as osteoclast activation or FLS activation. The precise contribution of Th17 cells in human autoimmune disease is currently unclear. There are, however, persuasive data from rodent models indicating that Th17 cells may be of primary importance as initiator and effector cells. Clinical trials that target IL-17A are ongoing in various rheumatic diseases.

IL-4 dominance during T cell–dendritic cell interactions in the presence of IL-18 leads to type 2 responses, which promote humoral immunity driven by Th2 cells synthesizing primarily IL-4, IL-5, IL-10, and IL-13. Resulting pathogenesis is more likely to be B cell mediated. Cytokines that predispose to regulatory T cell development are unclear, although high levels of IL-10 or TGF-β have been suggested in this context. Effector T cells can operate via secretion of cytokines to patterns determined by their prior activatory conditions.

**Cell-to-Cell Interactions**

In many inflammatory lesions, there is relative paucity of inducer T cell–derived cytokines (especially IL-17A or IFN-γ), despite abundant proinflammatory cytokine expression. Cell-cell membrane interactions between leukocyte subsets and between tissue cells and leukocytes have emerged as a dominant mechanism sustaining chronic inflammation. Cytokines contribute to these interactions at several levels (see Figure 26-3) including directly as membrane-expressed ligands, indirectly via activating cells, and synergistically by enhancing their subsequent cognate activities. The importance of cytokine-cell contact interactions is best studied in synovial tissues but applies to many inflammatory lesions. Many data now show that cognate interactions between T cells and adjacent macrophages constitute a major pathway driving cytokine release and that cytokines sustain this pathway (see Figure 26-3). Such interactions do not depend on T cell receptor–mediated T cell activation and provide a route to expansion of inflammation by T cells independent of local autoimmune recognition.

Vey and colleagues first observed monocyte activation via cell contact with mitogen-stimulated T cells. Freshly isolated synovial T cells activate macrophages by this mechanism, confirming that contact-induced cellular activation is a fundamental property of inflammatory T cells. Antigen-independent, cytokine-mediated bystander activation confers this capacity on human CD4+ memory T cells. Studies using synovial T cells from rheumatoid arthritis and psoriatic arthritis tissues reveal that exposure of memory T cells to synergistic combinations of cytokines are most potent in this respect, particularly IL-15, TNF, and IL-6. Cytokine activities also operate directly on macrophages to synergize with T cell contact. IFN-γ and IL-18 are most potent in this respect, acting via increased adhesion molecule expression.

Activated memory CD4+ and CD8+ T cells promote cytokine release from macrophages via diverse membrane ligands including LFA-1/ICAM-1, CD69, and CD40/CD154. After contact with T cells, macrophages release increased concentrations of TNF and IL-1, but not IL-10, and they exhibit reduced levels of IL-1Ra. Th1 cells promote relatively greater proinflammatory cytokine release than do Th2 cells after co-culture. This finding suggests that their functional phenotype extends beyond cytokine secretion to include a differential membrane receptor array.

This suggestion is borne out further in the relative phenotypic distinctions between Th1 (CD40L, CCR5, IL-18Rα) and Th2 (IL-33R/ST2, CXCRI3) cells. The role of Th17 cells in this context is unknown. Signaling pathways engaged in monocytes after such T cell–membrane interactions are distinct from the pathways activated by conventional cytokine-inducing agents. Distinct use of phosphatidylinositol 3-kinase, NFκB, and p38 mitogen-activated protein kinase pathways is observed. Similarly, discrete macrophage signals follow contact with
cytokine-activated T cells (which resemble synovial T cells) compared with T cell receptor–activated T cells. Such distinctions offer therapeutic potential in targeting cytokine-activated, T cell–driven pathways, leaving antigen-driven responses relatively intact. The activation state of memory T cells necessary for the previously discussed interactions to proceed remains controversial. Purified resting T cell subsets activate synovial fibroblasts to release IL-6, IL-8, matrix metalloproteinase 3 (MMP3), and prostanooids, in synergy with IL-17.

It is likely that T cells are activated by interactions with diverse moieties including extracellular matrix components and potentially autoantigens. Nevertheless, it is now clear that cytokines can promote chronicity by activating T cells to promote inflammation regardless of local (auto)antigen recognition and that this has enormous therapeutic potential (see Figure 26-3).

**Agonist/Antagonist Cytokine Activities in Chronic Inflammation**

Complex regulatory interactions exist to suppress ongoing inflammatory responses. This is often achieved via parallel secretion of antagonistic cytokines and soluble receptors to regulate cytokine effector pathways. Th1 responses are suppressed partly by cytokines of Th2 type (e.g., IL-4, IL-10), and consequently exaggerated Th1 responses arise in models in which the Th2 response is deficient. Th1 and Th2 cells similarly limit Th17 cell expansion. Similar regulatory loops operate for other leukocytes, exemplified by the yin-yang effects of TNF and IL-10 on macrophage cytokine release and effector function.

Inhibitory cytokine activities are usually defined with respect to a proinflammatory cytokine, and in other contexts they may have quite distinct function, rendering prediction of their net contribution to an inflammatory response difficult. IL-10 opposes many of the proinflammatory effects of TNF and IL-1β (e.g., reduces adhesion molecule expression, MHC expression, and MMP release), but it potently activates B cell activation and immunoglobulin secretion. Similarly, TNF, which is normally considered a proinflammatory moiety, may have an important role in regulating T cell function because T cells removed from sites of chronic inflammation exhibit suppressed capacity to signal via their T cell receptor that recovers on TNF neutralization. Such regulation is complicated further by the precise ratio of cytokine to soluble receptor such as TNF to sTNFR or IL-10 to sIL-10R within the local environment. Commensurate with this, administration of anti-inflammatory cytokines such as IL-4, IL-10, and IL-11 has generally proved disappointing in the context of clinical inflammatory diseases. Suppressed IL-10 production might be one unanticipated effect of p38 MAP kinase inhibitors, with efficacy of these agents and underscoring the importance of considering the complex interplay between proinflammatory and anti-inflammatory cytokines. An important caveat is the potential requirement of combinations of cytokines to suppress inflammation optimally (e.g., combinations including IL-4, IL-10, and IL-11). Further functional antagonism is exemplified in the antagonistic activities of IL-1β and IL-1Ra and of IL-18 and IL-18 binding protein in regulating macrophage activation.

The role of cytokines in regulating cognate interactions between leukocytes has emerged more recently. Although anti-inflammatory pathways are poorly induced after cell contact, cytokine-activated T cells can induce IL-10 release by monocytes. Rheumatoid arthritis synovial membrane IL-10 release, which is partially T cell dependent, feeds back to regulate TNF release. Cytokine production from adjacent cell lineages within an inflammatory lesion may also be suppressive. IFN-β reduces mitogen-activated, T cell–induced macrophage release of TNF and IL-1, whereas IL-1Ra release is enhanced. This provides a mechanism whereby type I IFNs could modify proinflammatory cytokine production. Regulation extends beyond conventional cytokine activities. Prostaglandins and lipoprotein moieties, particularly high-density lipoprotein, can suppress cytokine-mediated, T cell–macrophage interactions.

**Disease-Modifying Antirheumatic Drugs**

Conventional disease-modifying antirheumatic drugs can also act via modulation of cytokine production. Methotrexate modulates release of various cytokines in vitro, in part mediated via the adenosine–cyclic adenosine monophosphate pathway. The active metabolite of the dihydroorotate dehydrogenase inhibitor leflunomide, A77 1726, reduces TNF, IL-1β, IL-6, and MMP-1, but it does not reduce IL-1Ra release by monocytes after mitogen-activated T cell contact. Leflunomide may mediate these activities through modulation of inhibitor of NFκB (IκB)α phosphorylation and degradation and AP-1 and c-Jun N-terminal protein kinase activation. Sulfasalazine is an inhibitor of proinflammatory cytokine–induced NFκB. Biologic agents also potently modify cytokine expression in a variety of disease states. Signal transduction inhibitors such as agents that inhibit JAK and Syk not only inhibit cytokine action but also the subsequent production of additional cytokines that can exacerbate disease.

**Cellular Interactions across Diverse Tissues**

Cytokines promote cognate cellular interactions across a range of tissues. In contrast to T cell–macrophage and T cell–dendritic cell interactions, in which adhesion molecule and co-stimulatory pathways are often implicated, cell-cell membrane communications apart from leukocyte-leukocyte interactions are often mediated through membrane cytokine expression (see Figure 26-3). T cell contact–mediated activation of fibroblasts operates via membrane TNF and IFN-γ to enhance fibroblast cytokine release and MMP—but not tissue inhibitors of metalloproteinase—expression, favoring tissue destruction. The source and activation status of the fibroblast are vital; fibroblast-like synoviocytes, but not cutaneous fibroblasts, are potently activated by this route. Other studies have shown T cell contact–mediated activation of neutrophils, keratinocytes, mesangial cells (via a combination of membrane cytokine and CD40L expression), platelets, and renal tubule epithelial cells. Cytokine-activated macrophages (via IFN-γ or sCD40L) may interact via cell contact with mesangial cells to activate adhesion molecule and chemokine release by the latter. Cell-cell contact between cells of the immune system and beyond likely represents a ubiquitous mechanism whereby
perpetuation of chronic inflammation is potently influenced by local production of cytokines.

**B Cells and Cytokine Release in Chronic Inflammation**

Cytokines are crucial to B cell maturation, proliferation, activation, isotype switching, and survival (see Chapter 14). Cytokine-mediated B cell activation is important in immune complex generation, B cell antigen presentation, B cell–T cell interactions, and germinal center formation. Particular importance has been placed on the TNF superfamily cytokines, BLyS and APRIL. These cytokines are crucial for B cell development, survival, and optimal activation. Their hierarchical place is, however, now disputed at least in terms of RA biology because clinical targeting of both using atacicept (which neutralizes both BLyS and APRIL) has proven ineffectivc in RA. Neutralization of BLyS with a monoclonal antibody has been similarly disappointing in RA, although more promising data have emerged in the treatment of SLE. B cells in turn represent a potent source of cytokines such as IL-6 and IL-10. B cells also have been considered important inducers of macrophage-derived cytokine release. This process may operate primarily via immune complex formation or through regulation of T cell activation (with B cell help). Complex regulatory feedback loops involving cytokine expression and B cells are likely important in a range of rheumatic diseases in which B cells are of paramount pathophysiologic importance. This may be one core mechanism whereby B cell depleting strategies in rheumatic diseases (e.g., via use of rituximab) are mediating their effects.

**Innate Cell Lineages in Chronic Inflammation**

Cytokines potently activate innate response cells that contribute to the chronic inflammatory lesion of a variety of rheumatic diseases. Tables 26-1 through 26-8 document relevant examples in which neutrophils, natural killer cells, eosinophils, and mast cells may be recruited and activated by the presence of appropriate cytokine combinations.

**Growth Factors in Chronic Inflammation**

Many data document the importance of growth factor families in chronic inflammation. TGF-β superfamily members including TGF-β isoforms and bone morphogenetic protein family members warrant particular reference. TGF-β is critically involved in processes of cell proliferation, differentiation, inflammation, and wound healing. Bone morphogenetic proteins, in addition to regulating inflammatory responses, are paramount in determining cartilage and bone tissue development and remodeling. As such, they are of increasing interest in the pathogenesis of several rheumatic diseases.

**CYTOKINE EFFECTS BEYOND IMMUNE REGULATION**

A striking feature of the cytokine field concerns the broad functional pleiotropy exemplified in the effects of cytokines in normal physiologic and adaptive processes. Cytokine activities are found in muscle, adipose tissue, central nervous system, and liver, mediating normal regulation of metabolic pathways and modulation imposed by altered tissue conditions. Examples are found not only in the release of adipokines that regulate adipose metabolic pathways but also in the release of conventional cytokines by fat pads in inflammatory synovitis. Because cytokines thereby likely mediate normal and pathophysiologic activities in many tissues, it is increasingly recognized that they may underlie the co-morbidity that is observed in vascular, central nervous system and bone tissues in several rheumatic diseases. Thus cytokines or their receptors arising from the primary target tissues (e.g., joint, kidney) may “leak” into the circulation and promote excess pathology in other tissues. Commensurate with this targeting, such cytokines may modulate this co-morbid risk. This is now exemplified in the reduction of vascular morbidity in patients receiving TNFi.

**SUMMARY**

Cytokines represent a diverse family of glycoproteins active across a broad range of tissues. Their pleiotropic functions and propensity for synergistic interactions and functional redundancy render them intriguing therapeutic targets. So far, single cytokine targeting has proved useful in several rheumatic disease states. Further elucidation of the biology and functional interactions within this expanding family of bioactive moieties is likely to prove informative in resolving pathogenesis and in generating novel therapeutic options.

**References**


The references for this chapter can also be found on www.expertconsult.com.
Cell Survival and Death in Rheumatic Diseases

KEITH B. ELKON

Further developments in apoptosis paralleled advances in molecular biology, genetics, and biochemistry. The detection of a nucleosomal ladder was of considerable importance, because it defined a biochemical event (i.e., nucleosomal cleavage) and provided a simple electrophoretic test for detection of apoptotic cell death that remains a standard in the field (see Laboratory Detection of Apoptosis later in this chapter). Another landmark was the discovery in the 1980s that the death of cells during development of the nematode Caenorhabditis elegans was under strict genetic control. Remarkably, the death of these cells could be perturbed by mutations of a small number of genes called ced (for cell death abnormal) genes. Horvitz and colleagues determined that two ced genes, ced3 and ced4, encoded death effectors, whereas ced9 was an antiapoptotic gene. Most of the remaining ced genes were responsible for engulfment and removal of “corpses.” This simple model in which CED-3 is the main death protease that is activated by CED-4 and inhibited by CED-9 has served as a paradigm for defining apoptotic pathways in mammalian cells (Figure 27-2). In 2002, Robert Horvitz was awarded the Nobel Prize for discoveries concerning the genetic regulation of programmed cell death.

Mammalian cells are much more complex and, as will be discussed in detail, have multiple defined pathways that follow the basic C. elegans model. The molecules within these pathways, the downstream effectors of apoptosis, the caspases (cysteine aspartate proteases), and the proteins implicated in the clearance of apoptotic cells are discussed in detail here. Regulation of cell death is of seminal importance in a number of diseases, including cancers, autoimmune diseases, and degenerative disorders. The relevance of apoptosis to rheumatic disorders is summarized at the end of the chapter.

Programmed Cell Death

Whereas apoptosis originally referred to the appearance of dying cells in certain contexts, as explained earlier, the concepts of atrophy, cellular or tissue involution or regression, and degeneration had been appreciated for hundreds of years, yet the two phenomena were not associated until relatively recently. Perhaps the most precise descriptions of cells that died in an orderly and apparently programmed fashion were documented in developmental biology. Examples included the involution of cells between digits, the metamorphosis of insect larvae, and the death of specific cells during development of C. elegans.
Figure 27-1  Electron microscopic morphology of cell death. A, A cytotoxic T cell (lower left) conjugated to its target, P815 (a murine mast cell), before initiation of cell death. B, Induction of apoptotic changes in P815. Note the reduction in target cell size, the nuclear condensation, and the vacuoles with relative preservation of organelles. C, Osmotic lysis and necrosis in P815 induced by antibody and complement. Note the increased size of the nucleus and the apparently random fragmentation of the chromatin. Organelles are severely disrupted. D, Electron microscopic appearance of autophagy. An autophagosome (AP) is observed in a normal rat kidney cell. The autophagosome is surrounded by two cisterns of rough endoplasmic reticulum (ER). A Golgi stack (Go) is visible next to the autophagosome. (C, Adapted from Russell JH, Masakowski V, Rucinsky T, et al: Mechanisms of immune lysis III: characterization of the nature and kinetics of the cytotoxic T lymphocyte induced nuclear lesion in the target, J Immunol 128:2087, 1982, with permission. D, Figure kindly provided by Eeva-Liisa Eskelinen, Department of Biosciences, Helsinki, Finland.)

Figure 27-2  Caenorhabditis elegans paradigm of apoptosis. Genes involved in the regulation, execution, and clearance of apoptotic cells during development of C. elegans and their mammalian homologues are shown. In this figure, only the most closely related homologues are indicated, but as described in the text and shown in the figures, the complexity in mammalian cells is much greater. Note that at least half of the cell death abnormal (CED) proteins are involved in engulfment and removal of apoptotic corpses. Two distinct but partially overlapping pathways of engulfment are present in C. elegans: CED-2, -12, -5, and -10, which most likely regulate cytoskeletal changes, and CED-1 and -6, which may be involved in recognition (CED-1) and upstream signal transduction (CED-6). CED-7 is homologous to the mammalian adenosine triphosphate (ATP)-binding cassette transporter-1 (ABC-1), which likely affects membrane dynamics. The lower right panel shows a schematic of the two proposed pathways for apoptotic cell ingestion in mammalian cells. In pathway (1), an unknown receptor(s) triggers activation of the GTPase, RhoG, which leads to transport of the scaffolding protein, ELMO, to the cell membrane. There, Dock 180, a guanine exchange factor, promotes Rac activation, leading to activation of the cytoskeleton and phagocytosis. In pathway (2), activation of the CED-1 homologue (most likely CD91) interacts with the adapter protein, GULP, and downstream activation of the cytoskeleton.
**Autophagy**

*Autophagy*, which means “to eat oneself,” is essentially a protective process that occurs after growth factor withdrawal or nutrient depletion that may or may not result in cell death. The process is highly conserved from yeast through humans. Under circumstances of nutrient depletion, cells switch to a catabolic program in which cellular constituents are degraded for energy production as a survival mechanism. In these cells, a double membrane vesicle (autophagosome) forms around organelles or protein aggregates and encapsulates them (see Figure 27-1D). A characteristic feature of this process is the lipidation of LC3 (microtubule-associated protein 1 light chain 3) to form LC3-II, which can be identified as coarse dots on the autophagosomes seen by microscopy (see Figure 27-6I, later), or by changes in molecular mass on Western blot analysis. The autophagosomes then fuse with the lysosomes, leading to degradation of their contents (Figure 27-3).

The biochemical pathways by which autophagy is initiated and executed are complex and are schematically outlined in Figure 27-3. In brief, lack of nutrients leads to a reduction in phosphoinositide (PI)-3 kinase activity, resulting in loss of suppression of mTOR (mammalian target of rapamycin) and activation of autophagic (ATG) proteins. The ATG gene family comprises at least 20 members and includes beclin1 (ATG6), a protein that is critically involved in the regulation of autophagic cell death.7 Beclin1 is monallelically deleted in a variety of human cancers, and may function in control of cell growth and in tumor suppression.9 Autophagic death shares a number of morphologic features with necrotic cell death and may be seen in neuronal cell death associated with polyglutamine repeats. Many reports indicate that autophagy is important in immune function. Examples include intracellular host response to pathogens, survival of lymphocytes after growth factor withdrawal, Toll-like receptor (TLR) stimulation in plasmacytoid dendritic cells (pDCs), and major histocompatibility complex (MHC) class II presentation of antigen.10,11 The strongest link between autophagy and autoimmunity or autoinflammatory disease is the association between a genetic variant of ATG16L1 and another protein that regulates autophagy, the immunity-related guanosine triphosphatase gene (IRGM), and Crohn’s disease. The precise mechanisms that account for this association remain to be determined.11

**Necrosis**

Necrosis (from the Greek nekros, meaning “corpse”) is distinguished from apoptosis predominantly by morphologic appearance.12 Necrotic cells are swollen, and electron microscopy reveals disorderly fragmentation of chromatin and severe damage to the mitochondria (see Figure 27-1). The cellular membrane loses integrity and becomes permeable to vital dyes such as trypan blue and propidium iodide (see Figure 27-6, later). The distinction between apoptosis and necrosis remains important from a number of perspectives. In contrast to the genetic and biochemical programs that regulate apoptosis, necrotic cells usually result from death “by accident,” that is, from thermal or drug injury, infection, or infarction of an organ. Because of uncontrolled release of intracellular contents, necrotic cells induce a proinflammatory immune response, whereas apoptotic cells usually elicit an anti-inflammatory response.

The same inducers (e.g., ischemia, hydrogen peroxide) may produce apoptosis or necrosis, depending on the severity of the injury and the rapidity of cell death. The fate of the cell is determined in part by cellular energy reserves, especially adenosine triphosphate (ATP).13 ATP is generated by oxidative phosphorylation in mitochondria, as well as by glycolysis in the cytosol. Some inducers initially may cause apoptosis followed by necrosis (postapoptotic necrosis). This is likely to occur when removal of apoptotic cells is delayed and is thought to be important in stimulating an inflammatory response to self-antigens. A unique type of necrosis in neutrophils, called NETosis, is implicated in SLE (see Defective Uptake and Processing of Apoptotic Cells).

**Pyroptosis**

*Pyroptosis* (from the Greek pyro, meaning “fire”) is distinguished from other forms of cell death first and foremost by the activation of caspase 1 (interleukin [IL]-1β–converting enzyme) and secretion of the inflammatory cytokine, IL-1β. Pyroptosis is most strongly associated with infection by intracellular bacteria such as *Salmonella*, *Yersinia*, and *Legionella*, although it may also be seen following tissue infarction.14 Cells undergoing pyroptosis demonstrate nuclear condensation associated with DNA damage, cell swelling, and ultimately cell lysis associated with release of IL-1β.14 The mechanisms responsible for this process involve intracellular sensors of bacterial products and formation of the

---

**Figure 27-3** Pathways of autophagy (ATG). Autophagy is initiated by starvation or by activation of intracellular immune sensors. ATG proteins are activated, leading to the nucleation of an isolation membrane (green rectangle) and conjugation of phosphatidylethanolamine to LC3. Continued concerted activities of ATG proteins lead to formation of the autophagosome from the isolation membrane, which then fuses with lysosomes to form autolysosomes. The contents of the autolysosomes are degraded, and the products recycled for energy utilization.
CHAPTER 27 | CELL SURVIVAL AND DEATH IN RHEUMATIC DISEASES

inflammasome. These topics are briefly described in the following sections and elsewhere in the text (see Chapter 18).

BIOCHEMISTRY OF APOPTOSIS

A schematic diagram of the cell death program is shown in Figure 27-4. A brief outline of each major functional component within the program, from signals for death to removal of apoptotic cells, is provided here, but space limitations preclude a detailed analysis of the layers of regulation at each step of the pathway. For example, post-translational protein modifications such as phosphorylation, nitrosylation, and oxidation present additional complexities that are under intense study. A series of reviews in *Apoptosis and Its Relevance to Autoimmunity* offer more detailed information on the biochemistry of apoptotic pathways and their relationship to immune function.

The specialized proteins involved in apoptosis and its regulation contain a number of modules or domains that are predominantly involved in promoting protein-protein interactions (Table 27-1). These domains may be found

---

**Table 27-1 Modular Components of Proteins Involved in Apoptosis and Inflammation**

<table>
<thead>
<tr>
<th>Module*</th>
<th>Component of Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>BH (1-4)</td>
<td>Bcl-2 family</td>
</tr>
<tr>
<td>BIR</td>
<td>IAP family</td>
</tr>
<tr>
<td>CARD</td>
<td>Caspases, adapters, NODs</td>
</tr>
<tr>
<td>DD</td>
<td>Death receptors, adapters, kinases</td>
</tr>
<tr>
<td>DED</td>
<td>Adapters, caspases</td>
</tr>
<tr>
<td>LRR</td>
<td>Pyrin family, NODs, TLR</td>
</tr>
<tr>
<td>NBS/NOD</td>
<td>Pyrin family, NODs</td>
</tr>
<tr>
<td>Pyrin</td>
<td>Pyrin family</td>
</tr>
<tr>
<td>RING finger</td>
<td>IAP family</td>
</tr>
</tbody>
</table>

*See Mariathasan and Monack for further discussion.

BH, Bcl-2 homology; BIR, baculovirus AIP repeat; CARD, caspase recruitment domain; DD, death domain; DED, death effector domain; IAP, inhibitor of apoptosis; LRRs, leucine-rich repeats; NBS, nucleotide-binding site; NOD, nucleotide oligomerization domain; P-P-I, protein-protein interaction; TLR, Toll-like receptor.

---

**Figure 27-4** Mammalian apoptotic pathways. Cell death can be initiated by multiple pathways, including an extrinsic ligand-induced pathway (left), an intrinsic pathway mediated by the mitochondria (middle), and an intrinsic pathway mediated through the endoplasmic reticulum (ER) (right). Examples of stimuli that can induce each of these pathways are shown and are discussed in further detail in the text. Note that these pathways differ in the upstream caspases activated but converge to cleave the effector caspases, such as caspase 3, during execution of apoptosis. Tumor necrosis factor receptor (TNFR) and other “death receptors” can also signal cell survival by activation of nuclear factor kappa B (NFκB). During ER stress, unfolded proteins release the ER chaperone protein, Bip, from binding to the stress sensor proteins, IRE, PERK, and ATF6. PERK attenuates protein synthesis, whereas ATF6 and JNK are transcription factors that upregulate the expression of proapoptotic proteins that contain UPR elements such as CHOP (GADD 153) and caspases. In mice, caspase 12 is the initiator caspase, whereas in human cells, caspases 4 and 7 have been implicated. Alterations that occur during dissolution of the cell are too numerous to mention, but a few are highlighted in view of their potential relevance to autoimmunity (see text). Exposure of phosphatidylserine (PS) on the cell surface (lower left area) provides a simple means of detection of apoptotic cells through binding of annexin V, as well as proteins, such as lamins and DNA-PK, which may be antigenic. AIF, apoptosis inducing factor; CAD, caspase-activated DNase; Cyt c, cytochrome-c; DD, death domain; DED, death effector domain.
in receptors, adapters, effectors, or inhibitors. Furthermore, as will be discussed later, these domains occur in proteins involved in apoptosis, as well as in inflammation (Supplemental Figure 27-1 on www.expertconsult.com). It has been suggested that death domain (DD), death effector domain (DED), caspase recruitment domain (CARD), and pyrin domain all evolved from the prototypic DD-fold, corresponding to an antiparallel six-helix bundle. In general, like domains bind so as to facilitate homotypic interactions, leading to oligomerization of the same protein or binding to different proteins in a signaling pathway. These changes usually produce conformational alterations, which lead to further protein recruitment. Other domains such as the nucleotide-binding site (NBS) specify nucleotide binding.

**DEATH LIGANDS, RECEPTORS, AND SIGNALS**

Death of a cell may result from intracellular stress activating an intrinsic death program, or it may be forced on the cell by the interaction of a death ligand with a death receptor (see Figure 27-4). Death receptors (DRs) belong to the tumor necrosis factor (TNF) receptor (TNFR) superfamily of proteins, which comprises approximately 25 members. This family of receptors is responsible for diverse biologic responses such as inflammation, proliferation, antiviral activity, and cell death. Receptors in the TNF family include at least six receptors capable of transmitting apoptosis (see later), as well as receptors such as CD40, CD30, BlyS/BAFF/TALL, and TACI, which trigger survival and/or proliferation in part through activation of nuclear factor κB (NFκB). Although most receptors of the TNFR family exert their effects primarily within the immune system, some members (e.g., p75NGFR, TNFR I and II) appear to serve important functions in the nervous system and in other organs. Some TNFR-like proteins, such as PV-T2, PV-A53R, and CAR1, are encoded by viruses and may contribute to their virulence.

**Death Receptors**

The DRs identified, including Fas, TNFR I, DR-3 (TRAMP/wsl/APO-3), DR4 (TRAIL, receptor for the Apo2 ligand), DR5, and DR6, share homology in their intracellular domains over a 70 amino acid region called the death domain (DD). Three decoy receptors have been identified—two (DcR1 and DcR2) that bind and inhibit their ligand, TRAIL, and one (DcR3) that binds Fas ligand. These decoy receptors presumably modulate the cytotoxic function of the ligands, but their biologic contexts remain to be fully defined. Use of alternative splice forms and shedding of the receptors and ligands also downmodulate their function.

TNFR family members are characterized by two to six cysteine-rich domains (CRDs) in their extracellular regions. The co-crystal structure of TNFR I and lymphotoxin-α indicates that CRDs project from the cell surface in a linear array, making distinct contact with ligands at subunit interfaces. The first CRD may also be responsible for preassembly of the receptor as trimers that undergo further conformational alteration upon ligand engagement.

The three-dimensional structure of the DD has been solved by nuclear magnetic resonance spectroscopy and has been shown to consist of six amphipathic α-helices that create a unique fold. Functionally, the DD appears to be a novel protein-protein association motif that facilitates homotypic interactions. For example, the DD of Fas and TNFR I self-associate, thereby recruiting adapter proteins that also contain DD and that directly or indirectly mediate receptor signal transduction (see Figure 27-4).

**Death Receptor Signal Transduction**

This section will focus predominantly on signaling from Fas and TNFR, because these are the best-characterized members of the death receptor subfamily, and it is likely that other death receptors signal through similar pathways. As illustrated in Figure 27-4, Fas and TNFR I share a common death pathway. Binding of Fas ligand to Fas causes conformational changes in the receptor cluster, leading to recruitment of intracellular adapter molecules. Initially, aggregation of Fas induces uptake of the adapter protein, FADD, to the DD of Fas. FADD has two structural domains: a C-terminal DD, which mediates Fas binding, and an N-terminal death effector domain (DED). The FADD DED allows recruitment of procaspase 8 and procaspase 10 through DED-DED interactions. Procaspases 8 and 10 have a bipartite structure comprising a DED and an enzymatic caspase domain, the latter linking Fas aggregation with the execution phase of apoptosis. Apposition of procaspases 8 and 10 to the activated Fas complex leads to autocatalytic cleavage and conversion of the proenzymes to activated proteases, which are released and are able to initiate a proteolytic cascade, leading to programmed cell death. In some cell lines, caspase 8 cleavage also results in cleavage of the proapoptotic molecule Bid, which activates the mitochondrial amplification cascade (type II pathway); Supplemental Figure 27-2 on www.expertconsult.com; see also Figure 27-4).

Although the six DD-containing receptors initiate cell death in certain contexts, all may signal cell survival/proliferation in different cell types and/or in different contexts. The ability to signal an opposite cell fate depends on the recruitment of proteins such as tumor necrosis factor receptor–associated factors (TRAFs), which activate NFκB, thereby promoting cell survival (see later).

Defective function of caspase 8 leads to a form of cell death called NECROPTOSIS that combines features of apoptosis and necrosis resulting in activation of NFκB and production of inflammatory cytokines such as TNF.

**Death Ligands**

FasL (CD178) is a 40-kD type II transmembrane protein that shares 15% to 35% amino acid identity with the TNF superfamily of ligands. FasL is expressed constitutively in the anterior chamber of the eye and in the testis but is induced when CD8, T helper-1 (Th1) CD4+ T cell, and some natural killer (NK) cell populations become activated. In lymphocytes, expression of ligand is tightly regulated and activity on the cell surface short-lived, because
Supplemental Figure 27-1  Components of inflammasomes. Inflammasomes contain pro-interleukin (IL)-1 but different sensors and regulators. In most cases, two stimuli (e.g., activation of a Toll-like receptor [TLR] or nucleotide-binding oligomerization domain [NOD] sensor plus potassium flux) are required to activate the inflammasome. The prototypic inflammasome containing caspase 1 or 5 (caspase 11 or 12 in mice) and the adapter protein, ASC (Pycard), is shown. A NALP3-containing inflammasome is of rheumatologic interest because it is activated by uric acid and certain nucleic acids. NOD2 is mutated in Crohn’s disease and Blau syndrome. Pyrin is mutated in familial Mediterranean fever (FMF). The domain structure of these proteins (see Table 27-1) is as follows: \( a \) = CARD domain; \( b \) = nucleotide-binding site; \( c \) = leucine-rich repeat; \( d \) = pyrin domain.
excessive, if unfolded proteins persist, or if post-translational protein modification is abnormal. Apoptosis is the result of three central players: inositol-requiring protein-1 (IRE1), activating transcription factor-6 (ATF6), and protein kinase RNA-like ER kinase (PERK) (see Figure 27-4). Unfolded proteins may activate these proteins directly, or by binding to the ER chaperone immunoglobulin-binding protein, Bip. Once activated, the three proteins cause selective RNA cleavage, termination of protein synthesis, and induction of proapoptotic responses (see Figure 27-4). In contrast to the death receptor or mitochondrial pathways, apoptosis is executed through caspase 12 in mice and possibly caspase 4 in mammals. However, many of the same molecules that regulate apoptosis in the mitochondria, the Bcl-2 family in particular, also influence ER-mediated apoptosis, possibly through control of intraluminal calcium. It is important to point out that induction of the UPR, in certain circumstances, leads to NFκB activation and inflammation, rather than cell death.

### Antiapoptotic Proteins: FLIP, Bcl-2, IAPs, and Akt

(See Supplemental Figure 27-2 on www.expertconsult.com.) Cellular homeostasis within each tissue is carefully regulated. Excessive cell growth or premature cell death translates into disease, as will be discussed in the last part of this chapter. The death and survival pathways are finely balanced, and each cell death program is regulated, often at multiple levels. Before we discuss how the death program is executed, we will describe the way in which cell death is modulated by inhibitors of apoptosis.

### Inhibition of Death Receptors

In most resting cell types that express Fas on their cell surface (e.g., lymphocytes), the receptor is nonfunctional. Resistance to death is explained both by low levels of expression of the receptor and by active inhibition by a protein called FLIP. FLIP resembles the structure of caspase 8 and competes with caspase 8 for recruitment to FADD. This prevents FADD from initiating apoptosis. When lymphocytes become activated, FLIP is usually degraded, allowing Fas signal transduction to occur unimpeded (see Supplemental Figure 27-2 on www.expertconsult.com). Similarly, a protein called SODD (silencer of death domain) attenuates TNFR1 signal transduction.

### Bcl-2 Family of Cell Death Regulators

The Bcl family comprises more than 18 members. Bcl-2 is the prototype antiapoptotic protein that was first discovered to be overexpressed in certain B cell lymphomas (Bcl). Of particular significance, Bcl-2 overexpression did not enhance cell proliferation (most cells were in the G2/M phase of the cell cycle) but rendered the cells more resistant to death. The antiapoptotic members—Bcl-2, Bcl-XL, Bcl-w, Mcl-1, and A1—contain three or four of the characteristic Bcl-2 homology (BH) domain motifs and most possess the N-terminal BH4 domain and the hydrophobic C-terminal membrane anchor, accounting for their attachment to mitochondrial, ER, and nuclear membranes (see Supplemental Figure 27-2 on www.expertconsult.com). Virus-encoded proteins—BHRF1, LMW5-HL, ORF16, KS-Bcl-2, and E1B-19K—and Bcl-2 have similar antiapoptotic functional properties. The proapoptotic members of this family can be subdivided into two groups: the Bax/Bak-like proteins (Bax, Bak, Bok, and Bcl-Xs), which contain two to three BH3 domains, and the BH3-only subset (Bad, Bik, Bid, Hrk, Bim, Noxa, Puma, and Bmf), which contain only the single domain. Bcl2s are regulated at transcriptional and post-transcriptional levels by a multitude of stimuli.

How do Bcl2s regulate apoptosis? One level of regulation is conferred by binding interactions (homodimerization or heterodimerization) between members via their BH1, BH2, and BH3 domains. Although outcomes vary for each specific pair, homodimerization of Bcl-2 or Bax potentiates their antiapoptotic or proapoptotic functions, respectively, whereas heterodimers may potentiate or abrogate function of one member of the pair. Bax and Bak have been shown to be pivotal downstream effectors of intrinsic apoptotic pathways. A possible model is that Bcl-2 (or homologues) usually heterodimerizes with Bax and Bak, preventing apoptosis. Increased expression of a BH3 proapoptotic protein causes binding to Bcl-2, thereby releasing Bax and Bak to induce apoptosis. It has also been suggested that BH3 proteins may bind and directly activate Bax/Bak.

Bcl regulation of cell death is closely connected to mitochondrial function. The physical association of Bcl-2 family proteins with the outer mitochondrial membrane, as well as the close structural similarity between BH1 and BH2 domains and bacterial pore-forming proteins such as colicin, allows them to regulate ion fluxes or the transfer of small molecules from the membrane. In vitro models suggest that Bax and Bak promote opening of the VDAC, allowing the release of cytochrome-c into the cytosol, whereas Bcl-2 binds directly to VDAC and closes it.

### Intracellular Inhibitors of Apoptosis (IAPs)

Intracellular inhibitors of apoptosis (IAPs) are a separate family of antiapoptotic proteins that are highly conserved through evolution. The neuronal apoptosis inhibitory protein (NAIP) was discovered through association of NAIP mutations in patients with the severe form of spinal muscular atrophy. Seven additional members of the family (c-IAP-1, -2, X-IAP, survivin, ILP2, ML-IAP, and Bruce) that share a baculovirus IAP repeat (BIR) domain have subsequently been identified; most contain a RING domain that functions as an E3 ligase. Ubiquitylation may target interacting proteins for proteasomal degradation, or they may be activated. IAPs such as X-IAP directly inhibit effector caspases, especially caspase 9 (see Supplemental Figure 27-2 on www.expertconsult.com), whereas cIAPs modulate cell survival through ubiquitylation of substrates such as RIP and proteins in the NFκB pathway (see Supplemental Figure 27-2 on www.expertconsult.com). IAPs block apoptosis induced by a variety of stimuli, including Fas, TNF, ultraviolet irradiation, and serum withdrawal, and survivin is overexpressed in certain cancers and in the rheumatoid arthritis (RA) synovium. In some cells, the antiapoptotic effect of IAPs is eliminated by release of the protein Smac/
Supplemental Figure 27-2  Inhibitors of apoptosis. Inhibitors of death pathways are shown in brown. Note that a number of tumor necrosis factor (TNF) family members (e.g., CD40L, BAFF/BLyS/TALL, receptor activator of NFκB [RANK] ligand) activate the nuclear factor κB (NFκB) pathway, resulting in increased transcription of antiapoptotic proteins, inhibitors of apoptosis (iAPs), Bcl-2 family members, and A1. In some cases, inhibitors of apoptosis are blocked or destroyed during apoptosis. For example, the protein Smac/Diablo is released from mitochondria and eliminates the antiapoptotic effect of the iAP family. Growth factors impart survival signals, in part through the phosphoinositide (PI)-3 kinase → Akt pathway (right side of diagram). Activation of the kinase, Akt, directly inactivates certain proapoptotic proteins such as Bad and caspase 9, and promotes survival by inhibiting the forkhead family (FKRL) of transcription factors and enhancing the expression of the antiapoptotic protein, NFκB. Inhibitor of caspase-activated DNase (ICAD) is a specific inhibitor of the DNase, CAD, which is degraded by caspases. Tetrapeptide inhibitors of specific caspases are shown in blue.
metalloproteases cleave the extracellular portion of the ligand into soluble, functional molecules. The zinc metalloprotease, TNF-converting enzyme (TACE), is a membrane-anchored member of the disintegrin family of proteases that cleave active TNF from the cell surface.23,24

**Function in Immune Regulation**

Although Fas and FasL interactions in the thymus are not thought to play a major role in negative selection, this pathway is involved in the maintenance of immune privilege in the eye and the testis, in the pathogenesis of graft-versus-host disease, and in immune evasion by tumor.25 The major physiologic function of Fas and FasL in the immune system is the preservation of peripheral tolerance. This is achieved by the phenomenon of activation-induced cell death (AICD), whereby CD8 T cells, Th1 CD4+ T cells, and possibly NK cells induce apoptosis of activated T cells, B cells, and macrophages. Deletion of activated immune cells removes the source of proinflammatory molecules, prevents the continued presentation of self-peptides by primed (high levels of co-stimulatory molecules) antigen-presenting cells, and eliminates B cells that have mutated to self-specificity within the germinal centers.26 More recently, it has been shown that Fas-FasL interactions promote a variety of additional functions, including early T cell proliferation, tumor survival, and cell migration. These topics are discussed in greater detail elsewhere,27,28 and the consequences of Fas deficiency are described later in this chapter.

It should be noted that, whereas TRAIL signals apoptosis through DR4 and DR5 predominantly in tumor cells, evidence suggests that TRAIL also plays a role in negative selection of thymocytes.29 Similarly, DR3 (TWEAK, receptor for the Apo3 ligand) has been implicated in negative selection.29 Finally, DR6 plays a role in immunologic homeostasis, as evidenced by enhanced T and B cell proliferation in DR6-deficient mice.30

**INTRINSIC CELL DEATH PATHWAYS: INITIATION AND EXECUTION OF APOPTOSIS**

Cells need constant sources of nutrition and depend on a variety of signals for active maintenance of survival. Loss of signals from neighboring cells31 or withdrawal of growth factors or cytokines results in initiation of a cell death program. Damage or stress to intracellular organelles may be induced from outside or within the cell. Here, injury or stress to DNA, mitochondria, and the endoplasmic reticulum is discussed.

**Genotoxic Injury**

Mutations occur frequently in mammalian DNA and usually are promptly repaired. However, if repair fails or DNA is severely damaged by radiation or drugs, the transcription factor, p53 (“guardian of the genome”), is upregulated and phosphorylated by DNA damage sensors such as ATR and ATM. Activated p53 induces cell cycle arrest through induction of the cyclin-dependent kinase inhibitor, p21. If DNA damage is repaired, cell cycle arrest is abrogated, whereas if the injury cannot be repaired, the cell undergoes apoptosis. The critical importance of p53 as a tumor suppressor is illustrated by the high frequency of p53 mutations in cancers.32 p53 induces apoptosis, in part through transcription of death effectors such as Bax that cause mitochondrial stress. In addition, activation of the transcription factor Foxo-1 upregulates the expression of Bim, FasL, and TRAIL.33

**Mitochondrial Stress**

Mitochondria are cytoplasmic organelles that contain their own 16-kb genome encased by inner and outer membranes, with a number of proteins, including cytochrome-c, situated between these membranes (see Figure 27-4). Mitochondria help to maintain redox potential and serve as the energy powerhouse of the cell through the generation of ATP by oxidative phosphorylation. These biochemical pathways create an electrochemical gradient (Δψ) that is positive and acidic on the outside and alkaline on the inside of the mitochondrial membrane. Spanning the inner membrane is the adenine nuclear translocator (ANT), which mediates ATP transport (with VDAC, see later) to the cytosol. On the outer mitochondrial membrane (OMM) is situated the voltage-dependent anion channel (VDAC), which is permeable to solutes of approximately 5000 kD.34

Genotoxic injury, reduced supply of nutritional or growth factors, raised intracellular calcium, reactive oxygen intermediates (ROIs), and exposure to certain chemicals such as staurosporine cause mitochondrial stress. These initiating factors lead to selective mitochondrial membrane permeabilization (MMP) with resulting dissipation of the proton gradient responsible for the Δψ permeability transition, permeabilization of the outer membrane, and loss of ATP production. Mitochondria themselves are the major producers of ROIs, which, in excess, damage nucleic acids, proteins, and membrane lipids.

Once MMP is initiated, cytochrome-c is released from the intermitochondrial space into the cytosol (see Figure 27-4). In the cytosol, cytochrome-c and the cofactors Apaf-1 and ATP or dATP assemble with caspase 9 to form a molecular aggregate called the apoptosome, which promotes the cleavage of procaspase 9 into its active form.35 Caspase 9 acts on effector caspases such as caspase 3, resulting in the caspase cascade that leads to cleavage and inactivation of a wide variety of substrates within the cell (see Figure 27-4). A caspase-independent apoptosis-inducing factor (AIF) is also released from the mitochondria and induces nuclear changes and cell death by less well defined pathways.36

**Endoplasmic Reticulum (ER) Stress**

The main functions of the ER are to regulate intracellular calcium flux and to promote proper folding of nascent proteins. In the contiguous conduit, the Golgi apparatus, post-translation modifications such as glycosylation and isoprenylation are executed. Elaborate mechanisms are in place to ensure that errors in protein folding do not occur, but if they do, an unfolded protein response (UPR) is initiated. The ER/Golgi initiates apoptosis if calcium flux is
Diablo or Omni/HtrA2 from the mitochondria (see Supplemental Figure 27-2 on www.expertconsult.com).

Akt

Akt is a cytosolic protein kinase (protein kinase B) that serves a special role in prevention of apoptosis, because it links cell activation through PI-3 kinase with multiple transcription factors. Phosphorylation of Akt is antagonized by the phosphatase, PTEN. When phosphorylated, Akt promotes cell survival by altering the function of intrinsic (mitochondrial) and extrinsic (death receptor) pathways of apoptosis. Specifically, this includes inactivation of proapoptotic molecules such as caspase 9 and Bad, and activation of survival pathways that include NFκB and forkhead transcription factors that inhibit FasL (see Supplemental Figure 27-2 on www.expertconsult.com).45 Antiapoptotic molecules may be cell type or context specific.

CASPASES

Caspases are cysteine-containing proteases that have an unusual substrate specificity for peptidyl sequences with a P1 aspartate residue.46,47 These proteases are 30 to 50 kD in size and comprise an amino-terminal prodomain with a large subunit domain and a small subunit domain (see Figure 27-4). Active site cysteine residue is contained within the conserved pentapeptide, QACxG, on the large subunit of the enzyme, whereas most of the substrate specificity is determined by the small subunit. The upstream caspases 8, 9, 10, 2, and 4 have large prodomains that interact with regulatory proteins such as FADD for caspases 8 and 10 and Apaf-1 for caspase 9 (see Figure 27-4). Clustering of these complexes allows autocatalytic cleavage of large and small subdomains to form the active tetramer. Effector caspases such as 3, 6, and 7 have small prodomains and are thought to be cleaved into their active forms by the upstream caspases.

Members of the caspase family can be divided into three functional subgroups on the basis of their substrate specificities.48 Group I members (caspases 1, 4, and 5) are potently inhibited by the serpin CrmA; group II members (caspases 2, 3, and 7) are specific for DExD; and group III members (caspases 6, 8, 9, and 10) are specific for I/V/LExD—a sequence that is also contained at the junctions of the caspase subunits themselves. Significantly, granzyme B produced by cytotoxic T cells has a substrate specificity similar to that of group III caspases and is capable of inducing apoptosis through this pathway. Identification of the substate specificity of caspases has led to a number of practical applications, including the ability to quantify activity using fluorogenic tetrapeptide substrates and blockade of proteolytic activity with noncleavable cell-permeable tetrapeptide analogues (see Supplemental Figure 27-2 on www.expertconsult.com).

Effector caspases are necessary for the execution of apoptosis. They cleave specific substrates such as the structural proteins fodrin, gelsolin, and lamins, which are key intracellular enzymes involved in DNA repair (e.g., poly ADP ribose polymerase, DNA-PK) (see Figures 27-4 and 27-6 [later]). These changes facilitate inactivation of synthetic functions of the cell, dissolution of the nuclear membrane, and packaging of cellular proteins into apoptotic blebs on the cell surface. Caspases also cleave regulatory proteins such as Bcl family members and the inhibitor of caspase-activated DNase (ICAD). Cleavage of ICAD leads to the release of active CAD, which enters the nucleus and cleaves nucleosomes at the linker region, yielding the characteristic “DNA ladder” (see Figures 27-4 and 27-6 [later]).49,50

As was already mentioned in the section on pyroptosis, not all caspases are involved in the execution of apoptosis. Human caspases 1, 4, 5, and 12 and mouse caspases 1, 4, 11, and 12 are most likely involved in inflammation. Caspase 1 and caspase 5 interact to form a multiprotein complex that has been called the inflammasome51 (analogous to binding of the apoptosome caspases 1 and 5 to the adapter proteins, ASC [PYCARD] and NALP1 [DECAP], respectively, by their CARD domains) (see Supplemental Figure 27-1 on www.expertconsult.com). ASC and NALP1 are multi-domain proteins that contain many of the protein interaction domains listed in Table 27-1. The N-terminus of the protein, pyrin, which is mutated in familial Mediterranean fever (FMF), binds to ASC. The C-terminus of pyrin, which is mutated in FMF, binds directly to IL-1β, suggesting that pyrin normally directly exerts an inhibitory effect on IL-1β.52

Caspases are tightly regulated by their own prodomains and by Bcl and IAP family members (see Supplemental Figure 27-2 on www.expertconsult.com). In addition, viral proteins such as serpin CrmA, produced by cowpox, and p35, produced by baculovirus, are potent inhibitors of caspases.

FINDING, REMOVING, AND RESPONDING TO DEAD AND DYING CELLS

Finding the Dying Cell

During apoptosis, the enzyme, calcium-independent phospholipase A2, is activated by caspase cleavage, leading to the generation of lysophosphatidylcholine (LPC) (see Figure 27-5 for complete apoptosis schema). LPC acts both as a soluble chemoattractant for macrophages and as an epitope on the cell membrane for natural immunoglobulin (Ig)M antibodies.53 Additional “find-me” signals include sphingosine-1-phosphate (S1P) and the nucleotides ATP and UTP, which are released by dying cells, especially as the plasma membrane becomes damaged (necrosis).54,55 Intravital microscopy coupled with biochemical studies revealed that cell necrosis led to release of ATP, activation of Nrlp3 inflammasomes, and release of IL-1β.56 Subsequently, a chemokine gradient attracted neutrophils to the site of cell damage and formylated peptides to the necrotic cell.

Eating the Dying Cell

Of the 14 C. elegans death genes (ced1 through ced14), at least half encode proteins that are required for engulfment of apoptotic cells (see Figure 27-2).57 CED-7 is present in the membrane of both the apoptotic cell and the phagocyte, whereas remaining proteins function in the phagocyte to execute two partially overlapping pathways of engulfment.
CED-1 is a receptor that recognizes changes in the apoptotic cell and signals through CED-6 to activate the phagocyte. CED-2, -5, -10, and -12 most likely form a functional complex that activates Rac, which then promotes the cytoskeletal changes required to engulf the apoptotic prey. The mammalian counterparts are described in Figure 27-2, and their signaling pathways are discussed in detail elsewhere.

Within the immune system alone, more than 10⁹ apoptotic cells are removed from the body each day. These apoptotic cells are generated in vast numbers in central lymphoid organs such as the thymus and bone marrow by out-of-frame rearrangements of antigen receptors, negative selection, or partial redundant function of each individual receptor. All identified receptors have other functions, perhaps reflecting the diversity is due to the heterogeneity of the opsonins, cell type selectivity, the context (homeostasis vs. inflammation), and partial redundant function of each individual receptor.

In contrast to live cells, which express cell surface proteins such as CD47 that prevent engulfment (“don’t eat me” signals), changes on dying cells promote their ingestion by phagocytes (“eat me” signals). Chief among these on early apoptotic cells is the appearance of phosphatidylserine (PS) on the cell surface membrane (see Figure 27-5). This membrane asymmetry (PS is usually located on the inner surface of the membrane) is caused by reduced function of a transmembrane transporter (ABC1), as well as secondary selection of high-affinity B cells in germinal centers. Specialized sites of selection (i.e., thymus, bone marrow, and lymphoid follicles) have remarkably efficient phagocytes that rapidly remove the dying cells.

As mentioned earlier, natural IgM antibodies and acute phase proteins such as C-reactive protein (CRP) bind to phosphorylcholine on the cell membranes of dying cells and amplify classic pathway complement deposition. Other serum opsonins or bridging molecules include thrombospondin, which bridges the αvβ3 and CD36 receptors, and collectins (mannose-binding protein, C1q, and surfactant proteins). Collectin-binding receptors are controversial (see reference 66 for discussion). The ER protein, calreticulin, is unique in that it is translocated from the ER to the cell surface of apoptotic cells but can also be detected at low concentrations on live cells.

Despite the detection of only limited chemical alterations on the apoptotic cell membrane, blockade of a large and diverse number of receptors on phagocytes can impair the uptake of apoptotic cells (see Figure 27-5). This diversity is due to the heterogeneity of the opsonins, cell type selectivity, the context (homeostasis vs. inflammation), and partial redundant function of each individual receptor. All identified receptors have other functions, perhaps reflecting an evolution from receptors designed to remove apoptotic cells during development to pattern recognition receptors useful for host defense. Many of the receptors are integrins comprising the vitronectin receptor, αβ6, αβ5, and complement receptors 3 (CD11b/CD18) and 4 (CD11c/CD18), as well as class A and B scavenger receptors. Non-integrin receptors include the PS-binding receptors, TIM-4 and BA11 mentioned earlier, the ATP-binding cassette transporter (ABC1), CD14, and the closely related Tyro 3 family receptor tyrosine kinases, c-Mer, TYRO, and Axl. CD91 (LDL receptor–related protein) is a multifunctional receptor that recognizes 30 different ligands, of which calreticulin is one. According to the “tether and tickle” model, some receptors, such as CD14 or CR3, serve as

Figure 27-5  A, Attraction and keep-out signals released by dying cells. Lactoferrin is a keep-out signal for neutrophils. LPC, lysophosphatidylcholine; S1P, sphingosine-1-phosphate. B, Receptors, ligands, and opsonins (bridging proteins) implicated in recognition or phagocytosis of apoptotic cells (see text for details). CRP, C-reactive protein; CRT, calreticulin; LyPtC, lysophosphatidylcholine; Ox, oxidized form; PS, phosphatidylserine; TSP, thrombospondin. Yellow circles are lipids or oxidized lipids.
Responding to the Dying Cell

Ingestion of apoptotic cells has significant effects on the phagocyte and potentially on the T cell response to ingested antigens. In vitro and some in vivo studies suggest that uptake of apoptotic cells by macrophages induces the expression of immunosuppressive cytokines such as transforming growth factor (TGF)-β1, prostaglandin E2, and possibly IL-10 by macrophages. These cytokines tend to dampen an immune response to self-antigens. Significantly, ligands for c-Mer, such as Gas6 and protein S on apoptotic cells, suppress production of IL-12, TNF, and interferon (IFN)-α by macrophages and/or dendritic cells. Apoptotic cells activate a specific transcriptional repressor of IL-12, GC-BP, which also suppresses adaptive immune responses. Oxidized lipids derived from the apoptotic cell membranes activate two transcription factors—peroxisome proliferator–activated receptor δ (PPARδ) and liver X receptor (LXR)—resulting in enhanced phagocytosis of apoptotic cells and suppression of inflammatory cytokines. Because some peptides derived from apoptotic cells can be presented to lymphocytes by dendritic cells and possibly by macrophages through cross-priming, questions of critical importance for studies of autoimmunity include whether self-peptides are presented after phagocytosis of apoptotic cells, and under what conditions they induce tolerance or immunity.

Ligands (DAMPs) and Sensors for Cell Debris. Engulfment of dying cells is the first step of the “clean-up” process, but swift degradation of cellular contents both within the phagocyte and extracellularly is equally important. Recently, much attention has been devoted to the self-molecules that activate (called DAMPs for damage-associated molecular patterns, as opposed to PAMPs, which are pathogen-associated molecular patterns) and the sensors that respond to products from dead and dying cells (Supplemental Figure 27-3 on www.expertconsult.com). A partial list of DAMPs includes high mobility group B1 (HMGB1), nucleic acids, the nucleotides ATP and UTP, uric acid, heat shock proteins, and SAP130. SAP130 is a U2-associated spliceosome protein, although it is not known to be a common autoantigen. As shown in Supplemental Figure 27-3 on www.expertconsult.com, the sensors for DAMPs (e.g., TLRs, retinoic acid inducible gene I [RIG-I]–like receptors [RLRs]) are, for the most part, the same as those that sense PAMPs. Therefore, the concept that self/nonself-discrimination can be attributed to differences in sensors that are specific for self–versus foreign antigens is no longer tenable. Because self-molecules can stimulate inflammatory responses, it is vital that DAMPs be efficiently processed and rendered noninflammatory. Serum contains a potent DNase, DNase I, as well as abundant RNases that serve these functions. Opsonins such as CRP scavenge nucleoproteins for rapid removal. Within the cell, a specific acid-activated DNase, DNase II, resides in lysosomes and degrades ingested nuclear DNA; multiple different endonucleases and exonucleases (see Defective Uptake and Processing of Apoptotic Cells, later) have been described. As discussed later, failure to degrade these molecules or activation of the inflammasome pathway likely explains why the debris from necrotic cells induces type 1 IFN, TNF, IL-1β, and other proinflammatory cytokines (see earlier). Additional details on receptors and sensors for each molecule can be found in recent reviews.

LABORATORY DETECTION OF APOPTOSIS

Numerous methods have been devised to detect cells undergoing apoptosis. These methods depend on biochemical changes in the cell as described previously and are depicted in Figure 27-6. Electron microscopic examination (see Figure 27-1) is still regarded as the “gold standard.”

Cell Membrane Alterations

Annexin V binds to negatively charged phospholipids in a calcium-dependent manner and therefore can readily detect the flip of phosphatidylycerine to the outer surface of the cell membrane (see Figure 27-6A). When annexin V is conjugated to fluorescein isothiocyanate (FITC), biotin, or other markers, it provides a convenient tag for detecting apoptotic cells by flow cytometry. Flow cytometry detection with annexin V is simple and sensitive and detects cells at an early stage of apoptosis. Annexin V will also bind to necrotic cell membranes before complete rupture of the cell. Entry of trypan blue, as seen by light microscopy, or propidium iodide, as quantified by flow cytometry, into the cell indicates profound damage to the cell membrane indicative of necrosis. The fluorescent chemical, 7-amino-actinomycin D (7-AAD), intercalates into double-stranded nucleic acids but will enter the cell only after membrane damage, and therefore is also a useful marker of cell death by flow cytometry.

Loss of Mitochondrial Membrane Potential (MMP)

As was discussed previously, multiple stimuli lead to apoptosis through the intrinsic mitochondrial pathway, resulting in loss of membrane potential. Several dyes, including rhodamine 123, tetramethyl rhodamine methyl ester (TMRM) (see Figure 27-6B), and DiOC6, bind relatively selectively to mitochondria, thus providing a fairly sensitive measure of mitochondrial potential.

Caspase Activation

As has been discussed, caspases are normally present in an inactive state, but once activated, they recognize specific tetrapeptide sequences (see Supplemental Figure 27-2 on www.expertconsult.com). Caspase activity therefore can be quantified directly in intact cells by flow cytometry analysis with cell-permeable fluorochrome tetrapeptide conjugates, or in cell extracts by enzyme-linked immunosorbent assay (ELISA) that detect release of colorimetric dyes conjugated to the tetrapeptides. Caspase activation can also be quantified indirectly by Western blot analysis of cleaved (activated) caspase 3, or of cleavage of specific caspase substrates.
Supplemental Figure 27-3  Sensors responsive to danger-associated molecular patterns. Sensors are shown as receptor structure, adapter proteins are shaded blue, and signaling pathways are abbreviated. Note that DNA may enter the cytoplasm from the nucleus, where exonucleases such as TREX1 degrade them. If this fails, a cytoplasmic sensor (possibly of the PYHIN type) stimulates the production of type 1 interferon. See text for details.
Cleavage of the nuclear protein, poly-(ADP-ribose) polymerase (PARP), is also used to evaluate activation of caspase 3 (see Figure 27-6C).

**Chromatin Condensation and DNA Fragmentation**

Condensation of chromatin can be seen by light microscopy following nuclear staining (see Figure 27-6D). This can be a sensitive screening method depending on the experience of the viewer, but confirmation with a more specific assay is usually required for verification. Chromatin condensation is more easily seen by staining with vital dyes such as Hoechst 33,342 bisBENZIMIDE (2′-4-ethoxyphenyl)-5-[4-methyl-1-piperazinyl]-2,5′-bi-1H-benzimidazole) or DAPI (4′,6-diamidino-2-phenylindole) (Hoechst AG, Frankfurt, Germany) and inspection under fluorescence microscopy. All of the dark blue circles represent the nuclei of apoptotic thymocytes ingested by the macrophage, whereas the macrophage nucleus is the large less-dense circle marked by the white arrowhead. F, Normal (left panel) or apoptotic (right panel) cells were permeabilized and incubated with RNase, and their DNA stained with propidium iodide, according to the method of Nicoletti and colleagues. The cells were analyzed by flow cytometry, and staining in the subdiploid peak (smaller than the G0/G1 peak and labeled M1 in the histogram) reflects the extent of apoptosis. G, Normal and apoptotic cells were lysed and the nuclei removed by centrifugation. Cytosolic extracts were then applied to agarose gels and components resolved by electrophoresis. Ethidium bromide staining of the extract made from live cells reveals high-molecular-weight DNA that remains close to the application well (left lane), whereas extract from apoptotic cells demonstrates loss of high-molecular-weight DNA and the appearance of the typical ladder of nucleosomes. H, Six-micron sections were made from a normal mouse thymus. The cells were permeabilized and incubated with biotinylated deoxyuridine triphosphate (dUTP) in the presence of terminal deoxynucleotidyl transferase (TdT). Nicked DNA incorporated the labeled deoxyuridine and was detected by staining with peroxidase-labeled streptavidin and substrate (dark color). I, A breast cancer cell line was induced to undergo apoptosis. Autophagic vesicles were detected with an anti-LC3 antibody by indirect immunofluorescence. (Figure kindly provided by Bassam Jarjri, Laboratory of Hemato-Oncology, Public Research Center for Health, Luxembourg, Belgium.)

As has been mentioned, DNA is cleaved by multiple DNases, leading to the cleavage of nucleosomes at the linker region between histone bindings, yielding the characteristic 180 base pair “DNA ladder” (see Figure 27-6G). DNA fragmentation results in free 3′-OH groups, which can be detected within the nuclei in tissue sections using biotinylated deoxyriboonucleotide triphosphates (dNTPs) (terminal deoxynucleotidyl transferase 2′-deoxyuridine, 5′-triphosphate [dUTP] nick end-labeling [TUNEL] assay; see Figure 27-6H). Whereas formation of the ladder is specific to apoptosis, generation of free ends of DNA is not, and may also be detected in DNA damaged by necrosis. Furthermore, it has been reported that TUNEL and in situ DNA incorporation methods may yield positive results in cells undergoing extensive DNA repair or rapid proliferation.

**Autophagy**

As was discussed, during autophagy, lipidation of LC3 to form LC3-II can be identified as coarse dots on the autophagosomes by immunofluorescence microscopy. Figure 27-6I demonstrates such staining on a breast cancer cell line induced to undergo autophagic cell death.

**APOPTOSIS IN RELATION TO RHEUMATIC DISORDERS**

The regulation of apoptosis is highly relevant to the pathogenesis and treatment of rheumatic disorders. Pertinent examples are discussed in the following sections and are reviewed in Nagata et al. Mice with mutations of Fas or Fas ligand (FasL) develop a syndrome characterized by lymphoproliferation (lpr) and generalized lymphadenopathy (gld), together with systemic autoimmunity. As might be expected from the key role described for Fas in AICD, lymphadenopathy and spleenomegaly are the consequences of failure of activated lymphocytes to die, resulting in an absolute increase in the numbers of T and B lymphocytes, as well as of the accumulation of an unusual subset of T cells that do not express CD4 or CD8 co-receptors (i.e., double-negative T cells). The nature and extent of systemic autoimmunity vary according to the strain into which the Fas or FasL mutation has been bred.
A syndrome of massive lymphadenopathy with systemic autoimmunity in children was reported by Canale and Smith in 1967. Subsequently, these and other lpr patients were found to have mutations in Fas.\textsuperscript{92–94} The syndrome (called Canale-Smith syndrome [CSS] or autoimmune lymphoproliferative syndrome [ALPS]) is characterized by lymphadenopathy or splenomegaly, autoimmune cytopenias (most commonly affecting platelets and red cells), and an increase (>5%) in circulating double-negative T cells. Patients had defective lymphocyte apoptosis in response to anti-Fas antibodies or FasL when tested in vitro, and most had heterozygous mutations in Fas affecting the death domain. These mutations impair Fas-mediated apoptosis through a dominant negative effect.\textsuperscript{95,96} Although Fas/FasL mutations are an exceptionally rare contributory factor to systemic lupus erythematosus (SLE), ALPS is informative because it suggests that defective apoptosis of cells of the immune system can cause systemic autoimmune diseases such as idiopathic thrombocytopenic purpura, autoimmune hemolytic anemia, and Guillain-Barré syndrome. It illustrates (as do the mouse models) that even when a single gene has a powerful effect on predisposition to systemic autoimmunity, the clinical expression of disease depends on the precise nature of the mutation\textsuperscript{95,96} and the interaction with modifying genes. Recently, somatic mutations in Fas have been detected in cases where patients present at an older age and do not have germline mutations.\textsuperscript{97}

Several other examples of genetic alterations in death or survival genes may lead to lupus-like diseases in mice (reviewed in Kim et al\textsuperscript{98}). Of particular interest is overexpression of the ligand for the BlyS/BAFF receptor, which promotes the survival of B lymphocytes,\textsuperscript{99} because increased expression of this ligand has been reported in SLE, SJögren’s syndrome, and rheumatoid arthritis.\textsuperscript{100} Mutations in the p55/TNFR1/CD120a receptor in humans results in a periodic autoinflammatory syndrome called tumor necrosis factor receptor–associated periodic syndrome (TRAPS). Mutations predominantly occur in the first two CRDs of the receptor, resulting in reduced shedding of the extracellular domain of the receptor and reduced neutralization of circulating TNF.\textsuperscript{101} Increased lymphocyte survival due to overexpression of Bcl-2, knockout of Bim, reduction in PTEN activity,\textsuperscript{102} and increased survival of dendritic cells\textsuperscript{103,104} causes lupus-like autoimmunity in mice.

In summary, defective apoptosis of B or T cells may lead to inappropriate survival of self-reactive cells in the central (thymus or bone marrow) or peripheral immune system. Enhanced survival of dendritic cells may promote the activation and expansion of low affinity self-reactive T cells.

**Defective Uptake and Processing of Apoptotic Cells**

Autoantibodies were discovered in the 1940s and 1950s, and their molecular and functional identities were characterized in the 1980s and 1990s. Why the immune system targets a select subset of self-antigens (mainly nucleoproteins) in each disease has never been satisfactorily explained. The fact that autoantibodies target nucleosomes in SLE, while certain antcardiolipin antibodies cross-react with PS, which translocates to the cell surface during apoptosis,\textsuperscript{105} supports the idea that autoantibodies target the products of apoptotic cells. Additional inferential evidence for this hypothesis comes from detection of lupus antigens in apoptotic blebs,\textsuperscript{106} modification of antigens by cleavage, and phosphorylation during apoptosis.\textsuperscript{107,108} Apoptotic blebs or microparticles may be released from dying cells; this has important consequences for immunoregulation.\textsuperscript{109}

Because apoptosis occurs on a vast scale in the central lymphoid organs and should tolerate the host, under what conditions would apoptotic cells immunize? If apoptosis occurs in the presence of an adjuvant agent (e.g., virus, bacterium, chemical agent), tolerance may be lost at least transiently. An interesting paradigm for the generation of the proinflammatory cytokine, IL-17, has been suggested in experimental animals in which TGF-β may be generated by apoptotic cells, together with IL-6 induced by certain microbes, thus promoting Th17 induction.\textsuperscript{110} Abnormalities leading to accelerated apoptosis of cells or reduced uptake of dying cells will allow cells to undergo postapoptotic necrosis; this, in turn, will provoke a proinflammatory cytokine response from phagocytes, as was explained previously. Indeed, an increase in the rate of apoptosis of SLE peripheral blood mononuclear cells has been observed,\textsuperscript{111,112} and it is suspected, but not proven, that ultraviolet exposure to the skin may induce excessive apoptosis in SLE patients. Reduced macrophage phagocytosis of apoptotic cells has been reported in SLE and increased apoptotic cell debris observed in germinal centers of lymph nodes obtained from SLE patients.\textsuperscript{113,114} This is of importance because B cells are positively selected in germinal centers, so an increase in self-antigen may enable selection of autoreactive B cells generated by somatic hypermutation.

It is well known that deficiencies of early complement components predispose to human SLE.\textsuperscript{115} Two overlapping explanations for this striking association are that phagocytosis of apoptotic cells is impaired in the absence of early complement components,\textsuperscript{62,72} and that C1q suppresses the induction of IFN-α by nucleoprotein-containing immune complexes.\textsuperscript{116,117} Knockout of members of the Tyro 3 receptor tyrosine kinases is associated with defective clearance of apoptotic cells and expression of a lupus-like disease,\textsuperscript{118,119} which may also be in part related to regulation of type I IFN.\textsuperscript{114} Sub and associates\textsuperscript{120} observed low levels of protein S, one of the ligands for the TAM receptors, in SLE patients. Mammalian nucleic acids can potently stimulate TLR-dependent and -independent pathways to generate inflammatory cytokines. Deficiency of DNase I led to lupus, and conditional deficiency of DNase II caused an RA-like disease in mice.\textsuperscript{121,122} Failure to degrade DNA or DNA:RNA intermediates in the cytosol results in activation of type I IFN in the brain (humans—Aicardi-Goutières syndrome) or myocarditis in mice.\textsuperscript{123} Approximately 2% of SLE patients have heterozygous mutations in TREX1.\textsuperscript{124} Nucleic acids contained within immune complexes have been shown to stimulate IFN-α by plasmacytoid dendritic cells (pDCs), providing a plausible explanation for increased IFN-α observed in SLE and amplification of disease activity.\textsuperscript{125} Some recent studies suggest that neutrophils may be a source of antigens, and that impaired degradation of neutrophil nets (neutrophil extrusions of DNA/histone complexes) may predispose to nephritis.\textsuperscript{126}


**Prolonged Exposure to Growth Factors**

Histologically, RA is characterized by an accumulation of inflammatory cells in the synovium, leading to pannus formation and destruction of cartilage and bone. Although Fas and FasL can be detected in RA synovium, and evidence of apoptosis has been detected in RA synoviocytes, the extent of synoviocyte apoptosis is not adequate to counteract ongoing proliferation. This imbalance is explained by a number of factors. First, FasL is expressed at relatively low levels on synovial T cells, and soluble Fas or FasL competitors may impede Fas-induced apoptosis. Cytokines such as TNF, IL-1β, and TGF-β1, which are overexpressed in the joints of patients with RA, favor synoviocyte proliferation and inhibit susceptibility to apoptosis. These and other signals reduce apoptosis of synoviocytes through activation of NFκB and increased expression of antiapoptotic proteins, including X-IAP and Akt. Growth of the pannus is compounded by inflammatory changes such as oxidation that result in upregulation and mutations of the growth suppressor protein p53, as well as IL-23, which promotes or stabilizes IL-17–producing Th cells (Figure 27-7). IL-23 has been shown to promote cell survival.

Observations regarding apoptosis regulators in RA are significant because they provide an opportunity for therapeutic manipulation. Local administration of anti-Fas monoclonal antibodies to human T cell lymphotropic virus-1 tax transgenic mice, or Fas ligand to collagen arthritis mouse models of RA, led to an improvement in arthritis. Several strategies used to modulate NFκB attenuate the growth of synovial cells. Administration of TRAIL also attenuated experimental arthritis, although this was not thought to be due to apoptosis, and an antibody to TRAIL-R2 (DR5) induced apoptosis of RA synovial fibroblasts. Accumulating evidence suggests that fibroblasts in scleroderma may also be more resistant to apoptosis, and that TGF-β may promote this phenotype.

**Tissue Injury in Organ-Specific Autoimmunity**

In contrast to systemic autoimmune diseases characterized by B lymphocyte stimulation, leading to antibody- and immune complex–mediated tissue injury, many organ-specific autoimmune diseases are caused by a cell-mediated attack, leading to the death of specific cell types within the organ. Cell targets are β cells of the islets of Langerhans of the pancreas in insulin-dependent diabetes mellitus, oligodendrocytes in the brain in multiple sclerosis, salivary and lacrimal glands in Sjögren’s syndrome, and myocytes in polymyositis. Programmed pathways of cell death (apoptosis) can be implicated in the pathogenesis of some of these diseases, as is illustrated by the resistance of Fas-deficient (lpr) mice to diseases such as diabetes and experimental encephalomyelitis. In many of these diseases, cell death at the site of injury can be directly demonstrated by DNA fragmentation (TUNEL staining) in situ. Apoptosis is usually considered noninflammatory, but, as has been discussed, the context of cell death influences the immune response. For example, cytotoxic lymphocytes (CTLs) that induce cell death predominantly by perforin-mediated cell lysis (necrosis, although apoptotic changes are also observed through caspase 3 activation) or macrophage defects leading to delayed clearance of dying cells will promote the release of proinflammatory cytokines such as TNF. Similarly, death by pyroptosis or necroptosis releases inflammatory cytokines.

In most organ-specific autoimmune diseases, especially those for which adoptive transfers have been performed in animal models, CD4+ T cells have been shown to be critically involved in disease pathogenesis. Disease-promoting CD4+ T cells are restricted by major histocompatibility complex class II molecules, and therefore are unlikely to exert direct cytotoxic action on the class I–bearing target cell (although CD4+ T cells can upregulate FasL). CD4+ T cells may arm other effectors through the production of

---

**Figure 27-7** Antia apoptotic phenotype of rheumatoid synovial fibroblasts. Cytokines and growth factors produced by macrophages and T cells lead to activation of nuclear factor kappa B (NFκB) and upregulation of antiapoptotic proteins such as B cell lymphoma-2 (Bcl-2). Inflammatory stimuli, release of nitric oxide (NO), and reactive oxygen intermediates (O*) upregulate and induce mutations of p53. Infiltrating lymphocytes have a Fas ligand “low” phenotype.
cytokines (IFN-γ); they may induce tissue injury through a “bystander pathway” involving macrophages, or may induce receptors for cell death on the target cell “assisted suicide pathway.” In Sjögren’s syndrome in humans, controversy continues regarding whether Fas or FasL is constitutively expressed in normal salivary glands, but co-expression of these molecules in patients with Sjögren’s syndrome presumably causes cell death of acinar and ductal cells.143

Inflammatory myopathies, such as polymyositis (PM) and dermatomyositis (DM), are autoimmune diseases that result in destruction of skeletal muscle fibers. Although Fas is upregulated on myocytes in these diseases, expression is also increased in nonautoimmune muscle disorders such as metabolic myopathies, denervating disorders, and muscular dystrophies, but not in normal human muscle tissue.144 Detection of FasL on mononuclear cells invading the muscles in PM and DM patients with apoptosis of muscle cells implicates Fas/FasL in tissue injury in myositis.145 Increased expression of the T cell cytotoxic mediator perforin in some PM and DM patients146 indicates that granzyme-mediated myocyte injury is also involved. The tRNA synthetase antigens or their cleavage products may perpetuate inflammation by exerting chemotactic recruitment of immune cells through chemokine receptors.147

Accelerated Apoptosis in Degenerative Rheumatic Disorders

Apoptosis of chondrocytes occurs during normal development of joints, and accelerated cell death may be important in diseases such as osteoarthritis (OA). The main mechanism underlying primary or secondary osteoarthritis is degradation of cartilage. Degradation is mediated by enzymatic and nitric oxide (NO)–induced breakdown of the extracellular matrix and insufficient new matrix synthesis. Normal and OA-derived chondrocytes in the superficial and middle cartilage zones—the major areas involved in early cartilage degeneration—express Fas and are sensitive to Fas-mediated death.148 Chondrocytes obtained from patients with OA have enhanced spontaneous apoptosis in these zones compared with normal controls.149,150 Although NO is also capable of inducing apoptosis in chondrocytes, it does not seem to act through the Fas pathway.148 In an experimental model of OA, transgenic mice lacking type II collagen, the main constituent of the extracellular matrix in cartilage, had high levels of apoptosis in their chondrocytes.151 Together, these findings suggest that apoptosis of chondrocytes plays a role in OA, and that inhibitors of NO synthesis may be of value in treating this disease (see reference 152 for review). The therapeutic use of intra-articular Fas agonists in RA may be deleterious to chondrocytes, whereas ADAM15 exerted a protective effect on chondrocyte apoptosis.153

Osteoporosis is a common disorder resulting from increased bone resorption, decreased bone synthesis, or a combination of the two. Several reports support the concept that estrogen exerts its beneficial effect in preventing osteoporosis through induction of apoptosis in bone-resorbing osteoclasts,154,155 and glucocorticoid-induced osteoporosis may be explained by an increased rate of apoptosis of osteoblasts and osteocytes.156 In the presence of macrophage colony-stimulating factor (M-CSF), osteoclasts differentiate from a myeloid precursor common to macrophages and dendritic cells. In addition to numerous factors influencing bone turnover,157 a soluble member of the TNFR family, osteoprotegerin (OPG) or osteoclastogenesis-inhibitory factor (OCIF), inhibits osteoclast activity after binding to its cognate receptor activator of NFκB (RANK) ligand (OPG ligand/TRANCE).158,159 RANK ligand is expressed on osteoblasts and on activated T cells. Engagement of the membrane form of the receptor induces activation of NFκB, thereby enhancing the formation, survival, and resorptive activity of osteoclasts.

Drugs That Affect Apoptotic Pathways

Up until very recently, therapy for inflammatory rheumatic disorders has been largely empiric. The types of drugs used include anti-inflammatory agents such as corticosteroids and nonsteroidal anti-inflammatory drugs such as cyclosporine, and cytotoxic drugs such as cyclophosphamide and azathioprine. Because most of these drugs impinge on critical biochemical events within the cell, it is not surprising that they have effects on pathways of apoptosis.

Anti-inflammatory Drugs

Glucocorticoids at high doses induce the death of lymphoid cells through transcriptional regulation by the glucocorticoid receptor. Corticosteroids modulate expression of a large number of molecules that affect apoptotic programs—cytokines, cell cycle control proteins, c-myc, Bcl-2—and inhibit NFκB activation, but the precise pathways that may operate in a cell-specific fashion160 and that are relevant to clinical efficacy remain to be defined. Patients on long-term steroid therapy are susceptible to osteoporosis and osteonecrosis, which may be explained by bone loss caused by apoptosis of osteoblasts and osteocytes.156

The major mechanism of action of NSAIDs consists of inhibition of cyclooxygenases (COXs), which reduce the production of proinflammatory cytokines and prostaglandins (see Chapter 59). NSAIDs also are effective in the chemoprevention of colorectal tumors in genetically susceptible individuals. Their antineoplastic properties may be explained by an increase in the prostaglandin precursor arachidonic acid, and by conversion of sphingomyelin to ceramide, a proapoptotic lipid.161,162

Immunomodulatory Drugs

Cyclosporine and the closely related macrolide antibiotics, FK506 (tacrolimus) and rapamycin (sirolimus), have potent immunosuppressive properties and are used to prevent allograft rejection. These drugs modulate T and B cell immune responses by interfering with nuclear factor of activated T cell (NFAT)-mediated IL-2 gene transcription, NO synthase activation, cell degranulation, and apoptosis.163 The reduced cytotoxic T lymphocyte activity is explained in part by impaired FasL induction secondary to the effect on NFAT,164,165 but effects on mitochondrial function have also been demonstrated. Certain cell types such as renal proximal tubules and synoviocytes or endothelial cells in RA may be more susceptible to the proapoptotic effects
of cyclosporine. Rapamycin appears to be an effective and well-tolerated treatment for ALPS patients with glucocorticoid-resistant disease.

**Cytotoxic Drugs**

Many cytotoxic or immunosuppressive drugs that induce the suicide of lymphocytes, mostly through the p53 pathway (see Figure 27-4), exert some anti-inflammatory effect. Although methotrexate has effects on adenosine receptors, even low-dose methotrexate induces apoptosis of activated lymphocytes in vitro and in RA patients, probably in a Fas-independent manner. Cyclophosphamide is an alkylating agent commonly used to treat many human cancers and severe autoimmune disease. Its efficacy has been attributed in part to apoptosis of tumor cells and perhaps of mesangial cells in glomerulonephritis. Induction of apoptosis may also account for certain adverse effects, such as oligospermia or azoospermia and pancreatic β-cell destruction.

Bisphosphonates are the most potent antiresorptive drugs available and are widely used to treat various metabolic bone diseases, such as Paget’s disease, bone tumor, ectopic calcification, and osteoporosis. Although individual members of this family differ somewhat in their effects, their general mechanisms of action include direct and indirect effects on osteoclast recruitment, function, and survival.

**Biologics**

The remarkable success of anti-TNF therapy for the treatment of RA, other arthritides, and Crohn’s disease is generally attributed to blockade of TNF stimulation of the proinflammatory NFκB pathway (see Figure 27-4). However, in Crohn’s disease, it has been reported that anti-TNF monoclonal antibodies ameliorate disease by binding to cell-associated TNF and inducing apoptosis of macrophages and T cells. Both etanercept and infliximab induce apoptosis of monocytes/macrophages in RA synovial tissue. A monoclonal antibody that blocks RANK ligand (denosumab) has been approved for the treatment of osteoporosis.

B cell depletion therapy with antibodies against B cell surface proteins such as CD20 (rituximab) and CD22 (epratuzumab) are being used increasingly for therapy. The mechanisms of action of anti-CD20 have been studied most intensively in chronic lymphocytic leukemia (CLL) B cells. Rituximab depletes B cells by induction of apoptosis associated with downmodulation of Bcl-2 and X-IAP, by activation of complement, and by antibody-dependent cell-mediated cytotoxicity (ADCC).

Belimumab, a human monoclonal antibody that attenuates the activity of BLyS on B cell survival, has met its phase III objectives in clinical trials in SLE patients (see also Chapters 63 and 64).

**OTHER OPPORTUNITIES FOR THERAPEUTIC INTERVENTION**

Understanding the biochemical pathways that regulate apoptosis offers new opportunities for therapeutic intervention, many examples of which are given in the preceding section (Figure 27-8). In antibody-mediated diseases, induction of apoptosis of B cells is effective. In cell-mediated diseases such as Sjögren’s syndrome and polymyositis, it may be beneficial to induce apoptosis of the cytotoxic effector cell. In diseases characterized by macrophage activation and inflammatory tissue growth, induction of macrophage apoptosis by anti-TNF reagents has already been shown to be effective in RA and in Crohn’s disease (see Figure 27-8).

If a death receptor is selectively expressed on the cell to be killed, a death ligand could be administered. Examples of such a strategy include the use of Fas agonists or anti-TRAIL receptor antibodies in arthritis. Proapoptotic pathways could also be initiated from within the cell by peptide mimetics of Smac/Diablo (see Supplemental Figure 27-2 on www.expertconsult.com) or through gene therapy approaches. Examples include blockade of NFκB and

Figure 27-8 Avenues for therapeutic manipulation of apoptosis: biologics in practice or clinical trials. Anti-tumor necrosis factor (TNF) reagents work, in part, by engaging membrane TNF (mTNF) on activated macrophages and inducing apoptosis. Anti-CD20 antibodies eliminate B cells, in part by inducing apoptosis. Anti-BAFF/BLyS reagents reduce B cell survival. Antibodies to receptor activator of NFκB ligand (RANKL) reduce differentiation or survival of osteoclasts. Estrogens promote osteoclast apoptosis, and bisphosphonates prevent it. Other approaches to modulation of apoptotic pathways are discussed in the text.
overexpression of Bax. Similarly, as was discussed earlier, a cytokine or an unwanted growth/differentiation promoter such as TNF in RA or OPG/RANK ligand in osteoporosis can be blocked by a monoclonal antibody or a soluble receptor fusion protein (see Figure 27-8).

In diseases in which apoptotic cell death leads to loss of organ function, the death ligand could be blocked by a monoclonal antibody or a soluble receptor fusion protein. Even when the death pathway has not been fully determined, attempts can be made to interfere with upstream components of apoptosis well before the “point of no return.” Antiapoptotic genes with limited (e.g., the protein FLIP blocks the Fas pathway) or broad (e.g., Bcl-2 family) specificity can be introduced into the cell. Further downstream, cell-permeable caspase inhibitors can block the execution phase of apoptosis in vivo, as has been illustrated experimentally. All of these approaches are feasible, but they are limited by their potential adverse effects. Therapy must be relatively specific for the target cell, because widespread prevention of cell death for sustained periods is likely to predispose to neoplasia.

Because apoptotic cells themselves induce immunosuppression, infusion of apoptotic cells or natural antibodies that bind to apoptotic cells (see Figure 27-5) has been shown to attenuate experimental arthritis.

**Conclusions**

Appreciation that death and survival of cells are highly regulated and dissection of the biochemical pathways activated by different modes of intracellular stress have made an enormous contribution to our understanding of the pathophysiology of human disease. Inherited mutations of apoptosis regulatory molecules and of the genes encoding sensors may cause systemic autoimmunity, dysregulation, or misdirection of other cell death, or survival molecules contribute to a whole range of musculoskeletal disorders. Many of the drugs used to treat musculoskeletal disorders exert potent effects on apoptotic programs. New biologic therapies that induce or prevent apoptosis of selected targets are already in use in rheumatic diseases, and it is likely that many more, perhaps with greater selectivity, will be developed. All of these findings suggest that improved understanding of the regulation of apoptosis will continue to have great impact for the pathogenesis and therapeutic manipulation of rheumatic disorders.

**Acknowledgments**

Help and discussion from current and past laboratory members are gratefully acknowledged.

**Selected References**


References

CHAPTER 27 | CELL SURVIVAL AND DEATH IN RHEUMATIC DISEASES


PART 3


Experimental Models for Rheumatoid Arthritis

RIKARD HOLMDAHL

KEY POINTS

Animal models are tools to mimic various aspects of rheumatoid arthritis (RA).

There are many different animal models for RA, including collagen-induced arthritis (CIA), collagen antibody–induced arthritis (CAIA), and adjuvant (pristane)-induced arthritis (PIA).

Arthritis can be induced in animals by immunization with cartilage components, nonspecific immune stimuli, bacterial or viral components, or genetic modification.

Animal models have a defined onset and are useful for the kinetic evaluation of arthritis, cell and mediator involvement, and detailed analysis of joint erosion.

Animal models provide direction for novel approaches to treatment such as cytokine inhibition.

For a deeper understanding of the complexity of the pathogenesis of rheumatoid arthritis (RA), the use of animal models is a necessity. Obviously a disease identical to RA cannot develop in experimental animals because they are different species with different genetics and environments, as compared with humans.

The advantages of using animal models are mainly the following:

1. Genetically controlling inbred animal strains.
2. Controlling their environment better than for humans.
3. Manipulating experiments. Researchers can change the genome of inbred strains by mutations, insertions, and deletions. They can also change the environment in a controlled way, such as immunizing or infecting animals, which may lead to arthritis. Controlled experiments can be performed.
4. Using animals is more ethical than using humans.

For animal models to be useful, they need to recapitulate some of the key features of RA, such as:

• The disease starts before the clinical diagnosis. An autoimmune and inflammatory process precedes the clinical onset by several years.
• Tissue specificity. RA is characterized by a tissue-specific inflammatory attack affecting diarthrodial, peripheral, and cartilaginous joints. Although systemic immune responses and manifestations are usually present, the inflammation is mainly directed toward peripheral joints.
• Chronicity. The disease is chronic and occurs in tissues in which no causative infectious pathogens have so far been demonstrated. Acute joint affections are common manifestations in both physiologic responses to infections and in connection with other inflammatory disorders, but in RA chronicity is an essential characteristic. The disease course may proceed with identifiable relapses, but there is usually a steady progression of joint destruction.
• Autoantibodies. The development of RA is preceded and associated with elevated levels of autoantibodies in serum. Antibodies to citrullinated protein (ACPA) have the highest specificity and sensitivity followed by antibodies to immunoglobulin (rheumatoid factors), but antibodies to other antigens also occur in subsets of patients such as antibodies to type II collagen (CII) and hnRNP-A2.
• Major histocompatibility complex (MHC) class II association. The genetic influence is significant but complex. Class II genes in the major histocompatibility complex make the largest genetic contribution by far. In particular, certain structures in the peptide-binding pocket of HLA-DR4 molecules are highly associated with RA. Several other loci confirm that involvement of adaptive immunity (PTPN22, CTLA4, IL-21) strengthens the view that RA is an autoimmune disease.

Taken together, these findings suggest that both activation of innate immunity and immune-mediated inflammation directed to peripheral joints play a role in the disease process. Because the disease process starts several years before clinical onset with enhanced levels of autoantibodies (ACPA and rheumatoid factor [RF]) together with a higher level of inflammation markers, it is likely that the etiologic factors are operating at this early time. Smoking and various chronic infections such as periodontitis have been suggested to be associated with the early disease process. This process is, however, not joint specific and a further spreading of the autoimmune reactivities toward joint specificities is likely to occur before onset of arthritis. One possible explanation for such a response is an occurrence of an infectious agent persisting in the joint such an agent has not been identified as an explanation for RA, although infectious agents have been shown to induce and promote arthritis. Alternatively, the immune reaction could be directed to molecular targets in the joints, in cartilage, or in the synovial tissue. Another explanation could relate to a defect in a gene related to peripheral joints (e.g., leading to cartilage fragility or a gene affecting immune recognition). However, such a genetic defect has not been found. Instead, genome-wide association studies suggest that the MHC class II region contains the most important genes supplemented with a large number of genes outside the MHC region, most of which are associated with adaptive immune responses. Thus the cause and
driving forces are polygenic and multifactorial, and the understanding of the disease will require a detailed basic analysis of disease mechanisms. Animal models are excellent tools for this type of analysis. Recent advancement of different animal models mimicking different aspects of human diseases, as well as the improvement in genetic techniques, has dramatically increased their usefulness. The present overview includes not only models for RA but also briefly adds models with related disease pathways such as psoriasis arthritis, reactive arthritis, ankylosing spondylitis, Lyme disease, and septic arthritis (Table 28-1). However, there is a focus on the classical models for RA that are currently commonly used in both industry and academia: the adjuvant arthritis models in the rat, the collagen-induced arthritis (CIA), and the collagen antibody–induced arthritis (CAIA).

**ARTHРИTIS CAUSED BY INFECTIOUS AGENTS**

Several infectious agents can invade joints, persist there, and cause arthritis. As with most persisting infectious agents, a balance between the parasite and the host is usually achieved. Thus inflammatory consequences may not only be caused directly by the parasite but also by an aberrant inflammatory response of the host. When microorganisms are present in the target tissue, chronic autoimmunity could be maintained by different mechanisms such as superantigen-mediated T cell activation, a cross-reactive immune response, or the presence of adjuvant material enhancing autoantigen presentation. Several such arthritogenic agents have been described in experimental animals, and some of these mimic a corresponding infectious disease in humans.

**Mycoplasma arthritides Arthritis**

Arthritis associated with mycoplasma infection is endemic among farm animals. It is also possible to induce arthritis in rodents after inoculation with *Mycoplasma arthritides*. However, mycoplasma bacteria are not easily found in RA joints, although they may cause arthritis in individuals with severe B cell deficiency. Inoculation of mice induces a mild chronic arthritis in conjunction with the persistence of the microorganism. In accordance with the observations in humans, B cell depleted mice are more susceptible to mycoplasma-induced arthritis.

**Lyme Arthritis**

*Borrelia* is a spirochete that may persist in joints and cause arthritis. The clinical picture is chronic and resembles RA, and it is genetically associated with MHC class II-DR4, as is RA. Clearly, live bacteria persist in the joints but in many patients it has been difficult to identify the spirochete in the arthritic joints. Mice infected with *Borrelia* develop arthritis similar to the human disease. As in humans, MHC controls the susceptibility to arthritis and the immune response associated with human MHC class II expressed in mice has been shown to be directed to *Borrelia*-derived antigens. The persistence of the spirochete seems to be a requirement for the development of the arthritis, although some mouse strains do not develop arthritis in spite of high levels of bacteria in the joints.

**Staphylococcal Arthritis**

Septic arthritis is most commonly caused by a persistent infection of *Staphylococcus aureus*. The bacteria tend to be encapsulated in tissues including joint synovia and persist for years. Inoculation with certain *S. aureus* strains induces septic arthritis in many mouse and rat strains. Severe and prolonged arthritis develops in infected joints mimicking the human situation. Interestingly, protection of the host is critically dependent on the innate defense such as neutrophils and complement, whereas the adapted immune response is not effective. Instead, the apparently aberrant adapted immune response promotes arthritis.

**Arthritis and Ankylosing Spondylitis Induced by Intracellular Bacteria**

Some bacteria with the capacity to invade cells on infection (e.g., *Yersinia*) are known to be related to postinfectious arthritides such as reactive arthritis and ankyllosing spondylitis. These diseases are genetically associated with HLA-B27, a MHC class I allele of the B locus. It has been possible to reproduce the human disease to a large extent in HLA-B27 transgenic mice and rats. In B27-transgenic rats, ankylosing spondylitis, balanitis, colitis, dermatitis, and arthritis occur spontaneously. However, if the rats are made germ free, the joint manifestations are no longer present, indicating the importance of a so far unknown infectious agent. A similar phenomenon has been shown to occur in B27 transgenic mice, in which arthritis occurs only in conventional animal facilities.

**Arthritis Caused by Bacterial Fragments**

Postinfectious arthritis can develop after bacterial infections. The occurrence of arthritis can be dependent on several different bacteria-derived compounds such as cell wall fragments, DNA, and heat shock proteins. Bacterial cell wall fragments are difficult to degrade and may cause prolonged activation of macrophages and synovial macrophages. The first animal model for RA to be described was the so-called adjuvant arthritis (mycobacteria cell wall–induced arthritis, MCWIA) induced in rats after injection of mycobacterium cell walls suspended in mineral oil (i.e., complete Freund’s adjuvant [CFA]). Only rats (and not mice or primates) develop arthritis after mycobacterium challenge, although it has been reported that joint-related granuloma formation has occurred in humans treated with mycobacterium-containing vaccine. CFA is a potent adjuvant that activates a multitude of pattern recognition receptors, activating antigen-presenting cells enhancing T cell immunity. Subcutaneous injection of CFA in rats leads to granulomatous inflammation in many organs (e.g., the spleen, liver, bone marrow, skin, and eyes) and causes profound inflammation in peripheral joints. MCWIA is severe but self-limited, and the inflammation subsides after 5 to 7 weeks. The mycobacterium cell wall fragments are most likely disseminated throughout the body and engulfed by tissue macrophages, which have difficulties in...
Table 28-1  Overview of Animal Arthritis Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Species</th>
<th>Genetics</th>
<th>Disease Characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arthritis Caused by Infection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycoplasma-induced arthritis</td>
<td>Rats and mice</td>
<td>More pronounced in B cell-deficient mice</td>
<td>Mild chronic arthritis</td>
<td>2, 3</td>
</tr>
<tr>
<td>Borrelia-induced arthritis</td>
<td>Mice</td>
<td>MHC</td>
<td>Severe and erosive arthritis with spirochetes in the joints</td>
<td>4, 5</td>
</tr>
<tr>
<td>Staphylococcus-induced arthritis</td>
<td>Rats and mice</td>
<td>MHC</td>
<td>Severe arthritis</td>
<td>7, 8</td>
</tr>
<tr>
<td><strong>Arthritis Caused by Bacterial Fragments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium-induced arthritis (MCWIA)</td>
<td>Rats</td>
<td>MHC, non-MHC genes (LEW &gt; F344)</td>
<td>Acute and generalized inflammatory disease including erosive arthritis</td>
<td>15</td>
</tr>
<tr>
<td>Streptococcal cell wall--induced arthritis (SCWIA)</td>
<td>Mice and rats</td>
<td>Non-MHC genes (LEW &gt; F344), (DBA/1 = Balb/c &gt; B10)</td>
<td>30, 31</td>
<td></td>
</tr>
<tr>
<td><strong>Arthritis Induced by Adjuvant Stimulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mineral oil--induced arthritis (OIA)</td>
<td>DA rats</td>
<td>non-MHC loci on chromosomes 4, 10</td>
<td>Acute and self-limited inflammation of peripheral joints</td>
<td>33, 48</td>
</tr>
<tr>
<td>Pristane-induced arthritis (PIA)</td>
<td>Rats</td>
<td>MHC, non-MHC loci on chromosomes 1, 4, 6, 12, 14 MHC (q, d)? Balb/c, DBA and C3H gene backgrounds</td>
<td>Chronic and erosive arthritis in peripheral joints</td>
<td>35, 47, 50</td>
</tr>
<tr>
<td>Pristane-induced arthritis (PIA)</td>
<td>Mice</td>
<td>MHC (q and r), non-MHC loci on chromosomes 1, 2, 3, 6, 7, 8, 10, 15 MHC (q, u)</td>
<td>Chronic and generalized inflammatory disease also affecting joints</td>
<td>53, 140</td>
</tr>
<tr>
<td>Unmethylated DNA--induced arthritis</td>
<td>Mice</td>
<td>DBA/1</td>
<td>Mild arthritis after intra-articular injection of unmethylated DNA or CpG oligonucleotides</td>
<td>26</td>
</tr>
<tr>
<td><strong>Arthritis Induced by Cartilage Protein Immunization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CII (heterologous or homologous CII in mineral oil)--induced arthritis (CIA)</td>
<td>Rats</td>
<td>MHC (a, l, f, and u), non-MHC loci on chromosomes 1, 4, 7, 10</td>
<td>Chronic and erosive arthritis in peripheral joints.</td>
<td>64, 80, 150</td>
</tr>
<tr>
<td>CII (heterologous or homologous CII in CFA)--induced arthritis (CIA)</td>
<td>Mice</td>
<td>MHC (q and r), non-MHC loci on chromosomes 1, 2, 3, 6, 7, 8, 10</td>
<td>Erosive arthritis in peripheral joints</td>
<td>65, 77, 79, 81-84</td>
</tr>
<tr>
<td>CXI (rat CXI in IFA)--induced arthritis</td>
<td>Rats</td>
<td>MHC (q, u)</td>
<td>Severe, chronic arthritis</td>
<td>102</td>
</tr>
<tr>
<td>Human proteoglycan (in CFA)--induced arthritis</td>
<td>BALB/c mice</td>
<td>MHC (d), several non-MHC loci</td>
<td>Chronic arthritis</td>
<td>103</td>
</tr>
<tr>
<td>COMP (in mineral oil)--induced arthritis</td>
<td>Rats</td>
<td>MHC (u)</td>
<td>Acute arthritis</td>
<td>63</td>
</tr>
<tr>
<td><strong>Passively Induced Arthritis Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collagen antibody--induced arthritis (CAIA)</td>
<td>Mice, rats</td>
<td>Balb/c, DBA/1, C57Bl6</td>
<td>Acute self-limited arthritis</td>
<td>74</td>
</tr>
<tr>
<td>K/BxN serum--induced arthritis (SIA)</td>
<td>Mice</td>
<td>Balb/c, C57Bl6</td>
<td>Acute self-limited arthritis</td>
<td>131</td>
</tr>
<tr>
<td>Fibroblast transferred SCID mouse</td>
<td>Mice</td>
<td>Local injection of fibroblasts into immunodeficient SCID mouse</td>
<td>Sustained destructive arthritis</td>
<td>137</td>
</tr>
<tr>
<td><strong>“Spontaneous” Arthritis Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-B27 transgenic animals</td>
<td>Mice and rats</td>
<td>B27 heavy chain transgene</td>
<td>Ankylosing spondylitis, colitis, balanitis, arthritis</td>
<td>12, 14</td>
</tr>
<tr>
<td>The MRL/lpr mouse (mutation in the fas gene controlling apoptosis)</td>
<td>Mice</td>
<td>lpr</td>
<td>Generalized inflammation as a part of lupus disease, which also affects joints</td>
<td>119</td>
</tr>
<tr>
<td>Stress-induced arthritis</td>
<td>DBA/1 mice</td>
<td>Non-MHC genes</td>
<td>Enthesopathic response with no evidence for immune involvement. A model for psoriasis arthritis.</td>
<td>115</td>
</tr>
<tr>
<td>TNF transgenic mice (overproduction of TNF)</td>
<td>Mice</td>
<td>TNF transgene</td>
<td>Erosive arthritis, as well as generalized tissue inflammation Arthritis</td>
<td>120</td>
</tr>
<tr>
<td>IL-1Ra--deficient mouse</td>
<td>Balb/c mice</td>
<td>ILRa deficiency</td>
<td>Arthritis</td>
<td>126</td>
</tr>
<tr>
<td>Gp130 IL-6R mutated mouse</td>
<td>C57 Black mice</td>
<td>IL-6R mutation</td>
<td>Severe arthritis due to transgenic encoded glucose 6 phophoisomerase specific T cells</td>
<td>109</td>
</tr>
<tr>
<td>K/BxN arthritis</td>
<td>Mice</td>
<td>TCR transgenic on NOD</td>
<td>Severe arthritis with autoreactivity</td>
<td>55</td>
</tr>
<tr>
<td>ZAP70 mutation</td>
<td>Balb/c mice</td>
<td>Spontaneous mutation in ZAP70 in Balb/c mice</td>
<td>Erosive arthritis</td>
<td>136</td>
</tr>
<tr>
<td>HTLV transgenic mouse</td>
<td>Mice</td>
<td>HTLV transgene</td>
<td>Generalized tissue inflammation</td>
<td>135</td>
</tr>
<tr>
<td>HTLV transgenic rat</td>
<td>Rats</td>
<td>HTLV transgene</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
degrading the bacterial cell wall structures and are therefore transformed into an activated state, which trigger inflammation.

MCWIA can be abrogated by elimination of the classical α/β type of T cells and spleen-derived T cells\(^{19,20}\) can transfer the disease. The specificity of such T cells has, however, not been reproducibly demonstrated, although some possibilities have been suggested including bacterial structures and cross-reactive self-components.\(^{21,22}\) Although a role for heat shock proteins as T cell antigens has not been confirmed, they play a regulatory role for the development of arthritis.\(^{23}\) In the search for the minimal arthritogenic structure in mycobacterium, it was observed that one of the essential structural elements of the mycobacterium peptidoglycan, muramyl dipeptide, could induce arthritis.\(^{24}\) Interestingly, T cells do not recognize this structure, but it has potent adjuvant capacity as it activates the inflammasome by stimulating innate immune receptors (NOD2) and antigen-presenting cells.\(^{25}\) The unmethylated DNA of bacteria has also been shown to independently trigger arthritis in mice\(^{26}\) and contribute to arthritis severity of MCWIA in rats.\(^{27}\) The bacterial DNA triggers Toll-like receptors on both antigen-presenting cells and inflammatory macrophages and will therefore interact with both T cell–dependent and inflammatory pathways. Another T cell–dependent arthritogenic pathway is triggered by the mineral oil in which the mycobacteria is suspended, as is discussed later in more detail.\(^{19-24}\) Thus this classical “adjuvant-induced arthritis” (MCWIA) is mediated by different and interacting pathways, dependent on both different mycobacterium cell components such as peptidoglycans, DNA, and heat shock proteins but also dependent on adjuvant activity mediated by the oil used to suspend the mycobacteria.

Postinfectious arthritis has also been observed to occur following streptococcal infections. A rapidly developing form of arthritis has been observed after systemic inoculation of streptococcal cell wall fragments in rats\(^{28}\) and mice\(^{29}\) but not in primates.\(^{30}\) Peptidoglycans from the cell wall rapidly disseminate throughout the rat including the joints.\(^{31}\) These structures are difficult to degrade for the macrophages, and as a consequence synovial macrophages are persistently activated. T cells are necessary for the initiation and perpetuation of the arthritis.\(^{32}\) Although the precise mechanisms are not known, it is possible that there are mechanisms shared with MCWIA.

### Nonbacterial Adjuvant-Induced Arthritides

Research has found that the induction of arthritis in rats was dependent on not only the mycobacteria but also the oil into which the mycobacterium fragments were suspended. Interestingly, some oils supported the induction of arthritis, whereas others did not.\(^{33}\) Many years later it was noted that the mineral oils that supported the induction of arthritis were in fact arthritogenic by themselves.\(^{33}\) It was also found that subcutaneous administration of nonbacterial adjuvant compounds such as pristane, hexadecane, and squalen were highly effective in inducing arthritis.\(^{34-36}\) These adjuvant compounds in most cases produce inflammation confined to the joints and offer more appropriate experimental models for RA than the earlier commonly used “adjuvant arthritis.”

Mineral oil–induced arthritis (OIA),\(^{33}\) avridine-induced arthritis (AvIA),\(^{34}\) pristane-induced arthritis (PIA),\(^{35}\) hexadecane-induced arthritis,\(^{37}\) and squalen-induced arthritis\(^{36}\) in the rat share many common features but differ by the degree of chronic development (Figure 28-1). They are induced with adjuvant compounds lacking immunogenic capacity (i.e., no specific immune responses are elicited). Instead they are rapidly spread throughout the body after a single subcutaneous injection, penetrate through cell membranes into cells, and interact with yet unknown cell surface receptors and intracellular proteins. One or two

---

**Figure 28-1** Pristane-induced arthritis in dark agouti rats. Induced with pristane subcutaneously. Development of severe and chronic relapsing arthritis starting 10 to 14 days after injection.
weeks after injection, arthritis suddenly develops. The arthritis appears in the peripheral joints, with a similar distribution as seen in RA. Occasionally other joints are involved, but systemic manifestations in other tissues have so far not been reported.

In certain rat strains, especially in the PIA model, the arthritis proceeds as a chronic relapsing disease. Interestingly, a systemic immune response leading to production of antibodies to RA33 and rheumatoid factors occurs, whereas no consistent immune response to specific cartilage components or citrullinated proteins has yet been observed.

A role for cartilage proteins in regulating disease activity is, however, still possible because the disease can be prevented and in fact therapeutically ameliorated by nasal vaccination with various cartilage proteins. Both the initiation and the chronic progression of the arthritis is T cell dependent as shown by in vivo administration of antibodies to α/β T cells. Together with the observation that the chronic disease course is associated with the MHC region, this could implicate an activation of T cells recognizing joint-derived proteins. However, T cell transfer of the disease seems to be oligoclonal rather than monoclonal and has so far failed to identify antigen-specific T cells. A role for environmental infectious agents is not likely because no difference in disease susceptibility could be seen in germ-free rats, although only conventional rats respond to heat shock proteins.

To date there is no evidence for recognition by lymphocyte receptors or receptors involved in the innate immune system. Surprisingly, some of the arthritogenic adjuvants are in fact components already present in the body before injection. For example, pristane is a component of chlorophyll and is normally ingested by all mammals including laboratory rats. Pristane is taken up through the intestine and spread throughout the body. However, they all share the capacity to penetrate into cells where they could change membrane fluidity and modulate transcriptional regulation and in higher doses induce apoptosis. The injection route and dose are critical (i.e., they determine which cell is first activated and to what extent).

The PIA model has been subjected to genetic analysis showing that the disease is polygenically controlled by at least 20 quantitative trait loci (QTL). These QTLs are often shared between the various forms of adjuvant arthritis and to a lesser degree also with CIA. Interestingly, they seem to control distinct phases of the disease such as arthritis onset, clinical severity, joint erosion, and chronicity. An approach to understand the complexity of the adjuvant arthritis, as well as the arthritis process in general, will be to eventually elucidate the underlying genetic polymorphism of these QTLs. This is, however, labor intensive and only a few genes and gene clusters have been positioned. The strongest effect is mediated through a polymorphism of the Ncf1 gene spread in both inbred strains and wild rat populations. The Ncf1 gene codes for the p47phox gene and controls the oxidative burst. Surprisingly, a higher oxidative burst capacity was associated with more severe arthritis. The effect was found to operate before T cell activation and therefore also controls the degree of autoimmunity, linking innate and adaptive immunity. Importantly, the MHC region, which controls the adaptive T cell response, and a C-type lectin gene cluster (APLEC), which is likely important in uptake of antigen to antigen-presenting cells, have also been identified to control PIA.

Adjuvant arthritis is not easily inducible in species other than rats. Of the previously mentioned adjuvant-induced arthritis models, only PIA has been described in the mouse. However, the induction of PIA in mouse requires repeated intraperitoneal injections of pristane, which triggers a widespread inflammatory disease with a late and insidious onset. In fact, the induced disease mimics systemic lupus erythematosus (SLE) rather than RA. The disease is clearly different from PIA in the rat; the same induction protocol does not induce disease in the rat, and the disease course and characteristics are different. Another adjuvant-related model is the induction of mild arthritis after intrarticular injection of agents activating macrophages such as unmethylated DNA. However, it has more recently been found that several mouse models earlier believed to be spontaneous are critically dependent on adjuvants and should therefore be classified as adjuvant-induced arthritis. From the observation that a BALB/c substrain in Japan spontaneously developed arthritis, a mutation in the Zap70 was identified. The Zap70 mutation (W163C) caused weaker TCR-mediated signaling, and the development of arthritis was preceded by increased levels of IL-17-producing autoreactive T cells. The Zap70 mutation led to defective positive selection and the emergence of autoreactive T cells attacking the joints. As a result, both rheumatoid factors and CIA reactive antibodies were detected in the mice. However, the arthritis did not develop in specific pathogen-free-housing conditions and it could be shown that injection of β-glycan or mannan induced the arthritis. Thus this model seems to be an adjuvant-induced arthritis in the mouse.

Other spontaneous arthritis models, the KxB/N model and the IL-1R deficient mouse, have been shown to be mediated by a likely adjuvant component because arthritis did not develop or was dramatically attenuated under germ-free conditions. The causative arthritogenic effect could be shown to be mediated by intestinal bacteria, segmented filamentous bacteria (SFB), and lactobacillus, respectively.

Taken together, there is today a set of useful adjuvant-type arthritis models in both mice and rats.

**Cartilage Protein–Induced Arthritis**

Arthritis is inducible with several different cartilage proteins such as aggrecan, link protein, type XI collagen CXI, cartilage oligomeric matrix protein (COMP), and CII. These various models have different characteristics and genetics. CIA, induced with CII, is today the most commonly used model for RA. It was first demonstrated in the rat and was later reported using other species such as mouse and primates. Today CIA is the most commonly used model for RA.

**Collagen (II)–Induced Arthritis**

Immunization with the major collagen in cartilage, type II collagen (CII), leads to an autoimmune response and as a consequence, sudden onset of severe arthritis. Although it is usually necessary to emulsify the CII in adjuvant such as mineral oil in the rat and complete Freund’s adjuvant in the
mouse, the disease can be distinguished from the various forms of adjuvant arthritis. However, the CIA model varies considerably depending on the experimental animal strain, the adjuvant used, and whether CII used is of self or nonself origin.

In both rats and mice immunized with heterologous CII, a severe, erosive polyarthritis develops 2 to 3 weeks after immunization but usually subsides within 3 to 4 weeks (Figure 28-2). The most commonly used DBA/1 strain thus develops a severe but only an acute disease. However, a genetic influence is obvious because mice on C57Bl/10 backgrounds develop a milder arthritis that later may develop into a more chronic relapsing disease course (Figure 28-3). In all of the models the erosive inflammatory phase is followed by a healing phase with pronounced formation of new cartilage and bone that clinically can be difficult to distinguish from active inflammation. The disease is critically dependent on both T cell and B cell responses to CII, and pathogenic antibodies play a role in the inflammatory attack on the joints. The CAIA is inducible with certain CII-specific monoclonal antibodies (Figure 28-4). These CII-specific antibodies bind to the cartilage and destabilize the cartilage matrix. Subsequently the inflammatory response is triggered with infiltration of antibodies into cartilage matrix, fixation of complement attraction of neutrophilic granulocytes, and activation of FcR expressing inflammatory cells in a process independent of the immune system. Interestingly, the epitopes recognized on CII contain arginines that potentially can be citrullinated. Recently it could be shown that one of the
major CII epitopes is citrullinated, and monoclonal antibodies specific for the citrullinated peptide induce arthritis showing an important link to RA.

The disease induced after immunization with homologous CII in both rats and mice is not as easily inducible, but once started it is severe and tends to be more chronic than the disease induced with heterologous CII. The pathogenic events in the chronic disease phase are largely unknown but are most likely dependent on both autoreactive B cell and T cell activity. Nevertheless, the CIA model is the most extensively investigated model for RA and has given valuable insights into the genetic control of the arthritic process and of the autoimmune interactions with cartilage. It has also been useful for the development of new therapeutic approaches and for drug screening.

Genetic Basis of Collagen-Induced Arthritis

Susceptibility to CIA varies dramatically between different inbred strains. The CIA is a complex, polygenic disease, similar to the adjuvant arthritis models described earlier. In the CIA model the autoimmune process is already determined by the induction through immunization with a defined antigen. Not surprisingly, the MHC class II polymorphism is important for determining susceptibility, but there is also a major influence by a large number of genes outside the MHC region. The major gene regions have been identified through genetic segregation experiments in both mice and rats, which have given an overall picture of the genetic inheritance of the susceptibility. As in other complex diseases, these genes operate in concert and can only be identified through isolation in a controlled genetic and environmental context. So far MHC class II genes (Aq), Ncf1, and complement C5 have been positioned from genetic analysis. The Ncf1 gene was defined through analysis of PIA and CIA in rats. A spontaneous mutation in the mouse Ncf1 gene, when combined with the CIA-susceptible MHC class II allele Aq in the C57Bl/10 mouse, develops a chronic relapsing form of CIA. In addition, these mice tend to develop a severe form of chronic arthritis in the postpartum period, with the spontaneous development of autoimmunity to CII. Another genetic polymorphism of importance is the complement C5, which is deficient in many mouse strains. The deficiency leads to a relative resistance to CIA, suggesting a role for complement pathways in arthritis, which in fact is opposite of its role in mycoplasma-induced arthritis. A role for alternative complement pathways and Fc receptor–mediated pathways has been demonstrated using both CIA and the CAIA models. The rapid progress of genome-wide association analysis of large human cohorts today gives direct information on involved genes and gene clusters in common diseases such as RA. However, the animal models are necessary to understand their functional relevance; for this, the genes controlling the corresponding disease in the animals need to be identified. Knowledge of animal model genetics will facilitate creation of humanized models through genetic modification.

Role of the Major Histocompatibility Complex

Early observations using the CIA model in both mice and rats indicated a role for the MHC region. In the mouse, CIA induced with either heterologous or homologous CII is most strongly associated with the H2d and H2e haplotypes, whereas most other haplotypes such as b, s, d, and p are relatively resistant. The major underlying gene within the H2d haplotype has been identified as Aβ. Moreover, the immunodominant peptide derived from the CII molecule bound to the arthritis-associated q variant of the A molecule has been found to be located between positions 259 and 271 of CII. This peptide can be glycosylated on the central lysine side chain and is recognized by most of the CII-reactive T cells. Interestingly, the peptide is also bound by DR4 (DRB1*0401/DRA) and DR1 molecules (i.e., the shared epitope), which are associated with...
RA. Mice transgenically expressing DR4 or DR1 are susceptible to CIA and respond to CII259-271 peptide, and CII-reactive T cells from RA patients seem to predominantly recognize the glycosylated forms of the CII259-271 peptide. These findings suggest a model for studies of RA that not only mimic some basic pathogenic events but may also share some critical structural similarities.

Arthritis is also inducible in mouse strains that do not express q or r. A commonly used model is to induce arthritis with high doses of chicken or bovine CII emulsified in Mycobacterium tuberculosis–containing CFA. T cell autoreactivity to CII has not yet been reproducibly demonstrated, however, and it is possible that the T cell reactivity is directed to a contaminant in the preparation such as another matrix protein or the pepsin used for the preparation. Such T cell reactivity could help B cells produce antibodies to CII, explaining the development of arthritis.

Autoimmunity in Collagen-Induced Arthritis

It is important to emphasize that the identified structural interaction between MHC class II–positive peptide complexes and T cells does not give us the answer to the pathogenesis of CIA (or RA), but rather a better tool for further analysis. An important question is how the immune system interacts with the peripheral joints (i.e., how autoreactive T and B cells are normally tolerized and what happens in the pathologic situation, after their activation by CII immunization). Most of the T cells reactive with the rat CII259-271 peptide do not cross-react with the corresponding peptide from mouse CII. The difference between the heterologous and the homologous peptide is position 266 in which the rat has a glutamic acid (E) and the mouse an aspartic acid (D), which leads to a weaker binding of the mouse peptide to Aα. The importance of this minor difference was demonstrated in transgenic mice expressing CII mutated to express a glutamic acid at this position. When mutated CII was expressed in cartilage, the T cell response to CII was partially but not completely tolerized. The mice were susceptible to arthritis, but the incidence was low, similar to what is seen in mice immunized with homologous CII. This finding shows that a normal interaction between cartilage and T cells leads to the activation of T cells, but with less capacity to induce arthritis or with regulatory properties. These CII autoreactive T cells may under extreme circumstances (such as CIA immunization) be pathogenic. In contrast, B cells reactive with CII are not tolerized and as soon as the T cells are activated, even in a partially tolerized state, they may help B cells to produce autoreactive and pathogenic antibodies. It is possible that a similar situation in humans could explain the difficulties in isolating CII-reactive T cells compared with the relative ease in which CII-reactive B cells can be detected in the joints.

Induction of Arthritis with Other Cartilage and Joint-Related Proteins

Type XI Collagen-Induced Arthritis

The type XI collagen (CXI) is structurally similar to CII and is to a large extent co-localized. CXI is a heterotrimer with three different α chains, one of which is shared with CII (the α3 chain). Both heterologous and homologous CXI have been reported to induce arthritis in rat strains. Interestingly, the induction with homologous CXI gives a chronic relapsing disease, which is distinctly different from the heterologous CXI-induced disease and CII-induced CIA.

COMP-Induced Arthritis

Another cartilage protein is cartilage oligomeric matrix protein (COMP). Homologous COMP induces arthritis in both rats and mice. In comparison with CIA, the resulting disease is self-limited, is less erosive, and has a different genetic control.

Proteoglycan (Aggrecan)-Induced Arthritis

Other major components of joint cartilage are proteoglycans, of which the largest is aggrecan. Immunization of Balb/c mice with fetal human aggrecan induces chronic arthritis. Both B and T cells are involved in the pathogenesis. Autoreactive T cells have been isolated and respond to the GI-domain of aggrecan in which neoepitopes are created and T cell receptor transgenic mice spontaneously develop arthritis at high age. The disease has been genetically mapped and shown to share many gene regions in common with CIA.

Antigen-Induced Arthritis

Antigen-induced arthritis is a classical model of RA that is induced by immunizing animals with a foreign antigen, usually bovine serum albumin, and subsequently injecting the same antigen into a joint. As a result, a pronounced T cell–dependent, immune complex–mediated, and destructive arthritis develops in the injected joint. The model is well controlled and has been used to understand the effector phase of the cartilage destruction.

Glucose-6-Phosphoisomerase–Induced Arthritis

The successful induction of arthritis after immunization with recombinant glucose-6-phosphoisomerase (G6PI) in adjuvant stems from the identification of the K/BxN model in which transgenic T cells specific for GPI lead to spontaneous arthritis in NOD mice. The GPI-induced arthritis is MHC dependent and associated with the H2q haplotype (as CIA), and it has been shown that GPI peptides with an Aα binding capacity can induce arthritis. Interestingly, GPI has a unique affinity for cartilage as it binds with high affinity to cartilage proteoglycans and spontaneous immune activation activation in the K/BxN mouse seems to primarily arise in lymph nodes draining joints. Thus although GPI is ubiquitously expressed, the immune system may recognize its presence in the joints.

Spontaneous Arthritis

Many of the classical inbred mouse and rat strains tend to spontaneously develop arthritis, in particular under certain environmental influence (Table 28-2). In some strains such as DBA/1, the grouping of males easily induces...
intermale aggressiveness and such stress seems to be associated with development of severe arthritis. This stress-induced arthritis is different from inflammatory arthritis models like CIA and has less inflammatory synovial infiltrate, but enthesopathy and new cartilage and bone formation, more similar to psoriasis arthritis than RA, dominate the joint pathology.

In addition, a number of genetic mutations strongly enhance arthritis development. One such mutation has been shown to occur in the Fas gene, of critical importance for apoptosis. In the MRL mouse background, arthritis develops together with a severe lupus disease. More recent research found that a mutation in the T cell receptor–signaling molecule Zap70 was associated with severe arthritis in Balb/c mice housed under conventional but not specific pathogen-free conditions.

**Spontaneous Arthritis in Genetically Modified Strains**

Importantly, models developing spontaneous arthritis have been possible to create by genetically modifying mouse strains. The prime example of this is the demonstration that overexpression of tumor necrosis factor (TNF) leads to severe arthritis. This model has been extremely useful in delineating the role of TNF in mediating arthritis. Other genetic mutations leading to overexpression of TNF also lead to spontaneous development of severe arthritis such as deletion of an upstream regulatory element controlling TNF secretion in fibroblast or the deletion of DNasII leading to aberrant secretion of TNF by chronically stimulated macrophages. Mice overproducing TNF develop arthritis irrespective of a functional immune system, which is thus operating entirely through innate inflammatory mechanisms.

Another important lesson from the TNF-overproducing mice was that it primarily led to the development of arthritis, although in some situations colitis and encephalomyelitis could also develop. Subsequently several other genetically modified mouse strains developing arthritis have been reported. A mouse deficient for the IL-1 receptor antagonist developed arthritis that not only affected the downstream effector functions but could also be shown to be dependent on T cell activation.

A mutation in the gp130 IL-6 receptor was observed to lead to arthritis in elderly mice. Interestingly, the mutation led to an accumulation of polyclonal autoreactive CD4+ T cells secreting inflammatory cytokines, indicating a regulatory role of the IL-6 receptor of the adaptive immune system.

Another type of spontaneous arthritis was observed in a T cell receptor transgenic mouse in which the TCR recognizes a peptide derived from the ubiquitously occurring protein G6PI bound to the MHC class II protein Ag7. Importantly, however, the arthritis does not develop if the mice lack the SFB bacteria in the intestinal flora, indicating that this disease is not strictly spontaneous but rather induced by a bacterial adjuvant stimulation. Research indicates that the pathogenic effector pathway in this model depends on antibodies reactive with G6PI. The arthritis can be transferred with such antibodies and bind to the cartilage surface, mimicking the pathogenesis of anti-CII antibodies in the CAIA model. The use of the G6PI antibody–induced arthritis has been instrumental in finding early inflammatory steps in the joint attack, which involves complement activation through the alternative pathway, mast cell activation, and neutrophil infiltration. Clearly the joints are specifically targeted in the disease, and it remains to be determined how T cells and antibodies recognize this systemically expressed autoantigen in a joint-specific context.

A transgenic model in which spontaneous arthritis has been observed is in mice and rats transgenic for the envelope protein of human T cell leukemia virus 1. In this case there is not only joint inflammation and autoimmunity to CII but also widespread inflammatory infiltrates in skin, salivary glands, and vessels.

Another type of model is the induction of arthritis after transplantation of human synovial fibroblasts into immunodeficient SCID mice. The same type of arthritis develops after transfer of murine fibroblast cell lines. This model is likely to reflect inherent properties of fibroblast-mediated mechanisms showing different features as compared with other arthritis models like CIA and PIA.

These models most likely represent various aspects of the processes leading to arthritis, which will be determined by the transgene or defective gene or due to transplantation of specific cells.

**USING ANIMAL MODELS**

**Increasing Knowledge of Disease Pathways**

An ideal model for human RA should mimic the complexity of the human disease in being polygenic and dependent on environmental factors. The animal models have the advantage that both genetics and environment can be better controlled. RA is a syndrome, likely composed of several distinct disease entities as are the animal models.

Ideally, the animal models should mimic the various subtypes of RA and with increasing knowledge of RA such as the identification of ACPA as a predictive biomarker and the identification of new genes associated with RA, there will also be new demands on the animal models. Thus there is not yet an animal model reflecting the production of ACPA, nor have proper humanized mice been developed that pathophysiologically mimic the genetic polymorphism identified in humans. For introducing new genes (e.g., human MHC class II) or environmental factors (e.g., smoking), a proper and well-controlled genetic and

<table>
<thead>
<tr>
<th>Environmental Effect</th>
<th>Effect on Arthritis</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermale stress</td>
<td>+</td>
<td>115</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>–</td>
<td>151</td>
</tr>
<tr>
<td>Postpartum</td>
<td>+</td>
<td>70, 144</td>
</tr>
<tr>
<td>Estrogen</td>
<td>–</td>
<td>143</td>
</tr>
<tr>
<td>Darkness</td>
<td>+</td>
<td>152</td>
</tr>
<tr>
<td>Infections (segmented filamentous bacteria)</td>
<td>+</td>
<td>58</td>
</tr>
</tbody>
</table>
Environmental context is critical. For example, the introduction of the human MHC class II as transgenes in the mouse has led to a number of artifacts, some of them related to the nonphysiologic interactions between human and mouse genes. However, as long as the mouse models have been studied with a well-controlled genetic and environmental context, they have and will contribute with detailed information on the molecular pathogenesis leading to arthritis. For this work, the possibility to genetically modify both mice and rats is a powerful tool and the possibility to make controlled experiments is a critical advantage.

Developing New Therapeutic Strategies

To test new drugs and therapies it will be necessary to select from the different models available. Obviously there is no optimal model for RA, and there will never be one. The models described, however, are useful because they represent different aspects of RA pathogenesis. Thus depending on the questions to be asked or symptoms to be treated, different models may be used. The most common model today used for testing new therapeutic approaches is the CIA model, which should be included as a reference model. The usefulness of this model has been confirmed with the anti-TNF treatment, which was subsequently introduced in RA.

Recapitulating the three hallmarks for RA discussed earlier—tissue specificity, chronicity, and MHC association—reasonable criteria should be that the animal models should display these hallmarks. A common mistake is to only use acute models and to only use disease prevention and not established chronic disease as a readout. It is also of critical importance to be aware of the specific environmental influences on arthritis development in rodents (Table 28-3). Of particular importance are stress effects, which are easily produced by mixing mice from different litters in the same cage and which will lead to cage-dependent effects. Other important factors are sex hormones and most likely also neurohormones, which play an important role in modulating disease activity—seen as effects by estrous cycling, pregnancy, and light effects. Clearly, environmental and genetic effects need to be controlled. The control of genetics is usually achieved by testing standardized inbred strains. The problem is that these vary considerably between different colonies mainly due to genetic contamination. In spite of these problems, there is no question that both environment and genetics can be better controlled in experimental animal models than can be achieved in studies directly involving the human population.

### Table 28-3 Some Environmental Effects on Rat Arthritis

<table>
<thead>
<tr>
<th>Environmental Effect</th>
<th>Effect on Arthritis</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise stress</td>
<td>++</td>
<td>153</td>
</tr>
<tr>
<td>Predator stress</td>
<td>−</td>
<td>154</td>
</tr>
<tr>
<td>Estrogen</td>
<td>−</td>
<td>155</td>
</tr>
<tr>
<td>Testosterone</td>
<td>−</td>
<td>155</td>
</tr>
<tr>
<td>Infections</td>
<td>−/+</td>
<td>13, 46, 156, 157</td>
</tr>
</tbody>
</table>

+, increased arthritis; −, decreased arthritis.

Ethical Considerations. One important drawback of using experimental models for RA is suffering of animals. However, in the light of various human activities that use animals, the use of them in research is readily defensible. In fact, it could be unethical not to use them for research because it would prohibit further understanding of human diseases, thereby letting humans suffer from something curable or preventable. It should also be emphasized that the recent development of animal models for RA has refined them to be of more specific use, which has decreased animal suffering. For example, the historically most commonly used model for RA, the Mycobacterium-induced adjuvant arthritis, is a systemic and severe inflammatory disease, whereas pristane-induced arthritis and collagen-induced arthritis are more specific diseases of the joints.

CONCLUSIONS

Experimental animal models are essential tools for not only investigating the basic mechanisms leading to RA but also developing new therapies. Many models have been described, and each represents different aspects of the disease. Therefore it is important to use different models. The models for RA described so far can be divided into four principal different groups: (1) adjuvant induced, (2) cartilage protein induced, (3) passively induced, and (4) spontaneous. Of emphasis is that the models used should reflect essential hallmarks of RA such as tissue specificity, chronicity, and MHC class II gene association, and they should reflect the fact that RA is a polygenic disease triggered by unknown and multifactorial environmental factors.

Selected References


Full references for this chapter can be found on www.expertconsult.com.
References


PART 3  EFFECTOR MECHANISMS IN AUTOIMMUNITY AND INFLAMMATION


Neural Regulation of Pain and Inflammation

RAINER H. STRAUB

KEY POINTS
The sensory nervous system regulates peripheral inflammation, including release of mediators such as substance P and calcitonin gene–related peptide. The sensory nervous system processes nociceptive stimuli through pain pathways, which can be amplified by inflammation.

During inflammation, an efferent systemic response increases activity of the hypothalamic-pituitary-adrenal axis and the sympathetic nervous system while decreasing activity of the hypothalamic-pituitary-gonadal axis and the parasympathetic nervous system.

Inflammation leads to partial loss of sympathetic nerve fibers in inflamed tissue and in secondary lymphoid organs, which enhances proinflammatory activities via \( \alpha_1/2 \)-adrenergic pathways.

Recruitment or activation of neurotransmitter/neuropeptide-producing cells in inflamed tissue is an anti-inflammatory signal.

For over 2000 years (since Celsus and Galen), clinicians recognized that cardinal features of neurogenic responses, such as redness, warmth, swelling, and pain, are rapid sequela of inflammation. Neurogenic vasodilatation reported in 1876 by Stricker and in 1901 by Bayliss; the inflammatory axon reflex with erythema observed in the 1910s by Bruce and by Breslauer; the flare response reported by Lewis around 1930 with erythema, hyperalgesia, and edema; rediscovery of the antidromic vasodilatory flare response and dorsal root reflex by Chapman; and Kelly's and Janco's more extended concepts of neurogenic inflammation in the 1960s all were expressions of the same principle: the influence of sensory afferent nerve fibers on acute inflammation and on cardinal clinical signs of inflammation. In the past two decades, our view has expanded to include the sympathetic and parasympathetic efferent nervous systems in inflammatory/immune control.

The concept of neuronal regulation of inflammation is supported by reports of patients with hemiplegia and chronic inflammatory diseases, in whom the paralytic side is protected from inflammation (Table 29-1). Cases have been reported in which hemiplegia manifested long after the outbreak of chronic inflammatory disease or long before, leading to protection independent of the time point (see Table 29-1).

In addition, clinical observations demonstrate neuronal regulation of inflammation in that symptoms of many chronic inflammatory diseases have diurnal variation, with greater activity in the night and early morning hours. Because the rhythm of circadian changes in clinical signs depends on superordinate control of the hypothalamic nucleus suprachiasmaticus, a functional connection is revealed between the central nervous system (CNS), efferent pathways of the CNS (hormonal and neuronal), and the inflammatory/immune response.

Neuronal regulation of inflammation is dependent on a robust innervation of lymphoid organs and the direct influence of neurotransmitters/neuropeptides on immune cells. Although sympathetic nerve fibers usually follow arteries (also branching into vessel-free regions), sensory nerve fibers have their own routes along vessels or independent of the vasculature. In addition, nerve fibers of the parasympathetic nervous system innervate many tissues in the head, neck, and trunk of the body, and upper and lower limbs are excluded. The greatest support for a neuroimmune contact comes from innervation of lymphoid organs, where nerve fibers are responsible for neuronal regulation of immune responses.

The receptors for neurotransmitters are present on almost all immunocompetent cells. Some exceptions are known, such as the absence of \( \beta_2 \)-adrenergic receptors on T helper type 2 cells. The differential and time-dependent expression of receptors can shape the neuroimmune cross-talk. Sometimes receptor expression is increased or decreased in the context of an inflammatory response. The time-dependent involvement of different immune cells and receptor expression in the course of a given disease is probably important for neuronal regulation of inflammation.

In addition, intracellular signaling pathways of neurotransmitter receptors are dependent on environmental conditions; this has been demonstrated for G protein–coupled receptors. Another control process of these receptors involves regulators of G protein signaling (RGS), and tumor necrosis factor (TNF) can lead to increased desensitization of \( G \alpha \) protein–coupled receptors. For additional details on some receptors of neurotransmitters on immune cells, the reader is referred to the literature.

In the study of neuronal regulation of inflammation, the biphasic role of neurotransmitters is important. For example, norepinephrine binds to \( \alpha \)- and \( \beta \)-adrenergic receptors, which exert opposing effects on intracellular signaling cascades (\( \alpha_1 \): increase in diacylglycerol and protein kinase C; \( \alpha_2 \): decrease in cyclic adenosine monophosphate [cAMP]; \( \beta \): increase in cAMP). Although norepinephrine binds to \( \alpha \)-adrenergic receptors at between \( 10^{-9} \) and \( 10^{-5} \) mol/L, it binds to \( \beta \)-adrenergic receptors only at concentrations equal to or higher than \( 10^{-8} \) mol/L (typical serum level: \( 10^{-9} \) mol/L; typical tissue concentration: \( 10^{-7} \) mol/L). Because \( \alpha \)-adrenergic effects (proinflammatory) are markedly different from \( \beta \)-adrenergic effects (anti-inflammatory),
the concentration of norepinephrine in the environment of an immune cell is very important for noradrenergic effects. The situation is very similar for adenosine via A1 adenosine receptors (e.g., α-adrenergic) and A2a/b adenosine receptors (e.g., β-adrenergic). It is important to note that this behavior is typical for many neurotransmitters/neuropeptides because more than one receptor can be a binding partner, and different receptors have opposing intracellular signaling pathways. In conclusion, neuronal regulation of inflammation is an important aspect of the inflammatory process.

The role of neuronal reflexes can be explored by examining three key phases of the inflammatory process: (1) phase 1 includes first inflammatory actions within the first 12 hours; (2) phase 2 describes inflammation from several hours to several days until resolution of inflammation (the normal wound healing process); and (3) phase 3 starts with the onset of chronic inflammatory disease that does not properly resolve. During evolution, mechanisms were positively selected that serve to overcome acute transient inflammatory episodes but not chronic lifelong inflammation, because of the negative selection pressure. Transient inflammatory episodes, for example, include infections, wound healing responses, foreign body reactions, immune reactions during pregnancy, and others. Mechanisms of these short-lasting episodes are also used in chronic inflammatory diseases. From this point of view, it is meaningful to start with an acute transient inflammatory response such as a wound response after injection of foreign material into the skin.

**ACUTE INFLAMMATION (THE FIRST 12 HOURS)**

**Recognition of Foreign or Pathogenic Material: Immune and Pain Pathways**

After injection of foreign material into the skin, there are two categories of recognition: (1) recognition by local cells and (2) systemic recognition. These two forms of recognition are interwoven, and the strength of the local response accounts for the magnitude of systemic involvement (see later).

Systemic recognition of foreign material occurs in highly specialized nerve endings of sensory afferent, nociceptive nerve fibers (the nociceptor). Nerve endings of sensory afferent nerve fibers possess an impressive array of receptors that are responsible for instant activation of the nerve fiber (Figure 29-1). Upon introduction of foreign material, infectious agents can pose a threat, which can elicit a neuronal response via pattern recognition receptors on polymodal nociceptors (e.g., the Toll-like receptors) (see Figure 29-1). In addition, factors such as bradykinin, prostanoids, and cytokines from activated mast cells and other cells can stimulate their respective receptors on sensory nerve terminals (see Figure 29-1). Under consideration of these mechanisms, peripheral recognition of foreign material by nociceptors is part of the innate immune response. Moreover, mechanical irritation, noxious cold/heat, and low pH concentration stimulate the sensory afferent nerve fiber (see Figure 29-1). Altogether, this leads to an orthodromic action potential that stimulates the dorsal root ganglion (DRG) and releases elements such as substance P into the wounded peripheral tissue (efferent function of sensory afferents). The spreading reaction is attributed to the axon reflex and the dorsal root reflex, which lead to antidromic activation of neighboring sensory afferents, resulting in local expansion of the immediate flare response.

Substance P is one of the strongest chemotactic and vasodilatory factors; it leads to instant plasma extravasation and accumulation of neutrophils, monocytes, and other cells. Substance P and other neuropeptides increase vascular leakage and interstitial fluid volume in connective tissue capsules, tendons, and muscles, leading to stiffness. In addition, substance P immediately stimulates activities of mast cells, monocytes, macrophages, dendritic cells, and neutrophils to reflexively increase local proinflammatory responses. Parallel to substance P, calcitonin gene–related peptide (CGRP) with strong vasodilatory and chemotactic activities is released. The third sensory neurotransmitter is the excitatory amino acid glutamate, whose proinflammatory effects have been described. The fourth neurotransmitter of sensory afferent nerve fibers is galanin, which possibly has dual proinflammatory and anti-inflammatory roles, depending on receptor subtypes (but data are limited with respect to effects on immune cells). All these neurotransmitters/neuropeptides are locally secreted into the vicinity of the peripheral nerve terminal.

In addition to local effects of these neurotransmitters/neuropeptides, pain signals are transmitted to the brain and elicit a systemic response (Figure 29-2). The pathways ascend through sensory nerve fibers (Aδ or C fibers), the
neurons in the DRG, the neurons in the spinal medulla, and the contralateral spinothalamic tract to reach the medial and lateral thalamus, cortical areas S1 and S2, the hippocampus, and other brain regions responsible for affective components of pain (anterior cingulate cortex, insula, and prefrontal cortex) (see Figure 29-2). All parts of the pain pathway can be sensitized under the influence of inflammatory stimuli. Sensitization means stabilization and amplification of nociceptive stimuli.

**Peripheral Sensitization**

Sensitization appears already during the earliest phase of inflammation, as demonstrated in the kaolin/carrageenan or similar instant chemical models. Nevertheless, sensitization is a dynamic process that changes over time, as demonstrated by inflammation-induced induction of transient receptor potential vanilloid-1 (TRPV1) receptors on DRG neurons, gradual infiltration of macrophages into the DRG, or bilateral long-term upregulation of bradykinin receptor B2 in the DRG and dorsal horn.41 This process involves infiltration of immune cells releasing inflammatory cytokines such as interleukin (IL)-6 and TNF into the joint, leading to expansion of the receptive field and TNF have the potential to sensitize afferent nerve fibers in the joint to mechanical stimulation contributing to mechanical hypersensitivity.43,44 These effects can be blocked by anticytokine therapy with biologic agents, and it is expected that this also inhibits release of proinflammatory substance P and other neuropeptides.

**Figure 29-1** The mechanisms of a polymodal nociceptor. The figure schematically depicts receptors on and neuropeptides of nerve fiber endings of sensory afferent nerve fibers. The list of receptors is not complete. αAR, alpha adrenoceptors; AMPA, α-amino-3-hydroxyl-5-methyl-4-isoxazole-4-propionic acid; ASICs, acid-sensing ion channels; BR, bradykinin receptor; CGRP, calcitonin gene-related peptide; EAAR, excitatory amino acid receptor; GFRs, growth factor receptors; HRs, histamine receptors; IL, interleukin; IRs, inhibitory receptors; NK1, neurokinin 1; NMDA, N-methyl-D-aspartate; NPY, neuropeptide Y; PAR, proteinase-activated receptor; PR, prostaglandin receptor; SP, substance P; TLRs, Toll-like receptors; TNF, tumor necrosis factor; TRPA1, transient receptor potential ankyrin 1; TRPM8, transient receptor potential menthol 8; TRPVI, transient receptor potential cation channel V1.

Most of these units are thin myelinated fibers (Aδ fibers) or unmyelinated fibers (C fibers). Furthermore, mechanosensitive and thermosensitive “silent” nociceptors are sensitized in inflamed tissue, and they start to respond to mechanical and thermal stimuli during inflammation.43,44 This class of receptors is characterized by long-standing responses to algogenic factors, and they are important in neurogenic inflammation.43,44 These mechanisms are summarized under the heading of peripheral sensitization (the “S” in Figure 29-2).

It is important to note that injection of proinflammatory cytokines such as interleukin (IL)-6 and TNF into the joint leads to a huge increase in the number of action potentials recorded from afferent fibers supplying the joint. Both IL-6 and TNF have the potential to sensitize afferent nerve fibers in the joint to mechanical stimulation contributing to mechanical hypersensitivity.43,44 These effects can be blocked by anticytokine therapy with biologic agents, and it is expected that this also inhibits release of proinflammatory substance P and other neuropeptides.

**Central Sensitization**

In the DRG and spinal cord, peripheral inflammation makes neurons hyperexcitable and more susceptible to input from sensory nerve fibers (the “S” in Figure 29-2). This amplifies the response through additional activation of adjacent and even remote spinal neurons far away from the inflamed region, leading to expansion of the receptive field.43,44 The peripheral inflammatory response increases expression of substance P, CGRP, and bradykinin with their respective receptors in the DRG and dorsal horn.43,44 In the dorsal
from inflammatory to neuropathic features of sensitization has been demonstrated by increased expression of a typical marker of neuropathic pain, ATF3, and by the favorable effects of gabapentin treatment in a postinflammatory phase of hypersensitivity.

Spinal sensitization is often a consequence of increased release of excitatory amino acids (glutamate, aspartate, and glycine), substance P, CGRP, neurokinin A, and galanin from nociceptor neurons and upregulation of the respective receptors in the spinal medulla. Enhanced release can be induced by peripheral inflammation.
Figure 29-3  Mechanisms of peripheral and central sensitization. A, Peripheral sensitization. Injection of an inflammatory stimulus leads to an increase in the number of action potentials as recorded from afferent fibers supplying the joint. Peripheral sensitization is mediated by a plethora of heterogeneous receptors on afferent fibers (see Figure 29-1). B, Spinal central sensitization. Left, In the normal situation, only Aβ fibers are activated (upon mechanical stimuli); these are low-threshold nonnociceptor fibers that release glutamate (Glu). On the postsynaptic neuron, only α-amino-3-hydroxy-5-methyl-4-isoxazole-4-propionic acid (AMPA) receptors are activated and opened. Right, In the inflammatory situation, previously high-threshold Aδ and C fibers are activated by pressure, leading to release of glutamate and neuropeptides (NPs) such as substance P and calcitonin gene-related peptide (CGRP). This leads to activation of the postsynaptic membrane via AMPA (AMPA-R) and N-methyl-D-aspartate (NMDA) receptors (NMDA-R), neuropeptide receptors, prostaglandin receptors, and cytokine receptors (particularly, IL-1β, IL-6, TNF). These changes lead to long-standing hypersensitivity. DRG, dorsal root ganglion; GABA, γ-aminobutyric acid; GPCR, G protein–coupled receptor; IN, interneuron; MIG, microglia; NMDA, N-methyl-D-aspartate; NO, nitric oxide; NP, neuropeptide; PG, prostaglandin; STT, spinothalamic tract. The star-formed cell is an astrocyte.
Non-N-methyl-D-aspartate (NMDA) receptors but also NMDA glutamate receptors are relevant in joint inflammation (Figure 29-3B). Sensitization can be mimicked experimentally by intrathecal administration of substance P or NMDA via an increase in prostaglandins or cyclooxygenase-2. In addition to the activating pathway, there exist inhibitory pathways, for example, N-methyl-D-aspartic acid (GABA) or glycine. Second, spinal sensitization is dependent on microglial cells and astrocytes, which can aggravate pathologic pain states in which cytokines and chemotactic factors play an important priming and perpetuating role (see Figure 29-3B). Cytokines such as IL-1β, IL-6, and TNF play a dominant role in cytokine-induced hypersensitivity, and these cytokines are induced in the spinal cord during experimental arthritis.

Several proinflammatory intracellular signaling pathways have been implicated in priming of microglia and pain-processing neurons. It is not easy to distinguish whether signaling cascades in neurons, microglia, or other cells are important, because experimental studies using microdialysis or intrathecal administration of pathway inhibitors do not target a specific cell type. Nevertheless, these studies clearly demonstrate the importance of factors such as nuclear factor κB (NFκB), protein kinase A, protein kinase C, c-Jun N-terminal kinase, JAK/STAT3 signaling pathway, p38 mitogen-activated protein kinase (MAPK), Src-family kinase, arachidonic acid pathways, and others. These pathways typically lead to intraneuronal calcium and sodium accumulation, which is an excitatory signal.

For example, the p38 pathway was demonstrated to be an important proinflammatory signaling cascade in spinal neurons and microglial cells in experimental arthritis. Phosphorylated p38 is increased in microglial and neuronal cells during the course of experimental arthritis. Intrathecal administration of a specific p38 inhibitor led to decreased synovial inflammation but also to suppressed articular cytokine and protease expression and joint destruction as measured by radiographic and histology scores. This effect was dependent on the presence of TNF in the spinal cord. TNF can be a signaling element upstream of p38 by activating p38-phosphorylating kinases, or downstream of phosphorylated p38 that induces TNF secretion. It was demonstrated that intrathecal, but not subcutaneous, TNF neutralization with etanercept inhibited TNF secretion. It was demonstrated that intrathecal, but not subcutaneous, TNF neutralization with etanercept inhibited p38 phosphorylation and peripheral inflammation. The positive effect of intrathecal spinal TNF neutralization was confirmed in another model of experimental arthritis. In this model, peripheral joint inflammation was decreased, and pain-related behavior was drastically reduced.

It is interesting that all mentioned pathways belong to proinflammatory cascades initially described in peripheral immune cells. In contrast, anti-inflammatory pathways such as adenosine A1, β-adrenergic, and δ-opioid receptor pathways are inhibitory in microglial activation paradigms. For instance, spinal administration of an A1 adenosine receptor agonist markedly reduced inflammation, as well as bone and cartilage destruction, in an experimental arthritis model. Administration of the A1 adenosine receptor agonist also decreased nuclear c-Fos expression in the superficial and deep dorsal horns of the spinal medulla. In addition, the A1 adenosine receptor agonist decreased the density of astrocytes in these areas. This indicates that, along with neurons and microglial cells, astrocytes are involved in sensitization.

Finally, understanding why peripheral and central sensitization was conserved during evolution is important. Sensitization of pain has a protective role because it warns about potential danger, enables us to remove noxious stimuli, and stimulates wound management. Furthermore, avoidance of painful situations in the future would be desirable. This is nicely indicated by the fact that peripheral inflammatory stimulation of sensory neurons can induce central IL-1β release in the hippocampus, a cytokine that is instrumental in hippocampal learning phenomena. Sensitization is an amplification factor that should last as long as painful or noxious stimuli are present (or even a little longer to stimulate wound management). Thus, sensitization is a supportive factor of innate immunity. It has been positively selected as an evolutionarily conserved learning phenomenon that will not be stopped until inflammation is terminated (i.e., the stimulus is removed).

**Neuroendocrine Systemic Response**

Parallel to the local inflammatory reaction, hormonal and neuronal systemic responses are engaged. The hormonal response system is mainly the hypothalamic-pituitary-adrenal (HPA) axis, which is stimulated through activation of sensory pathways or through circulating cytokines. Activation of the HPA axis can happen on adrenal, pituitary, and hypothalamic levels. Neuronal efferent response systems include the sympathetic nervous system and the parasympathetic nervous system.

The systemic response of the HPA axis and the sympathetic nervous system is an “energy appeal reaction” that serves allocation of energy-rich fuels from stores to the highly catabolic immune system (Figure 29-4). This re-allocation of energy-rich fuels is important throughout the inflammatory response. The energy demand of the immune system can contribute to a systemic response whose long-standing use is detrimental to homeostasis (see Figure 29-4).

In the course of inflammation, the sympathetic nervous system is activated via direct spinal interneurons that link sensory inputs to sympathetic output (see Figure 29-2, blue lines). This has the advantage that the input defines the location of the output, which leads to confinement of the response to the affected area. In addition to the central coupling of sensory and sympathetic pathways, sympathetic nerve fibers communicate with sensory nerve terminals by way of α2-adrenergic and prostaglandin cross-signaling at the level of the peripheral nerve terminal. This inflammation-induced cross-signaling leads to higher activity of sensory afferents. Systemically, the sympathetic nervous system, similar to the HPA axis, is activated through stimulation of sensory pathways or through circulating cytokines (e.g., IL-6). Activation of the sympathetic nervous system is the main factor in re-allo-locating energy to the immune system (see Figure 29-4).

Although systemically relevant inflammation is coupled to increased sympathetic nervous tone and increased activ-
Activation of the HPA axis and the sympathetic nervous system at the onset of inflammation prepares the immune system for most naturally occurring immune challenges. Activation of the HPA axis mobilizes immune cells, leading to redistribution of neutrophils, monocytes, and natural killer (NK) cells. The sympathetic nervous system can support the very acute inflammatory process in phase 1 because of six main mechanisms: (1) mobilization of immune cells from systemic stores (similar to the HPA axis), (2)...
support of plasma extravasation,86 (3) remodeling of tissue by inducing matrix metalloproteinases,87,88 (4) stimulation of nociceptors via α2-adrenergic and prostaglandin cross-signaling,81,82 (5) chemotactant activity of sympathetic neurotransmitters,89 and (6) liberation of free fatty acids and glucose necessary for the activated immune system (see Figure 29-4). In summary, during the first hours of inflammation, the HPA axis and the sympathetic nervous system are mainly proinflammatory.

Vagal afferents from the intestine and liver play an important role in modulating a systemic milieu that increases or decreases the magnitude of very acute inflammatory hyperalgesia, which depends on the agent, the stimulus strength, and epinephrine secretion from the adrenal medulla.80,81 The vagal tone determines the overall reflex modulation of very acute inflammatory processes, and vagal afferents are important in perception of inflammatory conditions in the abdomen.82,83 Reports in the last decade have demonstrated that lipopolysaccharide-induced inflammation can be inhibited by electrical vagus stimulation of the distal end of the dissected vagus nerve.84 These very acute vagal effects were dependent on the sympathetic innervation of the spleen.85 In addition, carrageenan-induced leukocyte recruitment into a pre-formed subcutaneous air pouch was inhibited by vagus nerve stimulation of the intact vagus nerve.86 This was done without dissection of the vagus nerve so that afferent and efferent vagus nerve effects cannot be separated.86

Administration of an intrathecal p38 inhibitor (mentioned earlier), which has favorable effects in experimental arthritis, largely increases vagal activity.87 Because spinal application of p38 inhibitors blocks aspects of central sensitization,88 one would expect blockade of segmental sympathetic outflow, as demonstrated in Figure 29-2 (blue lines). A decrease in central pain signaling, and thus decreased hypothalamic activation of the HPA axis and sympathetic nervous system, and diminution of segmental sympathetic outflow most probably increase parasympathetic reflex activity. Particularly in very early inflammation, this should be a favorable anti-inflammatory feature. These acute vagus experiments were complemented by experiments in long-standing chronic inflammation models (see later).

INTERMEDIATE INFLAMMATION (BETWEEN 12 HOURS AND SEVERAL DAYS/FEW WEEKS)

Almost all experimental systems dissecting neuroinflammatory pathways have been performed under very acute inflammation conditions (within minutes to 12 hours, reflecting an experimental working day). Much less information is available after 12 hours until termination of uncomplicated inflammation with normal wound healing. Within the mentioned time span, many additional immune/inflammatory responses appear, such as increased local cell accumulation, antigen transport to secondary lymphoid organs, antigen processing, clonal expansion of lymphocytes, release of lymphocytes from secondary lymphoid organs, and access of antigen-specific cells to the target tissue.

In this phase of inflammation, mechanisms discussed in the very early phases can still apply. However, they might change over time because tissue innervation is altered. Immune/inflammatory effector responses relevant in this phase are modulated by neurotransmitters, which can differ from very acute inflammation.

Local Cell Accumulation in Inflamed Tissue

Local cells accumulate in inflamed tissue as a consequence of cell mobilization and chemotaxis. The major neurotransmitters/neuropeptides of sensory afferents (substance P) and of sympathetic efferents (norepinephrine) are potent chemotactic factors for innate immune cells, such as neutrophils, monocytes, and eosinophils. The direct chemotactic effect of substance P has been demonstrated by injecting substance P into the skin; this leads to upregulation of the endothelial adhesion molecule E-selectin (CD62E) and, for example, attraction of eosinophils to the injection site.88 Similarly for the sympathetic nervous system, the lack of catecholamine production in animals with a deletion of the dopamine-β-hydroxylase gene leads to a strong reduction of leukocyte accumulation in the adventitia and periadventitia of vessels.89

Substance P and norepinephrine also have strong chemotactic effects in vitro.89,100 The sympathetic co-transmitter neuropeptide Y and the sensory co-transmitter CGRP also have chemotactic effects.89,101 The effects of substance P and norepinephrine can be amplified by increasing secretion of potent chemotactic factors such as IL-8,102,103 In addition, norepinephrine and substance P can upregulate matrix metalloproteinases to soften the tissue.85,89,104

Immediate Change in Neuronal Innervation

Upon entry of monocytes and neutrophils into the tissue, these cells become activated and can engulf pathogens or foreign material. Activation of these cells is mediated by pattern recognition receptors, as well as by other proinflammatory mediators and inflammasonme-derived products, leading to activation of neutrophils and differentiation of entering monocytes into macrophages or dendritic cells.105

In striking contrast, norepinephrine and its potent co-transmitter adenosine (made from sympathetically released adenosine triphosphate [ATP]) inhibit many proinflammatory effects of activated innate immune cells such as monocytes, macrophages, NK cells, and neutrophils via β-adrenergic and A2 adenosine receptor signaling.106 In this context, it is important to mention that ectonucleotidases (CD39 and CD73), which convert purine precursor neurotransmitters to adenosine, are increased in inflammation.107,108 A classic example of A2-adenosine–mediated or β-adrenergically induced inhibition of cells is the strong negative effect on neutrophil or monocyte/macrophage phagocytosis and on the function of dendritic cells. The important role of adenosine as an inducer of regulatory T lymphocytes has been carefully documented.110 Because innervation is balanced in tissue with similar densities of sensory and sympathetic nerve fibers, this dichotomy of substance P and norepinephrine/adenosine is counterproductive for innate immunity.

Activated macrophages and stimulated tissue fibroblasts start to produce nerve growth factor (NGF), which supports
the outgrowth of sensory and sympathetic nerve fibers equally well. In other words, NGF is not specific for sensory or sympathetic nerve fibers. Indeed, inflammatory tissue releases large amounts of NGF, as, for example, is substantiated in rheumatoid arthritis (RA) or experimental arthritis.\textsuperscript{111,112}

Activated macrophages and fibroblasts also produce nerve repellent factors such as semaphorin 3C and semaphorin 3F.\textsuperscript{113,114} These two factors specifically repel sympathetic nerve fibers and have no effect on sensory nerve fibers, which instead are repelled by semaphorin 3A.\textsuperscript{113,114} In addition, sensory nerve fibers sprout under the influence of NGF into inflamed tissue, leading to a preponderance of substance P over sympathetic neurotransmitters.\textsuperscript{115} Such a sensory hyperinnervation is also observed in skin wounds when sympathetic nerve fibers are absent.\textsuperscript{116} Loss of sympathetic nerve fibers is a rapid process that is observed soon after initiation of experimental inflammation.\textsuperscript{117-119} It can also be observed in vitro with the use of repellent factors in neurite outgrowth assays (within a few hours).\textsuperscript{114}

Repulsion of sympathetic nerve fibers and sprouting of sensory nerve fibers are important ways to initiate a proinflammatory environment in the later phase of inflammation. As shown in Figure 29-5, the appearance of two noradrenergic zones (\(\beta\): normal/healthy; \(\alpha\): inflamed tissue) is a consequence of this process. It is important to mention that loss of sympathetic nerve fibers is observed not only in inflamed tissue but also in the spleen\textsuperscript{117,118,120} and in the lymph nodes. In the former, loss of sympathetic nerve fibers is evident in the white pulp (T cell proliferation area); similarly, these fibers are not observed in B cell follicles.\textsuperscript{117,121} In the same animals, sympathetic nerve fibers sprout into the hilus area of the spleen and do not reach the distal white pulp (T cell)

![Figure 29-5](image_url)

**Figure 29-5** Loss of sympathetic nerve fibers and sprouting of sensory nerve fibers into inflamed tissue. Loss of sympathetic nerve fibers leads to generation of two distinct noradrenergic zones. In a zone with low concentrations of neurotransmitters (the red \(\alpha\)-adrenergic zone), only \(\alpha\)-adrenergic effects are possible because of the affinity of noradrenaline for the two receptor subtypes (high for \(\alpha\), low for \(\beta\)). However, in the vicinity of sympathetic nerve terminals, \(\alpha\)- and \(\beta\)-adrenergic effects can be expected (green \(\beta\)-adrenergic zone). Sympathetic nerve fibers on the healthy side of the demarcation line support \(\beta\)-adrenergic mechanisms such as release of free fatty acids (FFAs), whereas on the other side of the demarcation line, norepinephrine supports proinflammatory \(\alpha\)-adrenergic signaling and pain induction via \(\alpha_2\)-adrenergic receptors on nerve terminals of nociceptive neurons. In parallel, sensory nerve fibers sprout into inflamed tissue, leading to dissociation of innervation between sympathetic and sensory nerve fibers. The proinflammatory milieu is stabilized. The dissociation is a consequence of specific nerve fiber repulsion of sympathetic but not sensory nerve fibers. In the symptomatic phase of the disease, neurotransmitter-producing cells appear whose anti-inflammatory capacities are too small to overcome inflammation. CNS, central nervous system; FFA, free fatty acid; TNF, tumor necrosis factor.
or follicles (B cell). Thus, a proinflammatory milieu is established in secondary lymphoid organs, as in peripherally inflamed tissue.

**Role of Catecholamines in Antigen Transport to Secondary Lymphoid Organs and Immune Response**

After antigen capture, a further important aspect of inflammation is the transport of processed antigenic material to secondary lymphoid organs. Transport to lymphoid organs is mediated by lymphatic vessels, whose pumping efficiency is decreased by β-adrenergic pathways and is stimulated by α-adrenergic signaling. In addition, migration of antigen-loaded dendritic cells is stimulated via α1-adrenergic mechanisms. It is important to note that immature dendritic cells migrate upon α1-adrenergic influence, but CD40-stimulated mature dendritic cells do not (those that arrived in secondary lymphoid organs and encountered T cell contact via CD40–CD40 ligand). Thus, rapid establishment of an α-adrenergic zone in peripheral tissue is probably important in inducing migration of dendritic cells. In addition, substance P supports dendritic cell maturation and activity.

Catecholamine and its co-transmitter adenosine influence the direction of the immune response, whether T helper type 1 or type 2. Detailed in vitro experiments show that norepinephrine via β-adrenergic pathways inhibits T helper type 1 cell priming by inhibiting IL-12 and stimulating IL-10 of dendritic cells. The effects of catecholamines on the T helper type 17 immune response are not known. Tolerogenic effects of adenosine have been described. In addition to reactions on T cells, norepinephrine inhibits antigen presentation by epidermal Langerhans cells; this event is β-adrenergically mediated. Already in the late 1980s, it was demonstrated that surface expression of the antigen-presenting molecule human leukocyte antigen (HLA) class II was inhibited by β-adrenergic signaling.

In summary, the sympathetic nervous system has many inhibitory roles via β2-adrenergic and A2-adenosine receptors when T helper type 1 cell priming participates (e.g., in arthritis). The opposite occurs in a situation with T helper type 2 conditions because norepinephrine stimulates IL-4 and IL-10; this occurs along with many stimulating effects on B cells and antibody production (e.g., in systemic lupus erythematosus). Because norepinephrine and adenosine have additional strong inhibitory effects on secretion of TNF, interferon (IFN)-γ, and IL-2 via β-adrenergic and A2 adenosine pathways, the general inhibitory aspect of these neurotransmitters at high concentrations, along with vasoactive intestinal peptide (VIP) and CGRP, is well established.

At low concentrations of norepinephrine, when α1/2-adrenergic signaling is dominant, even stimulating effects on TNF occur. Thus, β-adrenergic influence in peripheral tissue and in secondary lymphoid organs should be reduced during proinflammatory T helper type 1 cell priming. Similar sympathetic nerve fiber loss and establishment of distinct α- and β-adrenergic zones belong to an adaptive process in secondary lymphoid organs to support or inhibit immune responses toward distinct antigens.

**Clonal Expansion of Aggressive and Regulatory T and B Cells**

The antiproliferative effects on T cells of norepinephrine via β-adrenergic receptors have been documented by many investigators. The proliferative response of CD8+ T cells is inhibited to a greater extent than that of CD4+ T cells, presumably because CD8+ T cells have a greater number of β-adrenergic receptors, and this effect is mediated via inhibition of IL-2 secretion. A proliferative effect of norepinephrine via β-adrenergic receptors is known for B cells. Similarly, the proliferative effect of substance P on T and B cells is common knowledge. The supportive effect of norepinephrine on antibody production has been demonstrated many times. These dichotomous effects of norepinephrine shape the immune response induced by T helper type 1 or type 2 priming antigens or autoantigens.

Catecholamine effects depend on the stage of T or B cell activation because naïve cells are influenced in different ways as compared with mature antigen-selected cells. Thus, timing of the neurotransmitter influence is mandatory. In addition, it is not clear how these neurotransmitters/neuropeptides act on the effector or tolerant version of T cells or B cells. Conflicting data exist on effects of norepinephrine on CD4+CD25FoxP3 regulatory T cells, and data for aggressive/regulatory B cells are not known. The present stage of knowledge does not really allow us to make a final statement as to how the sympathetic nervous system in secondary lymphoid organs supports aggressive or regulatory T and B cell responses. Experiments of the 1990s did not separate effects on aggressive or regulatory immune cells.

In summary, effects of the sympathetic nervous system depend on the given immune stimulus and on whether a T helper type 1 or type 2 cell response or involvement of B cells is prevailing. Many in vitro and in vivo studies indicate that norepinephrine supports the T helper type 2 cell immune response and the B cell response, but it suppresses the T helper type 1 cell response. Substance P, on the other hand, does not demonstrate such a dichotomy. For acetylcholine, acting through many receptor subtypes, detailed immunologic experiments have not been performed.

**Resolution of Inflammation and Tissue Repair**

Upon clearance of a pathogen, resolution of inflammation or reconstitution of normal tissue is the final step. Inflammation often leads to a preponderance of sensory nerve fibers (sensory hyperinnervation) over sympathetic nerve fibers, which are reduced in inflamed areas. In acute wounds, both nerve fibers disappear, but reappearance of sensory nerve fibers seems to start earlier than reinnervation with sympathetic nerves (Table 29-2). In general, reinstillation of sympathetic nerve fibers is a very long process, as substantiated in transplanted organs (>4 weeks), after tubal nerve crush (8 to 12 weeks), after chemical sympathectomy in the spleen (3 to 8 weeks), and after monophasic arthritis (4 to 8 weeks).

In a typical wound reaction, substance P promotes wound healing responses, and catecholamines have negative and positive effects on wound healing, such as inhibition of...
wound macrophages/neutrophils but support of later re-epithelialization (see Table 29-2). Moreover, stressful events that release sympathetic neurotransmitters and glucocorticoids lead to wound healing problems. From this point of view, a preponderance of substance P–positive nerve fibers over sympathetic nerve fibers would be favorable. Sensory hyperinnervation is probably supportive.

**CHRONIC INFLAMMATORY DISEASE**

**Neuronal Influences on Chronic Inflammatory Disease in Animal Models**

Chronic inflammatory disease occurs when inflammation fails to resolve and tissue repair is inadequate. The neuronal elements can contribute to this process. Most studies investigated the role of the sympathetic nervous system in the adjuvant arthritis model in Lewis rats.

The proinflammatory role of substance P and sensory nerve fibers in this model was demonstrated early in the 1980s. Additionally, substance P is proinflammatory for human synovioocytes and monocytes. Nociceptive fibers in the draining dorsal lymph nodes must play a critical role during this induction phase because local capsaicin treatment of these lymph nodes markedly decreases disease severity. Although the proinflammatory effects of substance P and other tachykinins are widely known, substance P–antagonistic therapies were not effective; this is probably a result of the redundancy in the tachykinin system.

In experiments conducted to study the role of the sympathetic nervous system in adjuvant-induced arthritis, most studies focused on a period of 14 to 40 days. From these studies, it is evident that overall peripheral sympathectomy or blockade of adrenoceptors (particularly via β-adrenoceptors) before or at the time of injection of Freund’s adjuvant diminishes the severity of joint inflammation during the entire observation period of 40 days. Similarly, the sympathetic nervous system plays an aggravating role in the collagen-induced arthritis (CIA) model, which is an autoantigen-driven chronic inflammatory disease. This might result from increased CD4+CD25+FoxP3+ T cells, as was recently demonstrated. When the sympathetic nervous system was destroyed before immunization and up to day 18 after immunization, sympathectomy markedly reduced the severity of arthritis. However, when the sympathetic nervous system was destroyed after outbreak of the disease, sympathectomy strongly aggravated arthritis.

The surprising dual role of the sympathetic nervous system might be explained as follows. In the very early induction phase after injection of adjuvant/antigen, mobilization, migration, and chemotaxis of proinflammatory cells such as neutrophils, NK cells, and monocytes directed to the site of adjuvant/antigen injection play a dominant role. In addition, targeted destruction of sympathetic nerves in secondary lymphoid organs supports antigen presentation and the switch to an aggressive effector immune response. Under conditions with sympathectomy in secondary lymphoid organs, arthritis becomes more severe owing to improved antigen presentation, stronger T helper lymphocyte type 1 immune reactions (aggressive phenotype for tissue-specific autoantigens), and probably downregulation of several regulatory elements such as IL-4 and IL-10 (see phase 2). Because the effect of prior sympathectomy is long lasting, these initial events are important for later inflammatory disease, which has also been demonstrated in atopic dermatitis and experimental colitis. However, in the chronic phase of the disease, the influence of the sympathetic nervous system largely changes.

One of the main changes is loss of sympathetic nerve fibers in inflamed tissue and in secondary lymphoid organs, as was already discussed (phase 2, Figure 29-5). Nerve fiber loss, starting with onset of disease, turns the essentially anti-inflammatory influence of sympathetic neurotransmitters at high concentrations into a proinflammatory influence at low concentrations (see Figure 29-5). In addition, recently described tyrosine hydroxylase–positive cells with anti-inflammatory capacities appear in lymphoid organs and arthritic tissue. In chronic inflammatory disease, the number of these cells in secondary lymphoid organs increases over time, and they appear shortly after disease outbreak in the inflamed joint. Because these cells can be eliminated by 6-hydroxydopamine treatment (the sympathectomy technique), the anti-inflammatory influence of these cells is soon lost after experimental chemical sympathectomy. Loss of tyrosine hydroxylase–positive cells probably leads to an overall proinflammatory situation, because these cells might have tolerogenic activities. The quite different effects of this sympathectomy tool are now explained by early destruction of sympathetic nerve fibers, which are proinflammatory (phase 1), and later destruction of anti-inflammatory catecholamine-producing cells in phase 3, which are anti-inflammatory.

Finally, the influence of the parasympathetic nervous system has attracted increasing interest. The alpha7 subunit of the nicotinergic acetylcholine receptor is especially

**Table 29-2**  Behavior of Nerve Fibers and Their Neurotransmitters/Neuropeptides in Wound Reactions*

<table>
<thead>
<tr>
<th>Nerve Fiber Type</th>
<th>Change during Wound Reactions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensory nerve fibers</td>
<td>Sensory nerve fibers are lost after 2 days but reappear after approximately 7 to 14 days</td>
<td>266-268</td>
</tr>
<tr>
<td></td>
<td>Substance P and calcitonin gene-related peptide promote wound healing</td>
<td>269-273</td>
</tr>
<tr>
<td>Sympathetic nerve fibers</td>
<td>Fast loss of sympathetic nerve fibers and reappearance after approximately 14 days</td>
<td>274</td>
</tr>
<tr>
<td></td>
<td>Catecholamines block wound repair via β-adrenoceptors</td>
<td>275, 276</td>
</tr>
<tr>
<td></td>
<td>Norepinephrine inhibits wound macrophages and neutrophils</td>
<td>277, 278</td>
</tr>
<tr>
<td></td>
<td>Catecholamines support later re-epithelialization</td>
<td>279-282</td>
</tr>
<tr>
<td></td>
<td>Adenosine supports the wound healing response via A2 receptors (mediated through increase in fibrosis)</td>
<td>283, 284</td>
</tr>
</tbody>
</table>

*Experiments with 6-hydroxydopamine, the sympathetic nerve fiber toxic substance, are not included because this substance affects not only sympathetic nerve fibers.
relevant in the regulation of inflammatory responses; this led to the concept of the cholinergic anti-inflammatory reflex. Additional experiments with agonists of the alpha7 subunit of the nicotinic acetylcholine receptor demonstrated the anti-inflammatory importance of this cellular pathway in animal experiments and human cells. It is important to note that this particular nicotinic receptor is highly expressed on macrophages and fibroblasts of patients with RA.

At the moment, it is unclear how the vagus nerve influences synovial inflammatory disease. Four different possibilities exist regarding how favorable cholinergic effects on joint inflammation can be explained: (1) The cholinergic influence supports sympathetic inhibition of splenic proinflammatory immune responses, (2) the cholinergic influence directly affects cells in draining lymph nodes of the trunk, (3) the cholinergic influence affects cells in the gut (which play an important role as substantiated in the HLA-B27 rat model of arthritis), and (4) nonneuronal acetylcholine release appears within inflamed synovial tissue.

**Neuronal Influence on Endothelial Cells and Angiogenesis**

Angiogenic effects of sympathetic neurotransmitters have been demonstrated in tumor models and in tumor cells. Because tumor cells often demonstrate quite different signaling pathways, generalizability of these findings for different forms of inflammation might be critical. Nevertheless, studies in this field are the only source of information. Norepinephrine has been implicated in angiogenesis by inducing vascular endothelial growth factor (VEGF) expression in tumor cells via beta-adrenergic effects, while direct trophic effects on endothelial cells were demonstrated via alpha1-adrenergic effects. These effects were potentiated by hypoxia. In addition, the sympathetic co-transmitter neuropeptide Y (NPY) has angiogenic activities shown in animals deficient in the NPY receptor type 2. It is important to note that dopamine via D2 receptors has universal inhibiting effects on angiogenesis.

As long as sympathetic nerve fibers are present, alpha- and beta-adrenergic and NPY effects are possible. When nerve fibers are lost, these typical effects of sympathetic neurotransmitters/neuropeptides can be expected only in border areas along the demarcation line in Figure 29-5 with still existing normal innervation, because tissue concentration would be high enough. In an area of lost sympathetic nerve fibers and replacement by catecholamine-producing cells, mainly alpha-adrenergic and dopaminergic effects predominate, because concentrations of norepinephrine are low and dopamine is the main neurotransmitter produced. Thus, angiogenesis might be supported by sympathetic neurotransmitters in the healthy border zone of inflammation but not in the middle of the inflammatory zone. Nevertheless, in the sympathetic alpha-adrenergic zone, sensory hyperinnervation exists, so that higher levels of substance P can be expected.

Substance P was demonstrated to support capillary growth in vivo in a rabbit cornea model and in a rat sponge assay via NK1 receptors. In addition, substance P stimulates proliferation of different endothelial cell types. In vivo experiments showed that endogenous substance P could be implicated in neoangiogenesis connected with inflammation. Thus, the two neurotransmitter systems of catecholamines/NPY and substance P probably influence angiogenesis in inflammation.

**Neuronal Influences on Fibroblasts and Adipocytes**

Sympathetic neurotransmitters modulate the function of fibroblasts by inducing proliferation, collagen gene and protein expression, and fibroblast migration via alpha1-adrenergic receptors. In contrast, fibroblasts undergo increased apoptosis and autophagy via beta-adrenergic signaling. In addition, norepinephrine induces secretion of IL-6, IL-8, and matrix metalloproteinase 2 from fibroblasts via beta-adrenoceptors. Under conditions with an alpha-adrenergic zone due to nerve fiber loss (see Figure 29-5), one can expect proliferative responses of sympathetic nerve fibers on fibroblasts. This can support the fibrotic process in chronic inflammatory lesions.

The proliferative effects of substance P on fibroblasts are well documented. Substance P supports the growth-promoting effects of IL-1 in cultured fibroblasts. These substance P effects were mediated through NK1 receptors. In addition, substance P induces migration of human fibroblasts in vitro. Although substance P demonstrates a proliferative effect on fibroblasts, CGRP has no similar role, but it stimulates fibroblast IL-6 secretion. In conclusion, establishment of an alpha-adrenergic zone together with sensory hyperinnervation induces a proliferative effect.

In recent years, adipocytes in the proximity of inflammatory lesions have gained enormous interest because of their proinflammatory activities. These fat cells might be important targets of neuronal influences to support the inflammatory process. Indeed, it is well known that the sympathetic nervous system via beta-adrenergic pathways is instrumental in releasing energy-rich free fatty acids that are used by different immune cells as energetic substrates (recently reviewed in the context of chronic inflammation in Straub et al). It is important to note that the fat tissue in the proximity of inflamed lesions is perfectly innervated by sympathetic nerve fibers. Thus, norepinephrine is present in adequate amounts to stimulate lipolysis via beta-adrenoceptors (see Figure 29-5). Alpha2-Adrenergic stimulation leads to inhibition of lipolysis. Thus, the balance between lipolysis-promoting beta-adrenoceptor activation and lipolysis-inhibiting alpha2-adrenoceptor activation dictates the degree of lipolytic activity and provision of energy-rich free fatty acids to consumers. In addition, the sensory neuropeptide substance P supports proliferation and antiapoptotic pathways in fat depots that might contribute to the development of inflammatory fat accumulation, which has been demonstrated in vitro and in vivo.

**Neuronal Influences on Osteoclasts and Osteoblasts**

Neurotransmitters of the sympathetic nervous system (catecholamines) and neuropeptides (vasoactive intestinal peptide, substance P, and CGRP) play a role in normal bone homeostasis. It is very important to distinguish the
different phases of osteoblast and osteoclast differentiation because the influence of neurotransmitters/neuropeptides depends on the differentiation stage. Catecholamines inhibit bone formation via \( \beta_2 \)-adrenergic receptors by stimulating osteoclast differentiation and by inhibiting osteoblast function.\(^{220,226} \) This can be opposed by agonists of \( \alpha \)-adrenoceptors.\(^{207,208} \) Because noradrenaline has a higher binding affinity for \( \alpha \)-adrenoceptors than for \( \beta \)-adrenoceptors, the local concentration of this neurotransmitter most probably determines the effect (low concentrations can support bone formation, high concentrations inhibit bone formation).

Vasoactive intestinal peptide and CGRP support bone formation by inhibiting osteoclastogenesis and by stimulating osteoblasts;\(^{209,210} \) this is different for substance \( P \), which preferentially stimulates osteoclasts and thus bone resorption.\(^{211,212} \) All these mechanisms have been detected in cells of healthy subjects or normal animals and/or in cell lines. No such studies have been carried out in primary osteoblasts or osteoclasts from arthritic animals or patients with arthritis. This must be the subject of future studies because in chronic inflammation, the situation might be quite different.

**CHANGES OF THE NERVOUS SYSTEM IN PATIENTS WITH CHRONIC INFLAMMATORY DISEASES**

**Increased Activity of the Sympathetic Nervous System**

Several studies have demonstrated that patients with chronic inflammatory diseases have an elevated sympathetic nervous system tone.\(^{213-216} \) Increased sympathetic activity could be related to increased risk of cardiovascular events, as has been observed in patients with RA.\(^{217} \) Such an increased sympathetic tone may be a consequence of relatively decreased serum levels of cortisol in relation to inflammation, because there exists cooperativity of cortisol and norepinephrine on a molecular level via the \( \beta \)-adrenergic receptor signaling cascade.\(^{218,219} \) Functional loss of one factor probably upregulates the other factor to maintain functions such as blood glucose homeostasis, regulation of the bronchial lumen, blood pressure control, and others. Because TNF is relevant for adaptation of the HPA axis leading to inadequately low cortisol secretion, its neutralization may change the increased sympathetic tone.

A recent study confirmed increased sympathetic tone in patients with RA and also in those with SLE; this was accompanied by relatively normal tone of the HPA axis (called “uncoupling of the HPA axis and the sympathetic nervous system” during chronic inflammation because an increase in the tone of both axes can be expected during acute inflammation).\(^{220} \) It was found that 12 weeks of anti-TNF therapy only slightly decreased sympathetic activity as measured by plasma neuropeptide \( Y \) levels.\(^{222} \) Thus, uncoupling persists after treatment, and it appears that TNF is not the sole factor responsible for this phenomenon. A similar increase in sympathetic activity has been demonstrated in patients with Crohn’s disease and ulcerative colitis.\(^{221} \)

It should be mentioned that elevated activity of the sympathetic nervous system might not cause increased local sympathetic neurotransmitters in the inflamed joint because sympathetic nerve fibers are lost.\(^{157} \) In addition to increased sympathetic nervous tone, one observes a decrease in parasympathetic outflow.\(^{151} \) Such a decrease in the parasympathetic system is probably an unfavorable signal because this will impede the anti-inflammatory activity of the vagus nerve.\(^{222} \)

**Loss of Sympathetic Nerve Fibers and Sprouting of Sensory Nerve Fibers**

Loss of sympathetic nerve fibers and sprouting of sensory nerve fibers have already been described in Figure 29-5. Loss of sympathetic nerve fibers has been described in the synovial tissue of patients with RA,\(^{167,221-223} \) in the oral mucosa of patients with lichenoid reactions,\(^{226} \) in the inflamed Charcot foot,\(^{227} \) in inflammatory endometriosis,\(^{228} \) in chronic pruritus and prurigo nodularis,\(^{229} \) in Crohn’s disease,\(^{165} \) and in other cases. Sprouting of substance \( P \)-positive sensory nerve fibers has been observed in gastritis,\(^{230} \) esophagitis,\(^{231} \) and psoriatic skin disease,\(^{232} \) and in the synovial tissue of patients with RA compared with osteoarthritis,\(^{225} \) Charcot foot,\(^{227} \) chronic pruritus and prurigo nodularis,\(^{229} \) Crohn’s disease,\(^{165} \) and other conditions. A marked preponderance of substance \( P \)-positive nerve fibers over CGRP-positive nerve fibers has been noted in RA synovial tissue; this supports inflammation in that CGRP would have some anti-inflammatory properties.\(^{233} \) Generally, all these findings support the concept of a nonspecific reaction that favors inflammation.

This concept is supported by additional results indicating that in later stages of inflammatory joint disease, \( \alpha_1 \)- and \( \alpha_2 \)-adrenergic receptors seem to gain a more prominent role. In patients with juvenile chronic arthritis, functional \( \alpha_1 \)-adrenergic receptors on leukocytes are induced, whereas the receptors are absent on the leukocytes of normal donors.\(^{15} \) These effects seem to be cell type specific in that vascular \( \alpha \)-adrenergic receptor–mediated vasoconstriction in inflamed joints is downregulated, leading to elevated perfusion.\(^{234,235} \) Suppression of \( \alpha \)-adrenergic receptors on vasoconstrictors and parallel induction on leukocytes would serve as a proinflammatory stimulus because this leads to vasodilatation (vessels), proliferation of lymphocytes, and secretion of proinflammatory cytokines such as IL-6 and TNF (leukocytes).

AP-1- and NF\( \kappa \)B–binding sites are present in the regulatory region of the \( \alpha_1 \)-adrenergic receptor promoter that may be responsible for the induction of \( \alpha_1 \)-adrenergic receptor expression in monocytes by IL-1\( \beta \) and TNF.\(^{156} \) In another study, synovial fibroblasts of patients with RA could be activated to proliferate by \( \alpha_2 \)-adrenergic agonists; this process is mediated via phospholipase C, protein kinase C beta II, and mitogen-activated protein (MAP) kinase.\(^{237} \) Ex vivo studies with peripheral blood mononuclear cells from RA patients revealed that only in patients with high disease activity did catecholamines mediate their effects via \( \alpha \)-adrenergic receptors.\(^{238} \) In addition, peripheral blood mononuclear cells of patients with RA demonstrate lower numbers of \( \beta \)-adrenergic receptors, supporting the overall \( \alpha \)-adrenergic preponderance.\(^{16} \)

From these data, it seems as though \( \alpha \)-adrenergic signaling becomes more relevant in the later stage of this chronic
inflammatory disease. This is referred to as the “beta-to-alpha-adrenergic shift” during disease progression (see Figure 29-5). 

**Cells Positive for Neurotransmitters/Neuropeptides Appear in the Tissue**

In recent years, several reports have described the presence of immune cells in inflamed tissue with the capacity to produce neurotransmitters or neuropeptides. These studies demonstrated the production of substance P from RA and osteoarthritis synovial fibroblasts. Production of this neuropeptide was linked to a proinflammatory situation.

RA fibroblast-like synoviocytes and macrophages possess the enzyme machinery for catecholamines. The production of catecholamines seems to exert anti-inflammatory effects in synovial cells of patients with RA. Catecholamine production was linked to a tolerogenic phenotype of T lymphocytes in patients with multiple sclerosis. Cells in inflamed lesions can also secrete endogenous opioids with anti-inflammatory activities. Even the production machinery for acetylcholine exists in cells of the synovial tissue and might also be an anti-inflammatory signal.

The role of these cells is not clear, but it seems that the anti-inflammatory activities of catecholamine-, opioid-, and acetylcholine-producing cells are insufficient to overcome the proinflammatory domination. It needs to be investigated whether a change in neurotransmitter/neuropeptide secretion can be achieved by therapeutic drugs, leading to higher production of anti-inflammatory neurotransmitters/neuropeptides.

**SUMMARY**

Various neuronal systems exhibit both proinflammatory and anti-inflammatory roles, depending on many influences: (1) point in time in relation to the start of the inflammatory process, (2) antigenic stimulus shifting the immune response, (3) tissue location and environment, (4) involved cell types and time-dependent receptor expression, (5) changing tissue innervation, and (6) appearance of neurotransmitter-producing cells. Thus, the general statement of “the neuronal anti-inflammatory reflex” is meaningless outside this context.

Three decades of experimentation and clinical investigation have demonstrated the proinflammatory role of the nervous system in chronic inflammatory disease (sensory vasoregulation, peripheral and central sensitization, neurotransmitter-induced chemotaxis and cell mobilization, sympathetic nerve fiber loss, sensory hyperinnervation, etc.). As the result of adaptive programs, the net effect is unfavorable in chronic inflammatory diseases with immunity against self-antigens or harmless foreign antigens. Unfortunately, many anti-inflammatory activities of the different nervous systems are switched off (β-adrenergic, adenosinergic, opioidergic, cholinergic, and others). Therapeutic targets that capitalize on the anti-inflammatory effects of the nervous system or neurotransmitters have considerable potential to modulate chronic inflammatory diseases.

**Selected References**


Full references for this chapter can be found on www.expertconsult.com.
References


OVERVIEW OF EPIDEMIOLOGIC METHODS

Epidemiology is the study of the distribution of disease and its determinants in populations. Its purpose is to describe the frequency of disease and to determine causes responsible for variation in disease occurrence. Comparison of the relative strengths of those causes and assessment of their generalizability can allow “truth” to be inferred. This chapter explains basic epidemiologic concepts and definitions; describes the major study designs, their strengths and weaknesses, and their usefulness in inferring causality; and demonstrates specific applications of these principles to the study of rheumatic diseases. For the purposes of this chapter, the term disease is used to represent a disease, death, or other health outcome of interest, and the term exposure is used to represent a risk or protective factor examined for its association with disease.

Measures of Disease Occurrence

Prevalence

The frequency of a disease in a population at any given time is its prevalence. It is measured at one point in time and is the proportion of individuals with a disease out of the total population under study. Importantly, the numerator makes no distinction between new and established cases of disease. Multiple estimates of prevalence over time are commonly used to determine trends in disease occurrence or need for health services.

Incidence

In order to determine the likelihood that disease will develop over time, repeated observations of the same people are required to determine who develops disease and who does not. Incidence proportion, or risk, is the frequency of new cases over a specified time, out of those at risk for, but without, the disease at baseline. During the observation period, a person may be at risk for the disease but not develop it, and he or she would contribute time at risk for the entire period. Alternatively, a person may develop the disease in question, die from competing risks, or be lost to
follow-up. All of these situations result in that person’s no longer contributing time at risk to the denominator. The concept of person-time includes the actual time at risk contributed by each individual. For example, consider the hypothetical example of incidence of systemic lupus erythematosus (SLE) over a 10-year period (Figure 30-1). Person A may not develop the disease during the observation period but may die at Year 5 of a competing cause; this person contributes 5 person-years to the denominator. Person B might develop SLE 4 years after the study begins and thus is no longer at risk of developing the disease; this person contributes 4 person-years of time at risk to the denominator. Person C joins the study at Year 2 and is lost to follow-up at Year 9, contributing 7 years of time at risk. Person D joins the study at Year 1 and is lost to follow-up at Year 5 for a total of 4 years of person-time at risk. Person E joins the study at Year 2 and develops SLE halfway between Year 8 and Year 9, contributing 6.5 person-years at risk. Incidence rate is defined by the following:

\[
\text{Incidence rate} = \frac{\text{New cases developing over time period of observation}}{\text{Total person-time at risk for each individual without the disease at study entry}}
\]

In the example, two new cases of SLE are observed and the total person-time from persons A through E at risk is 5.0 + 4.0 + 7.0 + 4.0 + 6.5 = 26.5 person-years. Thus the incidence rate is 2 cases/26.5 person-years, or 1 case/13.25 person-years. This may be expressed as 0.0754 cases/person-year or 7.54 cases/100 person-years.

**Measures of Effect**

More important than just the description of frequency of disease or its development is the relationship between the disease and exposures to potentially causative factors. One way to examine this is to compare the prevalence or incidence of disease in groups with a given exposure compared with those without that exposure. Critical to the ability to assess potential causality of an exposure/disease association is that the exposed and unexposed groups must be comparable. Measures to delineate this relationship between disease occurrence and exposure vary according to study design. Cross-sectional surveys and case-control studies use the odds ratio, which is a measure of the odds (prevalence odds = prevalence/1-prevalence; incidence odds = incidence/1-incidence) of disease in the exposed group compared with the unexposed group. Longitudinal designs can calculate a relative risk (ratio of risks) of the incidence of disease in the exposed compared with the unexposed groups.

**STUDY DESIGNS**

These include ecologic studies, cross-sectional surveys, case-control studies, cohort studies, and randomized controlled clinical trials—the last frequently considered the most rigorous study design and the one most closely representing a formal experiment. Each study design has its own inherent strengths and weaknesses (Table 30-1), and the choice of study design depends on the research question, the rarity of the disease under study, the availability of appropriate study and comparable control populations, resources available to conduct the study, and logistics.\(^2\)

**Observational Studies**

In observational studies the exposure is not randomly distributed in a population. The investigator observes the exposure rather than selects the exposure status of an individual.\(^4\) Types of observational studies include ecologic, cross-sectional, case-control, and cohort.

**Ecologic Studies**

In the ecologic study design, the unit of observation is a group, rather than an individual.\(^5\) Aggregate data on rates of disease and risk factors are compared to examine associations between disease frequencies and exposures. The ecologic study is frequently a design of expediency and can
generate hypotheses for more rigorous testing in studies using individual-level data. One of its chief drawbacks is its high susceptibility to confounding. This occurs when an extraneous factor, not on the causal pathway, masks the true relationship between exposure and disease, by virtue of its association with both. Further, associations in the aggregate may not necessarily hold for the individual. This concept is termed the ecologic fallacy. As a hypothetical example, rates of specific kinds of cancers may be higher in countries in which cigarette sales are also high. Whether those who are buying, and presumably smoking, the cigarettes are the same persons who develop cancer is not known from this study design.

Cross-Sectional Surveys

The goal of this study design is usually descriptive including all individuals, with and without the disease under study, in the population or a representative sample of them, at one point in time with no follow-up period. Surveys can estimate prevalence of a particular disease in the population and determine the need for health services and resource

Table 30-1 Common Epidemiologic Study Designs and Their Strengths and Weaknesses

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Definition</th>
<th>Measure of Effect</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecologic</td>
<td>Aggregate data on exposures and disease; unit of analysis is a group, not an individual</td>
<td>Odds ratio</td>
<td>Inexpensive, Short duration, Hypothesis-generating</td>
<td>Susceptibility to confounding, Ecologic fallacy</td>
</tr>
<tr>
<td>Cross-sectional survey</td>
<td>Data on exposures and disease obtained at one time from all individuals in an area (or a sample thereof) with and without disease</td>
<td>Prevalence Odds ratio</td>
<td>Can study several outcomes, Short duration, Can generate population prevalence estimates of disease and risk factor distributions</td>
<td>May not be able to determine whether disease preceded exposure, Potential survivor bias, Not practical for rare diseases, Cannot produce incidence or relative risk estimates, Inefficient for rare exposures, Potential bias from sampling cases and controls separately, May not be able to determine whether exposure preceded disease, Potential recall bias, Potential survivor bias, Cannot produce prevalence or incidence estimates, Frequently requires large samples, Not feasible for rare outcomes, More expensive, Long duration</td>
</tr>
<tr>
<td>Case-control</td>
<td>Study of exposure/disease relationship in cases with a disease and controls without that disease, who are selected from source population from which the cases arose</td>
<td>Odds ratio</td>
<td>Best for studying rare conditions or those with long latency, Short duration, Small sample*, Inexpensive*, Odds ratio can approximate relative risk</td>
<td></td>
</tr>
<tr>
<td>Cohort</td>
<td>Individuals without disease are followed over time to determine which characteristics predict who will get the disease and who will not</td>
<td>Incidence Relative risk</td>
<td>Can determine sequence of events, Less susceptibility to survivor bias and bias in measuring predictors, Can study multiple outcomes, Can generate population incidence, relative risk</td>
<td>Investigator control over selection of participants and measures, Increased expense, Long duration</td>
</tr>
<tr>
<td>Prospective</td>
<td>Study sample selected by investigator and followed forward in time for development of disease</td>
<td>Incidence Relative risk</td>
<td>Investigator control over selection of participants and measures</td>
<td></td>
</tr>
<tr>
<td>Retrospective</td>
<td>Study sample and measurement of exposures and disease over time have already occurred</td>
<td>Incidence Relative risk</td>
<td>Less expense, Short duration</td>
<td>Less control over selection of participants and measures</td>
</tr>
<tr>
<td>Nested case-control</td>
<td>Case-control study within the context of a prospective or retrospective cohort</td>
<td>Incidence Relative risk</td>
<td>Underlying cohort design, Relatively inexpensive, compared with measurement on entire cohort, Most closely emulates an experiment, Strongest design to produce evidence for cause and effect, Random assignment of intervention minimizes confounding, May be faster and cheaper for some study questions than observational studies</td>
<td>May require bank of samples that can be assayed at later date until or after outcomes occur, Costly in time and money, Some research questions not suitable because of rare disease or ethical barriers, May not be generalizable if highly controlled environment does not reflect “real world” common practice, May have narrow scope and study question</td>
</tr>
<tr>
<td>Randomized clinical trial</td>
<td>Exposure (pharmaceutical, nonpharmacologic device, educational intervention) manipulated by investigator</td>
<td>Relative risk Hazard ratio</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Relative to cohort study design.

Modified from Hennekens CH, Buring JE. Epidemiology in medicine, Boston, 1987, Little, Brown, and Company; and Hulley SB, Cummings SR. Designing clinical research: an epidemiologic approach, Philadelphia, 1988, Williams & Wilkins.
allocation. Typically, information about risk factors is obtained simultaneously. Such risk factor data may or may not represent the most relevant time of exposure, nor can it be determined whether the exposure preceded or resulted from the disease.

An example of a cross-sectional survey, conducted approximately once per decade in the United States, is the National Health and Nutrition Examination Survey. This survey samples a proportion of the residents in the contiguous 48 states and measures various health outcomes and habits such as blood pressure, serum lipids, height, weight, smoking, and dietary intake. These surveys have been used in rheumatology to determine the prevalence of radiographic knee and hip osteoarthritis (OA) in various age, sex, and race/ethnicity subgroups.

Case-Control Studies

Much maligned by the uninitiated because of its susceptibility to bias, the case-control study can be the study design of choice—or sometimes the only appropriate study design—in certain situations, particularly when the disease under study is rare. Usually, it includes fewer individuals—at much lower cost and higher efficiency—than would be required for a cohort study because it begins with individuals who already have the disease in question, rather than waiting for a small proportion of a large cohort to develop the disease over time. Most important in the design of a case-control study are (1) the choice of the control group, which must be comparable to the cases, and (2) recognition of potential biases that may threaten validity.

Strictly defined, the case-control study is a study in which those with the disease (cases) are compared with a control population without the disease, drawn from the same source population from which the cases arose. The source population may be the residents of a particular geographic area or a hospital’s referral base. The control group serves as an estimate of the distribution of the exposure in the source population, and consequently, the control group must be sampled independently of exposure status. For example, if one is interested in examining the possible association between smoking and progressive systemic sclerosis (PSS), the controls must be from the same source population that generated the cases, if this can be determined, and must be sampled without regard to their smoking status.

Selection of Controls for Case-Control Study

If the source of the cases is a well-defined population, the controls can be sampled directly from that population. If the source population is too large to allow a complete enumeration, controls may be matched to each case by their residence in the same neighborhood. Random-digit dialing can be used to select controls, but this labor-intensive method omits from selection those without telephones or those who cannot be reached. If the cases are drawn from a particular hospital or clinic, then the source population should represent people who would be treated in that hospital or clinic if they developed the disease under study, but frequently, this source population can be difficult to identify and is influenced by referral practices. Hospital or clinic controls can be used, but this method can have particular pitfalls because the controls might not be selected independently of the exposure in the source population. For instance, in a hospital-based study of smoking in SLE, individuals hospitalized for other diseases such as myocardial infarction or pneumonia might have exposures different from the source population in general, especially if the exposure, in this case smoking, causes or prevents the “control” disease selected. One way to avoid this is to exclude diseases known to be associated with the exposure under study, but this may create other biases. Another tactic could be to select hospital controls with diseases that are felt to be unrelated to the disease or exposures under study such as traumatic leg fractures or to use several control groups selected differently. The latter example might sample controls from hospitalized patients with other diseases than the disease under study, nonhospitalized patients in the same medical care system, or nonhospitalized individuals in the general population, comparing each control group separately with the diseased group.

Weaknesses of the Case-Control Design

It is not possible to derive incidence or prevalence estimates from a case-control study. The greatest threat to validity is the inherent susceptibility to bias that can exist in this study design, because the cases and controls are sampled separately and the assessment of exposure variables is retrospective. Matching the cases and controls on factors such as age, sex, or race/ethnicity can help ensure comparability of cases and controls to a degree. As mentioned earlier, more than one control group, selected in different ways, can be used to see if findings are consistent across control groups with different sampling biases. A nested case-control design, in which a case-control study is performed within a larger cohort study, has the advantage of minimizing sampling bias because the cases and the controls would have been previously sampled in identical fashion into the parent cohort study.

The other chief source of bias in the case-control study is recall bias, which occurs when exposures predating the disease may be differentially reported by the controls and the cases, the latter of whom may have incentive to remember and report exposures. This can be partially prevented by using exposure data measured before the disease occurred, if available, and by blinding the observer and the subject to the exposure under investigation or, if possible, blinding them even to the specific disease under study and therefore to case or control status. For example, in a case-control study examining racial/ethnic variation as the exposure variable of interest in SLE, race/ethnicity is immutable and therefore not subject to recall bias. In contrast, if study participants know or suspect that prior exposure to hair dye, for instance, is the exposure of interest in the same case-control study, those with disease may be more prone to “remember” their exposure than might those without disease. Investigators can obtain information about multiple potential exposures or even include several "dummy" exposures to mask the real hypothesis to try to minimize this type of bias.
Cohort Studies

Cohort studies follow groups of individuals without the disease in question over time to describe the development or incidence of disease and to compare the incidence of disease among groups with different risk factors or exposures. Cohorts can be prospective or retrospective.

Prospective Cohort Study

Prospective cohorts are characterized by the selection of the cohort and measurement of risk factors or exposures before the outcome has occurred, thereby establishing time sequence or temporality, an important factor in determining causality. This is a distinct advantage over the case-control study in which exposure and disease are assessed simultaneously.

The primary disadvantage to the prospective cohort study is its expense, in that it requires large numbers of individuals followed for potentially long periods of time. Biases can creep in, particularly if there is significant loss to follow-up. This study design is highly inefficient and inappropriate to study rare diseases, but its efficiency increases as the frequency of the disease in the population increases.

For example, a prospective cohort study would be inappropriate to study PSS because of its rarity but excellent to study a common condition such as OA.

Retrospective Cohort Study

In a retrospective cohort, individuals are followed over time, but the cohort selection and collection of data have already occurred, sometimes for a different purpose than the current disease under study. For example, a cohort of individuals with small vessel vasculitis seen at a particular hospital between 1990 and 1992 could be identified, and data regarding baseline serologies, physical examination findings, and biopsy results when the patients were first evaluated could be abstracted. Then examination of outcomes such as stroke or development of dialysis-dependent renal disease could be ascertained in 2000, by medical record review or by re-contact with the individuals so identified. Because exposure or risk factor assessment precedes assessment of outcome, this study design can establish temporality, as in a prospective cohort, and is less subject to recall bias that can plague case-control studies. By selecting the cases and controls from the same source population, this study design avoids some of the selection biases of case-control studies in which the cases and controls are sampled separately. The retrospective cohort design is cheaper and more efficient than a prospective cohort, but because the data collection has already occurred, inferences from such a study are highly dependent on the quality, completeness, and appropriateness of the original risk factor assessments to study their association with the disease in question.

Nested Case-Control Studies

These studies are case-control studies that occur within the context of a prospective or retrospective cohort and are particularly useful in the assessment of risk factor variables that would be too expensive to measure on all members of the cohort. In this design, members of a cohort who have developed a particular outcome during the observation period are selected and compared with a sample of individuals within that same cohort who have not developed the outcome. Then, for example, stored biologic specimens from baseline may be assayed for an exposure of interest such as vitamin D level and compared between those who developed the disease and those who did not.

Registries

Individual occurrences of a disease can be obtained from multiple sources within a defined geographic area to form a disease registry. The data from these sources are linked to avoid duplication of cases. Registries may be based on population, hospitals, or clinics. Hospital- and clinic-based registries may identify potential participants for clinical research. Examples of registries collecting longitudinal population-based data on rheumatic diseases include the National Data Bank for Rheumatic Diseases (www.arthritis-research.org) and the Arthritis Internet Registry (www.arthritis.org/arthritis-internet-registry.php).

Clinical Trials

The study designs described previously in this chapter were all observational designs; there was no experimental manipulation of the exposure or outcome. Experimental study designs or interventions include clinical trials, field trials, and community intervention trials. Inferences from such trials of treatments assigned randomly to a large enough sample are much less likely to suffer from biases and other threats to validity than are observational designs. Randomization in theory should eliminate most confounding, although some variation in risk factors between the intervention groups may occur by chance and should always be ascertained and addressed in the analysis if necessary. The validity of conclusions from a clinical trial depends in part on the avoidance of loss to follow-up or participant dropout.

Clinical trials can be conducted for pharmacologic or nonpharmacologic interventions such as dietary, physical activity, assistive devices, or educational interventions. Trials can include single or multiple dosages of the study intervention, placebo controls, active comparator controls in which the intervention of interest is compared with another agent whose efficacy is known, and combinations of interventions. For example, the Glucosamine/chondroitin Arthritis Intervention Trial (GAIT) compared glucosamine hydrochloride alone, chondroitin sulfate alone, and the combination of glucosamine and chondroitin with placebo and with an active comparator, celecoxib, for their effects on symptoms of OA of the knee. The Arthritis, Diet, and Activity Promotion Trial (ADAPT) was a nonpharmacologic intervention in which diet, exercise, and the combination of diet and exercise were compared with a control group. Such nonpharmacologic trials may include an “attention control” in which the control group does not get the specific intervention of interest but does get at least a minimal amount of attention from the investigator because...
it is known that even minimal contact with the participants in a study can improve outcomes.\textsuperscript{13}

Optimally, to minimize bias, the study should be double-blind, in which the assignment of treatment is unknown to the participant and to the data collector evaluating the participant’s response. A cross-over design is a within-patient design, which allows each participant to be his or her own control and receive either the active intervention followed by a “washout” period in which no active or inactive treatment is given, and then the control treatment, or vice versa. This design has some advantages, particularly in sample size requirements, but can be biased if there is a significant carry-over effect of the active treatment into the “control” observation period.\textsuperscript{2} Response to treatment may also differ depending on whether the active drug is received before or after the placebo or other comparator.\textsuperscript{14}

Other important considerations in clinical trials are the selection and means of assessment of primary and secondary outcomes, which must be prespecified. Outcomes can include measures of disease modification, symptom modification, and frequency of side effects or other poor outcomes. Symptom modification trials are frequently of short duration and less expensive than disease modification trials, which are generally interested in longer-term outcomes. In trials of biologics for rheumatoid arthritis, for instance, effects on symptoms can frequently be measured in weeks to months, whereas effects on prevention or healing of radiographic erosions may require longer follow-up times.\textsuperscript{15} Similarly, disease modification trials in OA currently require large numbers of individuals followed for at least 2 years, predominantly because the metric of change in minimal joint space on knee radiographs is imprecise and can be fraught with measurement error.\textsuperscript{16} It is expected that outcomes based on magnetic resonance imaging, once validated, will be more sensitive, less subject to measurement error, and require smaller sample sizes and shorter observation periods to demonstrate an effective response.\textsuperscript{17}

Other trial designs can apply interventions to entire communities or to health care workers with measurement of outcome in their patients. An example of the latter would be an educational intervention designed to increase physician prescription of physical therapy evaluation of all patients with knee or hip complaints. The physicians receive the intervention, but whether the physical therapy is prescribed or whether it improves patient symptoms is measured by assessing the patient.

Although clinical trials represent the “ultimate” study design closest to a controlled experiment, there are significant potential threats to its validity. One of the most important biases can occur when there is large loss to follow-up. In order to minimize this type of bias, every effort should be made to continue to obtain outcome information on all participants, even those who otherwise discontinue study assignment to therapy. Because all predictors of dropout cannot be known and because dropouts may differ from individuals who remain in a study in ways that cannot be controlled, conventional analyses of treatment status are likely confounded.\textsuperscript{2} Data may be analyzed in an intention-to-treat fashion, in which all randomized participants are analyzed as a member of the group to which they were initially randomized, regardless of whether they actually adhered to the group assignment, but this analytic method tends to be biased because noncompliance leads to a misclassification of treatment status.\textsuperscript{3} One method for dealing with treatment noncompliance suggested by Mark and Robins\textsuperscript{18} includes making the assigned treatment a fixed covariate and received treatment a time-dependent exposure in a structural failure-time model. Completer, or “according to protocol,” analyses are often also performed, in which only those who adhered to their assigned group treatment are included in the analysis. Prerandomization screening and run-in periods before randomization can help to avoid randomizing those unlikely to adhere to or complete the protocol, thereby minimizing expense and dilution of effects.\textsuperscript{9,20} Other issues to consider in the interpretation of results of clinical trials deal with generalizability and the difference between efficacy in a controlled environment and effectiveness in the real world of everyday practice. Postmarketing observations can often reveal side effects or unintended consequences of interventions that may not be apparent within the context of highly regulated trials. More detail regarding design of clinical trials can be found in Chapter 32.

**Noninferiority Trials**

A common type of randomized control trial is the *superiority* trial, in which investigators determine whether a new treatment is more effective than placebo, no treatment, a lower dose of the test treatment, or an established treatment that is widely used or that has known effectiveness. *Noninferiority* trials, on the other hand, are used to determine whether the effect of a new treatment is no worse than a reference treatment.\textsuperscript{21,22} This differs from an *equivalence* trial, which aims to demonstrate that the effect of a new treatment is similar to the effect of the reference treatment.\textsuperscript{21} Designing and interpreting noninferiority trials can be challenging due to several weaknesses of this study compared with superiority trials. Intention-to-treat analysis (a commonly used approach in superiority trials in which not all participants may have completed the treatment protocol) is not possible in noninferiority trials. Intention-to-treat analysis tends to bias results toward the null (treatment equivalence), which, in a noninferiority trial, would result in an inferior treatment being incorrectly labeled as non-inferior.\textsuperscript{21,22} Additionally, an inferiority margin must be predetermined, and this margin may be subjectively based on the expectation of a minimally important effect or, more objectively, on the effect of the reference treatment in prior studies.\textsuperscript{22,23} For the latter, the assumption is that the effect of the reference treatment in the noninferiority trial is similar to its effect in prior trials, which may not be true if the current and prior trials differ on the basis of critical factors (i.e., study population).\textsuperscript{21,22}

**Comparative Effectiveness Research**

Evidence of effectiveness of treatments is necessary for informing medical decisions by clinicians, patients, and caregivers; reducing health care costs; and improving outcomes. Comparative effectiveness research (CER) generates evidence on the effectiveness of treatments with the goal of determining which treatment is best for particular groups of people with certain conditions to improve the quality of

---

Chapter 32.

**BROAD ISSUES IN THE APPROACH TO RHEUMATIC DISEASE**
treatments and outcomes. Systematic reviews in which all results from existing studies are compiled and the benefits and risks of the treatments are evaluated across different populations may be conducted. Alternatively, new studies may be conducted to examine the effectiveness of a treatment including its benefits, side effects, and costs compared with other available treatments for a given outcome in a specified population.

The quality of health care has continued to improve over the past decade in the United States, yet the increasing costs and financial burden of health care, as well as regional differences in treatment use, cost, and outcomes, are concerning. For example, Fisher and colleagues reported that individuals in the highest-spending regions received 60% more health care than the lowest-spending regions, but this additional care did not result in improved mortality rates, functional status, or patient satisfaction. The rising health care costs and distressing regional differences have given rise to a push for focused efforts on CER nationally. Under the American Recovery and Reinvestment Act (ARRA) of 2009, Congress allocated $1.1 billion to spark the advancement of CER nationally to improve health care quality while reducing costs. As required by the ARRA, the Institute of Medicine (IOM) Committee on Comparative Effectiveness Research Pioritization selected 100 health topics requiring CER including osteoarthritis (musculoskeletal disorders) and rheumatoid and psoriatic arthritis (immune system, connective tissue, and joint disorders) on the basis of the input of public and private stakeholders. For additional information on the national priorities of CER, the IOM’s Initial National Priorities on Comparative Effectiveness Research is available at the National Academies Press website: www.nap.edu.

**Biases in Study Design**

Error in a study may be random (chance) or systematic (bias). Bias includes errors in the selection of participants, errors in measurement of a variable, or confounding. Bias may produce an incorrect conclusion about the association between an exposure and disease.

**Selection Bias**

The procedures used to select participants for a study or factors related to study participation may result in a different exposure-disease association between participants and nonparticipants. Selection bias may occur in any study design, most notably in retrospective or case-control studies, where the exposure and outcome both occur before selection of participants. Differential participation may arise in cohort studies or clinical trials with loss to follow-up, particularly if participants leave the study for reasons related to the exposure or the disease.

**Information Bias**

Errors in the measurement or collection of information may occur. If a variable is measured categorically, information about a participant may be placed in the wrong category or misclassified. Nondifferential misclassification of an exposure occurs when misclassification is not related to the presence of disease. Misclassification is differential if exposure differs by disease status. Likewise, misclassification of a disease is nondifferential if it does not differ by exposure status and differential if it varies by exposure status. Nondifferential misclassification biases an association between exposure and disease toward the null, except if the association is the null. Differential misclassification may bias the association in either direction.

**Recall Bias**

In case-control studies, those who are cases may have a different recall of their exposure history than those who are noncases. This difference in recall can introduce a bias that inflates the estimate of an association between exposure and disease. Recall bias is differential misclassification because the exposure is misclassified differently among cases and controls. Methods for reducing recall bias include structuring questions to improve recall for both groups, selecting a control group that would be more likely to have good recall of exposure history, or using information other than interview such as medical records.

**Confounding**

Confounding occurs when there is a “mixing of effects” among the exposure, outcome, and a third factor. Specifically, a confounding variable is a risk factor for the disease, is associated with the main exposure, and is not an intermediate step on the causal pathway from exposure to disease. For example, when examining leg length inequality as a risk factor for lower extremity OA in a cohort study, a likely confounding variable would be injury to the lower limb. Injury is a risk factor for OA, it is associated with leg length inequality (a severe injury to a lower limb can result in a shortening of that limb), and it precedes both OA and leg length inequality. Methods used to control confounding include stratifying data by the confounding variable or including the variable as a covariate in multivariable statistical models. Matching may reduce confounding in case-control studies. In experimental studies, randomization is a strategy for reducing confounding.

The amount of confounding is an important consideration in determining whether one should control for it in analyses. If the estimate of an association minimally changes after adjusting for a potential confounding variable (e.g., unadjusted odds ratio = 2.62, adjusted odds ratio = 2.58), the inclusion of the variable as a covariate in a multivariable model may not be necessary. However, if the estimate changes profoundly (e.g., unadjusted odds ratio = 2.62, adjusted odds ratio = 1.05), then methods to control confounding should be employed to reduce bias in the association.

**EFFECT MEASURE MODIFICATION**

Two factors are considered to be independent if the combination of their effects is equal to their joint effects. If the effect of one factor depends on the effect of another, effect measure modification exists. This concept is also known as statistical interaction. Examining effect measure modification allows for the investigation of whether the association
between exposure and disease differs across subgroups. For example, Krishnan and colleagues reported a strong association between past history of smoking and RA in men (odds ratio 2.0, 95% confidence interval 1.2 to 3.2), but not women (odds ratio 0.9, 95% confidence interval 0.6 to 1.3). On further exploration, this association was only seen among men with rheumatoid factor–positive RA. If effect measure modification is not considered, results could be biased or important groups for targeting interventions could be missed.

**SCREENING**

Screening is an important public health strategy in reducing morbidity and mortality. Screening tests classify a person who is asymptomatic as being likely or unlikely to have the disease. This differs from diagnostic tests, which determine whether a person with signs or symptoms of a disease truly has the disease. If a screening test suggests a high likelihood of disease, then further diagnostic evaluation may occur to confirm disease presence. Although not applicable in all diseases, the early detection of disease when a person is asymptomatic is felt to result in more effective treatment than if disease detection occurs later when symptoms develop with advanced disease. To determine the validity of a screening or diagnostic test, one must establish the sensitivity or specificity of the test. Often, a new test may be compared with a “gold standard” of the definition of a disease, although this standard may not encompass all signs and symptoms of that disease. For example, biologic markers may be useful screening tools for detecting early changes in joint health that lead to OA. The presence of radiographic features of OA may be used as a comparative “gold standard,” yet this definition of OA lacks other components used to diagnose the disease such as pain, aching, or stiffness in the joint.

**Sensitivity**

Sensitivity is the probability that a test will correctly classify a case. This is expressed as a proportion of the number of cases identified by the test out of the total number of individuals with the disease. In screening, sensitivity is the probability of correctly classifying an individual as a detectable, preclinical case. If a test correctly provides a positive test result for 37 out of 43 people with disease, the sensitivity of the test is 86% (Table 30-2).

**Specificity**

Specificity is the probability that a test will correctly classify a noncase. This is expressed as a proportion of the number of individuals without disease identified by the test out of the total number of individuals without disease. If a test correctly provides a negative test result for 62 out of 66 people without disease, the specificity of the test is 94% (see Table 30-2).

**Predictive Value**

Predictive values are used to interpret the results of a test by examining the correct classification of individuals by the test. This measure is valuable because whether a person is truly a case or noncase is difficult to know (for determining sensitivity or specificity), but a positive or negative result of a test is known. A positive predictive value is a proportion of the number of cases identified out of all positive test results. If 37 people truly have disease out of 41 with a positive test result, the positive predictive value is 90% (see Table 30-2). A negative predictive value is a proportion of the noncases identified out of all negative test results. If 62 people truly do not have disease out of 68 with a negative test result, the negative predictive value is 94% (see Table 30-2).

**SUMMARY**

Epidemiologic methods can be used to measure frequency or development of disease and evaluate risk or protective factors in disease occurrence. Choice of study design depends on multiple factors including the research question, disease under study, availability of appropriate study populations, and resources available. Each study design has its own set of advantages and disadvantages, with the clinical trial considered the most rigorous.

**References**

3. Hulley SB, Cummings SR: Designing clinical research, Baltimore, 1988, Williams & Wilkins.

**Table 30-2** Hypothetical Distribution of Patients by Disease and Test Result

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>No Disease</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive test</td>
<td>37</td>
<td>4</td>
<td>41</td>
</tr>
<tr>
<td>Negative test</td>
<td>6</td>
<td>62</td>
<td>68</td>
</tr>
<tr>
<td>TOTAL</td>
<td>43</td>
<td>66</td>
<td>68</td>
</tr>
</tbody>
</table>


The references for this chapter can also be found on [www.expertconsult.com](http://www.expertconsult.com).
Economic Burden of Rheumatic Diseases

EDWARD YELIN

Cost-of-illness studies are used to provide a measure of the impact of medical conditions and can overcome the natural tendency of policymakers to focus attention on those conditions with high mortality rates to the exclusion of those with a high impact on quality of life.

Studies of the economic impact of musculoskeletal conditions as a whole have been conducted since the early 1960s using a mélange of publicly available data sources. Such studies indicated that over 3 decades, the economic impact of musculoskeletal conditions increased from the equivalent of about half a percent of the U.S. gross domestic product (GDP) to more than 2.5%. Over the past decade, the availability of a well-designed individual data source in the United States, the Medical Expenditures Panel Survey, has permitted researchers to estimate the economic impact with greater precision and to show that over that decade alone, the economic impact increased by the equivalent of more than 1.5 percentage points of GDP, to more than 7%. In the Great Recession that began in 2008, GDP declined by about 7%, so the economic impact of musculoskeletal conditions can be said to approach the same level as that of the Great Recession, although the effect is chronic, not acute.

There is now a substantial literature on the economic impact of rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), and osteoarthritis (OA). The studies in this literature are largely based on clinical samples, most from tertiary care centers, and indicate that there are substantial costs associated with all three conditions; this was so for RA even before the development of biologic agents. The most recent studies of RA indicate that with the advent of these agents, the costs of RA have skyrocketed, with the costs due to the biologic agents alone eclipsing the total costs of RA before their development including all other medical care costs and all wage losses.

The economic burden of musculoskeletal conditions is rapidly increasing in developing nations. Each discipline may use different tools to assess the burden of disease. Rheumatology health professionals may measure the impact of these conditions in terms of the discomfort and disability they cause and that these professionals may alleviate by providing health care. Psychologists may measure the impact of musculoskeletal conditions on the achievement of mental equipoise, or its opposite, mental discomfort. Economists take the measurements provided by other disciplines and turn them into monetary equivalents to describe the magnitude of the burden associated with musculoskeletal conditions. Thus economists are concerned with the monetary value of the resources used to produce health and with the return on the use of those resources in terms of reduced impacts of illness on pain and functioning. Cost of illness is an accounting of the resources spent in pursuit of health and of the residual amount of discomfort and disability left after those resources are spent (Table 31-1).

Economics is far more than a positive science, that is a discipline that observes how things are rather than seeking to change them, however. The normative part of the discipline has dual and sometimes competing goals. The first of these is ensuring that resources are allocated equitably (i.e., making sure that they are distributed fairly across individuals). The second is to make sure that they are used efficiently (i.e., as well as contemporary technology permits to produce goods). In health, resources would be distributed equitably if persons with similar levels of health received relatively

### KEY POINTS

- Studies of the economic burden of diseases are especially helpful in allocation of health care resources for musculoskeletal conditions in which mortality rates are less prominent.
- The economic burden is enumerated by summing expenditures for medical care, termed “direct costs,” and earnings losses and value of productivity losses in other activities, termed “indirect costs.”
- Per-person direct costs of musculoskeletal conditions have risen by more than 25% over the past decade, to $6799 in 2007 dollars.
- Much of this increase is fueled by a 50% increase in the number of prescription medications used each year by persons with musculoskeletal conditions and by a 38% increase in the price per prescription.
- Per-person earnings losses have risen by almost 50% during the same time, to $4979 in 2007 dollars.
- Whites, non-Hispanics, and the insured experience substantially higher medical care costs than members of minority groups and those without insurance, suggesting that substantial disparities in access to care continue in the United States.
- The aggregate economic impact of musculoskeletal conditions in the United States has increased by about 30% over the past decade, from the equivalent of 5.6% to 7.3% of the gross domestic product.
- The increasing use of biologic agents for rheumatoid arthritis has led to a rapid increase in the direct costs associated with this condition; costs of these agents alone exceed the total costs of RA from the prebiologics era including direct and indirect costs.
Table 31-1  Principal Methods to Assess Costs of Illness

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person years lost</td>
<td>The method calculates the number of years that individuals lose due to illness.</td>
</tr>
<tr>
<td>Direct costs</td>
<td>Includes costs for medical care services such as hospitalizations and medications.</td>
</tr>
<tr>
<td>Indirect costs</td>
<td>Includes costs associated with lost productivity due to illness.</td>
</tr>
</tbody>
</table>

There are two principal methods to assess the costs of illness: (1) the human capital approach, developed by Dorothy Rice when she was at the Social Security Administration and later the National Center for Health Statistics, and (2) the willingness-to-pay approach. The two methods do not differ in the way that they assess the direct costs of medical care. For the indirect costs associated with loss of function and the intangible impacts of disease, the human capital approach uses the market value of the labor to reduce the impacts (e.g., by hiring a replacement worker).

In a variant of the human capital approach called the “friction method,” the losses are estimated from the perspective of the employer and only last until the replacement worker is hired and then achieves the same productivity as the worker who left as a result of disease. At that point, an employer would be said to incur no additional costs from the onset of the prior incumbent’s disease. The willingness-to-pay approach values the loss of function as the amount the affected individual would pay to restore the function, which may be more, the same, or less than the amount it would take to replace the worker in the labor market. The human capital approach is no doubt more reliable in estimating the economic impact of the lost productivity of affected individuals because the cost of labor is well established in all advanced societies and, therefore, easy to estimate. The human capital approach, however, usually only enumerates the intangible impacts of disease (e.g., the burden associated with the experience of intense pain) but does not translate them into economic terms. The willingness-to-pay approach is theoretically capable of incorporating all of the costs of disease in those terms, although as a practical matter there are problems associated with attempting to do so.

The willingness-to-pay approach is theoretically capable of incorporating all of the costs of disease in those terms, although as a practical matter there are problems associated with attempting to do so.

The evidence that the distribution of health care services shown to reduce the burden of disease and which, based on a thorough knowledge of the treatment options, they chose for themselves. Similarly, in health, resources would be used efficiently if they produced the health outcomes people wanted while leaving the maximum amount to be used for other purposes, what economists refer to as “the opportunity cost” of those resources.

The pursuit of equity and efficiency may work in harmony. If resources are used efficiently to provide health care, then more can be allocated to ensure that everyone has access to the same set of services they desire. But the opposite is also true: if resources are used inefficiently, then the “wastage” cannot be redistributed to ensure that everyone has equal access.

The evidence that the distribution of health care resources in the United States is inefficient has been increasing for several decades and comes from studies conducted within the United States and from those comparing the United States to other nations. The inequitable distribution, however, has also recently become a central focus of analysis and policy. The signposts for inefficiency might be said to include the use of any test or procedure for which evidence has emerged, suggesting that the test or procedure does not work or does not work as well as alternatives that are available. Every comparative trial that does not result in a decreased use of the comparator that worked less well is, thus, evidence in support of the argument that our system of care is inefficient. However, the most compelling evidence comes from studies that compare populations of regions for the use of select tests and procedures and fail to find that those areas that use more of one achieve better health outcomes. Similarly, international comparisons often show that the extra usage within the United States fails to improve health-related quality of life, let alone longevity in comparison with other nations. Such studies also fail to show that the extra usage engenders greater satisfaction with most aspects of care.

The methods and data sources to estimate the economic burdens of disease, the positive side of health economics, have certainly improved over the years since the first cost-of-illness studies for the nation as a whole were completed by Dorothy Rice and colleagues. Subsequent to the development of the methods of cost-of-illness studies by Rice and others, individual investigators used those methods and the availability of clinical samples to provide highly specific estimates of discrete rheumatic conditions, particularly RA. For national estimates, the U.S. federal government has developed an annual survey, the Medical Expenditures Panel Survey (MEPS) to provide reliable estimates of the cost of illness experienced by individuals. The prior work by Rice relied on separate data sources about ambulatory care, hospital admissions, and nursing homes. The same individuals were not included in each of these data sources; indeed, the data sources were based on samples of medical care encounters rather than of individuals with specific conditions. MEPS provides estimates about every kind of health care for the same individuals and does so through systematic sampling of the U.S. population. For fairly rare conditions, because of the uniformity of data collected in each year, it provides the opportunity to merge multiple years of data so that sample sizes are sufficient to provide estimates of the cost of these conditions reliably. Similarly, the batteries of items to measure impact in each domain (e.g., the magnitude of the ambulatory care used) have been vetted through reliability studies comparing survey responses to medical records and billing information. The one aspect of cost-of-illness studies that cannot be replicated using MEPS concerns the indirect costs associated with mortality. Thus a full accounting of the costs of illness would include an estimate of the discounted present value of lost earnings among those who die prematurely. Although mortality rates are elevated for those with inflammatory rheumatic conditions, in general musculoskeletal conditions are not associated with dramatically higher mortality rates. Therefore, the bias in estimates of the costs of these conditions using MEPS data is relatively small.

The comprehensive sampling and reliability of batteries notwithstanding, the findings from the studies of Rice and colleagues about all kinds of conditions in the nation as a whole and the studies about persons with discrete conditions in clinical samples stand the test of time. Thus the early and contemporary studies both highlight the effect of population dynamics, particularly the aging of the population; increase in the cost of care as a function of changes in the kinds of services provided and increase in the unit price of those services; and changing employment circumstances and attendant wage losses on the burden of disease experienced by persons with musculoskeletal conditions.

The tools of the normative part of economics, not the subject of this chapter, have also evolved over time (Table 31-2). The tools are normative because they are based on the notion that certain levels of health care expenditures unnecessarily divert resources from other uses and that...
society should have allocative mechanisms, through the market or regulatory means, that redirect resources away from specific conditions or, more frequently, specific treatments for those conditions. That direct medical care expenditures for RA approach “x billion dollars” is a positive statement; that society should not be spending “y billion dollars” on specific treatments for RA is a normative statement; that a treatment achieves an end (e.g., extending life) less expensively than another, the product of a cost-effectiveness analysis, is a tool that may provide a guide in reallocating resources from RA care, in putting in place a policy to implement the notion that spending “y billion” on specific treatments for RA is not a good use of those resources.

This chapter summarizes the evidence accumulated on the current burden of musculoskeletal conditions overall, as well as on specific conditions within that overarching rubric. This is akin to indicating the temperature outside but not in indicating whether that temperature is hot or cold and thus whether the tools to assist in redirecting resources such as cost-effectiveness analysis should be used. However, as the evidence base in support of individual treatments for musculoskeletal conditions, for example, use of one or more disease-modifying agents in RA, or of entire strategies of treatment, for example, the use of both pharmacologic and behavioral interventions for OA increases, we as a society begin to have the means to move from the positive aspects of health economics to the normative.

**STUDIES OF THE COSTS OF ALL FORMS OF MUSCULOSKELETAL DISEASE**

In this section, data from the United States and elsewhere about the economic impact of musculoskeletal conditions as an entire category are presented.* Table 31-3 summarizes information on health care utilization and medical care expenditures for persons with any form of musculoskeletal condition in the United States by averaging 3-year periods between the year in which the MEPS was initiated, 1996, and the last year for which complete data are available, 2007. Three-year averages smooth out individual discrepant values in any one year due to small sample sizes. The first triad of years, thus, incorporates data from 1996 through 1998, whereas the last incorporates data from 2005 through 2007. In the table, all costs are expressed in 2007 terms; all data, however, represent annualized figures. Between 1996-1998 and 2005-2007, there has been little change in the use of ambulatory or hospital care for persons with musculoskeletal conditions. In each triad of years, about 90% of such persons have one or more ambulatory visits and the average number of visits per person has increased only slightly, from 9.2 to 10.4 per year. Similarly, the proportion with one or more hospitalizations has remained relatively constant in the range of 11% to 12% a year, with the average number of admissions holding steady at 0.2 per person per year.

In contrast to ambulatory visits and hospital care, there has been substantial growth in the use of prescription medications. Between 1996-1998 and 2005-2007, the proportion with any use of prescription medications inched up, but the average number of medications per person increased by about 50%, from 13.1 to 19.6.

The second set of columns in Table 31-3 displays the change in the unit prices of each kind of service between the first and last 3-year periods. The unit price of ambulatory visits increased from $185 in 1996-1998 to $216 in 2005-2007, or by about 17% in real terms. Because we had previously seen that there was little growth in the average number of visits, this suggests that there may have been an increase in the intensity of services provided in ambulatory visits or improved reimbursement for similar services, or some combination of the two. MEPS data do not permit an analysis of the extent to which of these two factors contributed to the increase in the unit prices for ambulatory visits. There was relatively little change in the unit prices associated with hospital admissions among persons with musculoskeletal conditions, with unit prices of $12,225 in 1996-1998 and $12,129 in 2005-2007.

The third set of columns, per-person costs, is the product of the average number of units of services used and the unit prices. Total per-person costs increased by about 26% between 1996-1998 and 2005-2007, from $5378 to $6799. The overall increase was driven overwhelmingly by increases in prescription drug costs but with smaller contributions from ambulatory visits and hospital admissions. Prescription drug costs increased by more than 110% between 1996-1998 and 2005-2007 in constant dollars, from $740 to $1508. Ambulatory visit costs increased by a third during this time frame, from $1694 to $2256 while hospital costs held steady when taking normal variation in 3-year averages into account.

The data in Table 31-3 reflect all the medical care costs incurred by persons with musculoskeletal conditions, regardless of whether they were incurred for that set of conditions or other conditions. Another way to assess the cost of illness is to estimate the increment in total costs beyond those that would be incurred on behalf of persons just like those with musculoskeletal conditions but who do not actually have

---

**Table 31-2 Economic Methods to Assess the Value of Health Interventions**

Drummond and colleagues provide a concise review of the methods to estimate the relationship between health care expenditures and the returns of these expenditures in terms of health-related quality of life including one of its domains, employment. When one cannot show that alternative levels of health expenditures will result in improved outcomes, one merely attempts to reduce the wastage of health expenditures, the subject of “cost-minimization studies.” When alternative treatments for a condition are available, one uses “cost-effectiveness analysis,” which shows the relative returns from these alternatives in a common natural metric (e.g., longevity). When one is comparing alternative investments across conditions, one needs an outcome metric that applies to all conditions equally; often the easiest outcome to measure in common terms is the dollar value of lost wages, the subject of “cost-benefit analysis.” However, there are inherent problems in translating outcomes into dollar terms. Accordingly, economists have developed such common metrics as the quality-adjusted life year, which takes into account the value individuals in society place on achieving a common outcome (economists use the term utility for these evaluations and use the term cost-utility analysis for assessing the returns on alternative health expenditures).

*The results are based on the author’s analysis of MEPS data. The methods were described in a 2001 publication and replicated in subsequent publications.*
Table 31-3  Units of Services Used, Unit Prices of Services, and Annual Costs per Case in the United States by Year*

<table>
<thead>
<tr>
<th>Years</th>
<th>Ambulatory Visits %</th>
<th>Hospital %</th>
<th>Rx %</th>
<th>Amb. Visits</th>
<th>Hosp.</th>
<th>Rx $</th>
<th>% Total</th>
<th>% Total</th>
<th>% Total</th>
<th>% Total</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996-1998</td>
<td>89.0%</td>
<td>9.2</td>
<td>11.1%</td>
<td>0.2</td>
<td>185</td>
<td>12,225</td>
<td>56</td>
<td>1694</td>
<td>31%</td>
<td>1956</td>
<td>36%</td>
</tr>
<tr>
<td>1997-1999</td>
<td>89.3%</td>
<td>9.2</td>
<td>11.6%</td>
<td>0.2</td>
<td>184</td>
<td>11,929</td>
<td>60</td>
<td>1690</td>
<td>31%</td>
<td>2028</td>
<td>37%</td>
</tr>
<tr>
<td>1998-2000</td>
<td>89.5%</td>
<td>9.3</td>
<td>11.6%</td>
<td>0.2</td>
<td>185</td>
<td>11,953</td>
<td>62</td>
<td>1727</td>
<td>31%</td>
<td>2032</td>
<td>36%</td>
</tr>
<tr>
<td>1999-2001</td>
<td>89.9%</td>
<td>9.5</td>
<td>11.8%</td>
<td>0.2</td>
<td>191</td>
<td>11,083</td>
<td>66</td>
<td>1812</td>
<td>31%</td>
<td>1995</td>
<td>34%</td>
</tr>
<tr>
<td>2000-2002</td>
<td>90.3%</td>
<td>10.0</td>
<td>11.7%</td>
<td>0.2</td>
<td>196</td>
<td>11,606</td>
<td>67</td>
<td>1963</td>
<td>33%</td>
<td>1973</td>
<td>33%</td>
</tr>
<tr>
<td>2001-2003</td>
<td>90.6%</td>
<td>10.5</td>
<td>11.8%</td>
<td>0.2</td>
<td>197</td>
<td>11,794</td>
<td>71</td>
<td>2063</td>
<td>33%</td>
<td>2005</td>
<td>32%</td>
</tr>
<tr>
<td>2002-2004</td>
<td>90.3%</td>
<td>10.7</td>
<td>11.7%</td>
<td>0.2</td>
<td>201</td>
<td>12,424</td>
<td>73</td>
<td>2149</td>
<td>33%</td>
<td>2112</td>
<td>32%</td>
</tr>
<tr>
<td>2003-2005</td>
<td>90.1%</td>
<td>10.7</td>
<td>11.8%</td>
<td>0.2</td>
<td>205</td>
<td>12,829</td>
<td>78</td>
<td>2194</td>
<td>32%</td>
<td>2181</td>
<td>32%</td>
</tr>
<tr>
<td>2004-2006</td>
<td>90.0%</td>
<td>10.6</td>
<td>11.5%</td>
<td>0.2</td>
<td>210</td>
<td>11,888</td>
<td>76</td>
<td>2236</td>
<td>33%</td>
<td>2021</td>
<td>30%</td>
</tr>
<tr>
<td>2005-2007</td>
<td>90.2%</td>
<td>10.4</td>
<td>11.6%</td>
<td>0.2</td>
<td>216</td>
<td>12,129</td>
<td>77</td>
<td>2256</td>
<td>33%</td>
<td>2062</td>
<td>30%</td>
</tr>
</tbody>
</table>

*Each row represents the average of 3 years of MEPS data.
†Residual includes all costs not enumerated in other three categories.
such conditions. There are two ways to make these estimates: by asking individuals to apportion health care episodes to various causes and by using regression techniques that take into account the health and demographic characteristics of the individuals because the attributions made by the individuals may not be reliable. Using the second method, approximately a third of the costs incurred by persons with musculoskeletal conditions can be attributed to those conditions.

The aggregate medical care costs associated with musculoskeletal conditions are the product of the costs per case, described in Table 31-3, and the number of persons with the conditions at any one time (Table 31-4). As can be seen in the table, the aging of the population is resulting in a substantial increase in the number of persons and percentage of the population with musculoskeletal conditions in the United States. Between 1996-1998 and 2005-2007, the number reporting one or more musculoskeletal conditions increased from 76.0 to 91.3 million, or by more than 20% in relative terms, while the percentage of the population with such conditions rose by 8.9% in relative terms, from 28% to 30.5%. When multiplied by the cost per person in each triad of years covered by the analysis, the aggregate costs rose from $408.6 billion to $620.9 billion in constant dollars, or by about 52% in relative terms. The dual importance of population aging and increases in the costs per case in the dynamics in the aggregate costs of musculoskeletal conditions can be seen by the fact that the percentage increase in the former, more than 20%, and the percentage increase in the latter, 26%, were both substantial.

The relatively large increase in the cost per case of all musculoskeletal conditions would have been even greater were it not for the fact that the vast majority of cases within the overall musculoskeletal diseases rubric are OA and similar mechanical diseases rather than such autoimmune conditions as RA or SLE. Treatments for OA, for example, have been relatively stable over time. As explained later, costs for individual autoimmune conditions have risen substantially over the past decade as new treatments such as the biologics for RA have taken hold.

Table 31-5 shows the per-person and aggregate medical care costs in 2005-2007, stratified by demographic characteristics. Musculoskeletal conditions are more prevalent among females; females with the conditions also experience per-person medical care costs that are about 20% higher than men, $7328 compared with $6130. When the higher prevalence and higher per-person costs are combined, aggregate costs are $373.2 billion among females and only $247.6 billion among males, a difference of about 50%. Per-person medical care costs associated with musculoskeletal conditions increase monotonically with age, from a low of $2994 among those younger than age 18 to a high of $11,128 among persons 65 or older. However, because of the number of persons in the 45- to 64-year-old group, aggregate costs are actually slightly higher in the latter group than among those 65 or older, $236.5 billion versus $213.5 billion. Non-whites with musculoskeletal conditions experience slightly lower per-person costs than whites with these conditions, $6436 versus $6865, a difference of about 7%. On the other hand, Hispanics with musculoskeletal conditions incurred medical care costs of $4969, fully 29% lower than the $6981 incurred by non-Hispanics.

In addition to the differences by race and Hispanic status, there are substantial differences in medical care costs by a measure of socioeconomic status—extent of formal education. Those with less than a high school education experienced costs 18% higher than the next lowest group, those who had completed a high school education, probably reflecting the former group’s greater need for medical care. Persons with musculoskeletal conditions who have never been married incurred substantially lower medical care costs than the currently or formerly married, 30% with respect to the former and almost 45% with respect to the latter.

Of note, persons with musculoskeletal conditions reporting public insurance actually incurred higher medical care costs than those with private insurance ($9306 vs. $6621, a difference of about 40%), but both groups had substantially higher costs than those without insurance who averaged only $2304 per person in medical care costs, a difference of 65% relative to those with any private insurance and 75% relative to those with any form of public insurance. Thus it cannot be said that those without insurance necessarily receive the care they need.

Overall, the results reported in Table 31-5 suggest that there are wide gulls in medical care costs in the United States by race, ethnicity, and insurance coverage.

### Table 31-4 Number and Percent of U.S. Population with Musculoskeletal Conditions and Annualized per Case and Aggregate Costs of the Conditions, in 2007 Dollars

<table>
<thead>
<tr>
<th>Years</th>
<th>Number</th>
<th>% U.S. Population</th>
<th>Cost per Person</th>
<th>Aggregate Cost (billion $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996-1998</td>
<td>75,978,133</td>
<td>28.0%</td>
<td>5378</td>
<td>408.6</td>
</tr>
<tr>
<td>1997-1999</td>
<td>75,173,840</td>
<td>27.5%</td>
<td>5513</td>
<td>414.4</td>
</tr>
<tr>
<td>1998-2000</td>
<td>74,077,194</td>
<td>26.8%</td>
<td>5576</td>
<td>413.1</td>
</tr>
<tr>
<td>1999-2001</td>
<td>75,600,394</td>
<td>27.0%</td>
<td>5802</td>
<td>438.6</td>
</tr>
<tr>
<td>2000-2002</td>
<td>79,748,298</td>
<td>28.1%</td>
<td>6038</td>
<td>481.5</td>
</tr>
<tr>
<td>2001-2003</td>
<td>84,297,419</td>
<td>29.3%</td>
<td>6321</td>
<td>532.8</td>
</tr>
<tr>
<td>2002-2004</td>
<td>87,575,871</td>
<td>30.1%</td>
<td>6593</td>
<td>577.4</td>
</tr>
<tr>
<td>2003-2005</td>
<td>88,946,833</td>
<td>30.3%</td>
<td>6705</td>
<td>596.4</td>
</tr>
<tr>
<td>2004-2006</td>
<td>89,652,587</td>
<td>30.3%</td>
<td>6695</td>
<td>600.2</td>
</tr>
<tr>
<td>2005-2007</td>
<td>91,320,095</td>
<td>30.5%</td>
<td>6799</td>
<td>620.9</td>
</tr>
</tbody>
</table>

Each row represents the average of 3 years of Medical Expenditures Panel Survey data. From author’s analysis of Medical Expenditures Panel Survey, 1996-2007.
CHAPTER 31 | ECONOMIC BURDEN OF RHEUMATIC DISEASES

445

CHAPTER 31

EConomiC BuRdEn of RHEumATiC disEAsEs

under 64 million in the first triad of years covered to more than 76 million in the last triad. Concurrently, the per-person earnings losses among working age persons with musculoskeletal conditions increased from $3350 to $4979, or by just less than 50% in real terms. As a result of the increase in the prevalence of the at-risk population and the per-person earnings losses among such persons, aggregate earnings losses among persons with musculoskeletal conditions rose from $213.9 billion in 1996-1998 to $379.9 billion in 2005-2007, or by more than three quarters. From Table 31-5, it is evident that, of the $620.9 billion in medical care costs incurred by all persons with musculoskeletal conditions in 2005-2007, about $358.5 billion occurred among those between ages 18 and 65. Thus among working age persons, earnings losses, estimated to be about $380 billion, exceed direct medical care costs by more than $20 billion.

Figure 31-1 summarizes the overall economic impact of musculoskeletal conditions since the first studies of the cost of illness were conducted by Dorothy Rice and colleagues in the early 1960s.4-28-31 It should be reiterated that the availability of the MEPS data makes it possible to study all of the costs incurred by individuals with the same dataset, as opposed to the prior studies by Rice and colleagues based on multiple data sources. This renders exact comparisons to

Table 31-5 Prevalence and Medical Care Costs among U.S. Persons with Musculoskeletal Conditions in 2007 Dollars, 2005-2007 (n = 26,373)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Prevalence (in millions)</th>
<th>Per Person ($)</th>
<th>Aggregate (billion $)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Total</td>
</tr>
<tr>
<td>Total</td>
<td>91.3</td>
<td>6799</td>
<td>620.9</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>50.9</td>
<td>7328</td>
<td>373.2</td>
</tr>
<tr>
<td>Male</td>
<td>40.4</td>
<td>6130</td>
<td>247.6</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18</td>
<td>8.6</td>
<td>2994</td>
<td>25.7</td>
</tr>
<tr>
<td>18-44</td>
<td>28.4</td>
<td>3693</td>
<td>105.0</td>
</tr>
<tr>
<td>45-64</td>
<td>33.1</td>
<td>7667</td>
<td>253.5</td>
</tr>
<tr>
<td>65+</td>
<td>21.3</td>
<td>11,128</td>
<td>236.5</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>77.3</td>
<td>6865</td>
<td>530.6</td>
</tr>
<tr>
<td>Nonwhite</td>
<td>14.0</td>
<td>6436</td>
<td>90.3</td>
</tr>
<tr>
<td>Hispanic Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>8.3</td>
<td>4,969</td>
<td>41.1</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>83.0</td>
<td>6,981</td>
<td>579.8</td>
</tr>
<tr>
<td>Education (n = 178 missing data)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;High school</td>
<td>15.4</td>
<td>8,115</td>
<td>125.1</td>
</tr>
<tr>
<td>High school</td>
<td>28.6</td>
<td>6,899</td>
<td>197.4</td>
</tr>
<tr>
<td>Some college</td>
<td>22.2</td>
<td>6,476</td>
<td>143.6</td>
</tr>
<tr>
<td>College graduate</td>
<td>14.3</td>
<td>5,841</td>
<td>83.4</td>
</tr>
<tr>
<td>Graduate school</td>
<td>10.4</td>
<td>6,565</td>
<td>68.0</td>
</tr>
<tr>
<td>Marital Status (n = 10 missing data)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>15.5</td>
<td>4707</td>
<td>73.1</td>
</tr>
<tr>
<td>Married or with partner</td>
<td>52.1</td>
<td>6695</td>
<td>349.0</td>
</tr>
<tr>
<td>Divorced/Widowed/Separated</td>
<td>23.6</td>
<td>8401</td>
<td>198.6</td>
</tr>
<tr>
<td>Insurance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any private</td>
<td>63.1</td>
<td>6621</td>
<td>418.0</td>
</tr>
<tr>
<td>Public</td>
<td>19.7</td>
<td>9306</td>
<td>183.4</td>
</tr>
<tr>
<td>None</td>
<td>8.5</td>
<td>2304</td>
<td>19.5</td>
</tr>
</tbody>
</table>

Each value is based on the average of 2005-2007 MEPS data.

Earnings losses associated with musculoskeletal disorders among the working age population, the component of indirect costs calculable in MEPS, are substantial and growing in real terms; Table 31-6 details their magnitude from 1996-1998 through 2005-2007. The number of persons with a work history between ages 18 and 64 increased from just under 64 million in the first triad of years covered to more than 76 million in the last triad. Concurrently, the per-person earnings losses among working age persons with musculoskeletal conditions increased from $3350 to $4979, or by just less than 50% in real terms. As a result of the increase in the prevalence of the at-risk population and the per-person earnings losses among such persons, aggregate earnings losses among persons with musculoskeletal conditions rose from $213.9 billion in 1996-1998 to $379.9 billion in 2005-2007, or by more than three quarters. From Table 31-5, it is evident that, of the $620.9 billion in medical care costs incurred by all persons with musculoskeletal conditions in 2005-2007, about $358.5 billion occurred among those between ages 18 and 65. Thus among working age persons, earnings losses, estimated to be about $380 billion, exceed direct medical care costs by more than $20 billion.

Figure 31-1 summarizes the overall economic impact of musculoskeletal conditions since the first studies of the cost of illness were conducted by Dorothy Rice and colleagues in the early 1960s.4-28-31 It should be reiterated that the availability of the MEPS data makes it possible to study all of the costs incurred by individuals with the same dataset, as opposed to the prior studies by Rice and colleagues based on multiple data sources. This renders exact comparisons to

Table 31-6 Number of U.S. Persons Aged 18 to 64 with a Work History and Annualized Mean and Aggregate Earnings Losses, in 2007 Dollars, by Year

<table>
<thead>
<tr>
<th>Years</th>
<th>Number with Work History</th>
<th>Earnings Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996-1998</td>
<td>63,840,865</td>
<td>3350</td>
</tr>
<tr>
<td>1997-1999</td>
<td>62,603,143</td>
<td>4142</td>
</tr>
<tr>
<td>1998-2000</td>
<td>61,508,942</td>
<td>4497</td>
</tr>
<tr>
<td>1999-2001</td>
<td>63,182,182</td>
<td>4451</td>
</tr>
<tr>
<td>2000-2002</td>
<td>66,886,334</td>
<td>4458</td>
</tr>
<tr>
<td>2001-2003</td>
<td>70,931,602</td>
<td>4574</td>
</tr>
<tr>
<td>2002-2004</td>
<td>73,343,698</td>
<td>5069</td>
</tr>
<tr>
<td>2003-2005</td>
<td>74,309,513</td>
<td>5397</td>
</tr>
<tr>
<td>2004-2006</td>
<td>74,832,482</td>
<td>5128</td>
</tr>
<tr>
<td>2005-2007</td>
<td>76,301,791</td>
<td>4979</td>
</tr>
</tbody>
</table>

Each row represents the average of 3 years of MEPS data.
the studies conducted by Rice impossible. Nevertheless, one can compare trends in the Rice studies and, subsequently, trends in the studies using MEPS. The vertical line in the figure indicates the disjuncture between the two sets of studies. In the period covered by the Rice studies, that is, from 1963 through 1995, total costs of musculoskeletal conditions rose from well under the equivalent of 1% of GDP to just under 3% of GDP. During this time, the proportion of total costs due to earnings losses varied between 38% and 59% of the total for a year. The variance in this percentage is a function of whether wages or medical care costs were rising faster.

The National Arthritis Data Workgroup has estimated that the change in methods associated with the initiation of the MEPS accounts for about half of the increase relative to the last of the studies conducted by Rice and colleagues, whereas the remainder constitutes a real increase (i.e., an increase not associated with changes in the methods of analysis). During the time since the MEPS analyses began, the total costs of musculoskeletal conditions have risen from the equivalent of 5.6% of GDP to about 7.3% in slightly more than a decade, or by more than 30% in relative terms. Medical care costs among persons of all ages with musculoskeletal conditions have risen from the equivalent of 3.7% of GDP to 4.5%, or by about 22% in relative terms, while the earnings losses, estimated here only for those aged 18 through 64, have increased from the equivalent of 1.9% to 2.8% of GDP, or by about 47% in relative terms.

A recession is said to occur when the economy contracts by 1% of GDP or more for two quarters or more. In the deep recession that began in 2007, GDP fell by about 7% on an annualized basis. The economic impact of musculoskeletal conditions can therefore be said to be the equivalent of a severe recession, although to judge by the increases over time, the onset occurs more slowly. However, the impact is permanent, at least until medical care expenditures either result in improved functionality and concomitant reduced earnings losses or, without that impact, are reduced by decreases in the amount of health care used and/or the unit prices of the services used to treat this group of conditions.

**OTHER NATIONS**

Studies on the economic impact of musculoskeletal conditions have been conducted in Canada, Australia, the United Kingdom, Sweden, and the Netherlands. All use methods broadly similar to the studies in the United States conducted by Rice and colleagues. The most recent Canadian study, using 2000 data but just published, showed that the economic impact of musculoskeletal conditions was approximately $22.3 billion CDN, or roughly 2.1% of that country’s GDP for that year, similar to the most recent of Rice’s studies in the United States. In the Canadian study, indirect costs associated with earnings losses were much greater than direct medical care costs for these conditions. The study in the Netherlands is particularly instructive. In that nation, it was found that musculoskeletal conditions ranked second among all major diagnostic groups in medical costs, exceeding coronary and other circulatory conditions and only being eclipsed by mental retardation.

The demographic structure of the developed nations in which these studies have been conducted are similar, with each facing a pandemic of musculoskeletal disease due to the aging of their populations. Increases are also probably associated with the increased prevalence of obesity in many of the developed nations; obesity is associated with OA. However, we are not aware of systematic studies of the cost of these conditions in the developing world.

**COST OF DISCRETE CONDITIONS**

The first studies of the cost of illness for discrete rheumatic diseases were conducted about RA and, reflecting the importance of this condition in the practices of rheumatologists and other rheumatology health professionals, the economic impact of RA continues to be the subject of extensive investigation. However, a sizable literature on the economic impact of SLE and OA exists, and there are at least a few studies about the costs associated with ankylosing spondylitis, psoriatic arthritis, gout, and fibromyalgia.
The costs of RA have been extensively reviewed. Several studies parse the literature for methodological differences among studies including the perspective of the analysis (individual with the condition, focusing on the share of total costs paid out of pocket; payer, focusing on the share paid by health insurance plans or national health insurance; and society, focusing on all of the resources devoted to health care for the condition, regardless of who pays for those resources, as well as the forgone productivity); type of case being studied (prevalent or incident); and effect of specific treatments including but not limited to the biologic agents.

Among studies of RA not of recent onset and preceding the development of the biologic agents, direct costs in most studies were in the range of $5000 to 7000 per year, while indirect costs averaged between two and three times direct costs. Within direct costs, approximately two thirds were attributable to hospital admissions even though less than 10% of persons with the condition are admitted to the hospital in any one year. Among those hospitalized, the vast majority were receiving surgical interventions, principally joint replacement surgery.

Among the approximately 90% of persons with RA not admitted to the hospital in a year, costs in the prebiologic era could be quite low. The modal costs for RA in that era were $685 per year in one study, whereas the median was only $2715. At the 90th percentile, however, costs were higher than $8000 per year while at the 95th percentile, costs were in excess of $30,000.

The high indirect costs of RA are a function of the prevalence of work disability, with rates varying from a low of 34% in one study to a high of 59% in another. One study reported on the duration of time from onset of the condition until work disability occurred: Of persons with RA, 10% stopped working in the first year after onset, about half had stopped working in the first decade after onset, and 90% left work before the normal age of retirement.

The impact of the biologic agents on costs of RA has been so profound that studies should be divided into those that predate and postdate the availability and diffusion of these agents. In some series, the proportion of persons with RA receiving biologic agents has reached a third or more. At approximately $20,000 per year per user, medical care expenditures for RA for biologic agents alone now exceed estimates of the total costs of the disease including indirect costs, from the prebiologic era. Several studies quantify the incremental costs associated with the use of the biologics beyond what would be expected for persons with similar health status. Michaud and colleagues reported that medical care costs for RA were more than $19,000 per year for persons taking biologic agents, but only about $6000 for those not, even after accounting for differences in the two populations. Fautrel and colleagues reported that the incremental costs associated with etanercept and infliximab were about $23,000 to $25,000 per year, or more than three times the annual direct costs of RA in the prebiologic era. Sorensen and Anderson studied the potential impact of the biologics on the Danish health care service, reporting that under varying scenarios the biologics would raise the cost of treating RA from 50% to 500%. In a study of elderly persons with RA, Weycker and colleagues noted that RA care averaged between $12,000 and $23,000 in the biologics era, or more than twice the average from the prebiologics era.

However, perhaps the best way to gauge the impact of the biologic agents on the costs of RA is merely to note that, in this disease with a prevalence of less than 1% of the population, pharmaceutical companies find it profitable to advertise the agents on national television shows in the United States. Assuming that the prevalence was 0.5% of the approximately 250 million adults in the United States, or 1.25 million, that about a quarter of these used the biologic agents, and that the agents cost $20,000 per year, then biologic agents would account for more than $6 billion a year in health care expenditures for RA.

The costs of RA may be substantial early after onset. Merkesdal and colleagues observed annual indirect costs due to lost wages within the first 3 years of disease. Of note, however, they observed that short-term sick leave costs were higher in the first year than the next two, but that permanent work disability costs increased to partially offset the decrease in sick leave. Similarly, Hallert and colleagues reported that direct costs in the first year after diagnosis averaged almost $5000, whereas indirect costs were more than twice as great. Among persons with seropositive RA of about 6 months’ duration, Newhall-Perry and colleagues reported that direct and indirect costs were about $200 and $300 per month, respectively. The relatively high indirect costs were a result of the high rate of work cessation (>10% among those employed at diagnosis) and sick leave (>5% in the same group).

The high medical care costs associated with RA, manifest even among those of recent onset, present a paradox for choice of therapy at the level of the individual patient-provider dyad and at the level of the payer and society. On the one hand, there is ample opportunity to reduce much of the nonmedication cost of treatment for joint replacement surgery and for indirect costs due to lost wages by timely use of the newly discovered and highly efficacious biologic agents. On the other hand, iron clad evidence that they can reduce the frequency of joint replacement and work loss is lacking information necessary to prove cost-effectiveness, in part because of methodological problems in the studies conducted including lack of head-to-head trials and uncertainty about the results of the studies that have been done. Thus we know that the costs of RA are high and growing due to the use of biologic agents, but we do not know whether using them can reduce the indirect costs and the functional losses that lead to joint replacement surgery.

Although the bulk of the now vast literature on the economic impact of RA derives from clinical samples, it is possible to make an estimate for the United States as a whole using MEPS data, with the caveat that there may be individual respondents to the MEPS surveys who are unaware of the specific form of arthritis they have even though a physician would diagnosis them as having RA and, on the other hand, some who inaccurately state that they have RA when that is not the case. The advantage is that MEPS samples from the community at large, eliminating the bias from sampling in tertiary care centers or select health plans. The MEPS estimates were made by merging
five annual waves of data, from 2000-2004. On average in the United States as a whole, direct costs among persons with RA averaged just under $10,000 a year, whereas earnings losses among those aged 18 to 64 totaled about $14,000 a year. In aggregate, direct costs and earnings losses associated with RA averaged $29.1 billion in the years of the analysis, or about 0.3% of the nation’s average annual GDP in those years.

Systemic Lupus Erythematosus

The first studies of the costs of RA appeared in the mid-1970s; the first studies of the costs of SLE did not appear until almost 2 decades later. Nevertheless, there is now a substantial literature spanning three continents and going beyond studies of undifferentiated SLE to focus on the impact of specific levels of disease activity and specific organ manifestations on direct and indirect costs. As we have seen, the costs of RA have been historically driven by work disability and joint replacement, but more recently expenditures for biologic agents have come to play an important role. SLE is also associated with high rates of work loss, and certain manifestations, particularly renal failure and neuropsychiatric impairment, result in high medical care costs due to hospitalization. However, the use of biologic agents in SLE is only now a prospect, so biologic agents are not yet a factor in the costs of SLE. The work loss costs may be profound in SLE both because a high proportion of persons with this condition experience temporary or permanent reduction of employment but also because the age of onset is, on average, a decade earlier than in RA.

Comparing the direct costs of SLE to those of RA in the prebiologics era for the latter condition, the magnitude of the direct costs are slightly higher in the former condition, but the distribution is not as heavily tilted to the inpatient environment. In the studies, medical care costs averaged about $7000 per year, with a range from slightly more than $4000 to just under $14,000. On average, the hospital accounted for considerably less than half of direct costs, despite high inpatient costs for the small proportion with complications of NSAIDs or surgery, average direct costs rise severalfold.

In an important study, Clarke and colleagues compared costs in the United States, United Kingdom, and Canada. To ensure that prices for services did not affect their estimates, they used the same unit prices for each country. They observed that costs of SLE in Canada and the United Kingdom were of similar magnitude, but both were about 10% to 15% less than in the United States. This extra level of expenditure does not, however, result in better outcomes. Indirect costs are measured using a greater heterogeneity of methods than for direct costs, but for those using similar methods, indirect costs exceed direct costs by an average of about 2:1. Indirect costs are high in SLE due to high rates of work loss; prevalence of work loss among those employed at onset of disease is estimated to be 15% within the first 5 years of diagnosis and 63% within the first 2 decades. Even though these rates of work loss are certainly worthy of concern in their own right, the labor market outcomes of persons with SLE may be worse than that of persons with some other diseases of lesser severity because symptoms such as fatigue, pain, and neurocognitive deficits may not be as obvious to the untutored eye.

In studies of the impacts of specific levels of disease activity and specific organ manifestations on direct and indirect costs, there is mounting evidence of the profound effect of lupus nephritis; other measures of renal damage; neuropsychiatric manifestations including memory impairment; and global measures of disease activity and severity and disease flare on costs. For example, in one study, Zhu and colleagues observed that SLE patients experiencing flares incurred twice the total costs of SLE as those without flares, with those experiencing renal or neuropsychiatric flares having the highest levels of costs. Effective treatment to prevent damage accrual to specific organs or to reduce the frequency and severity of flares may result in a substantial reduction in the costs of SLE.

Osteoarthritis

Osteoarthritis is a common condition affecting millions of people worldwide. The direct costs of OA have been in the range of $4000 to $6000 per year, with indirect costs considerably less than that, principally because many persons with OA are no longer in the normal ages of labor force participation. However, when hospitalization occurs for the complications of NSAIDs or surgery, average direct costs rise severalfold.

These may be underestimates. In a study using MEPS data, direct costs among persons with OA averaged about $10,000 and earnings losses among persons aged 18 through 64 were of about the same magnitude. Aggregate direct costs totaled just under $23 billion, whereas indirect costs were of about the same magnitude, $22 billion.

Even these may underestimate the economic impact of OA. In 2004 it was estimated that in excess of 1 million joint replacement surgeries occurred, with the number projected to increase with the aging of the population on the one hand and increasing willingness to operate on younger individuals on the other hand. Of these surgeries, 83% of hip replacements and 97% of knee replacements were done for OA. These operations alone cost in excess of $29 billion, about 25% greater than the estimates from MEPS including all other direct-cost items.

Another way to estimate the costs associated with OA in MEPS would be to subtract the costs of RA from the total enumerated in the overarching category “Arthritis” because
most persons in that category do have OA. If one were to do that, the direct costs of OA would total almost $270 billion, whereas the earnings losses among persons aged 18 through 64 would total about $28 billion. Together, the economic burden of OA including medical care expenditures and earnings losses among those aged 18 through 64 would be in the range of 3% of GDP.

**Back Conditions**

MEPS data are ideal for estimating the economic burden associated with back problems because the overarching category does not require a specific diagnosis from a physician but, instead, can be reported by an individual experiencing a back problem. Using MEPS, the author estimated that the prevalence of self-reported back problems increased by about 19% between 1996-1998 and 2002-2004, from 27.4 to 32.9 million; the fraction of the population with such a problem increased by 12% during this time, to 11.3%.

Direct costs per case associated with back problems increased by about 25% over the period covered, from $4756 to $5923 in 2004 terms. This increase was largely fueled by an 88% increase in the real value of prescription medicines used for back problems. Overall, direct costs for back problems increased from the equivalent of 1.2% to 1.7% of GDP between 1996-1998 and 2002-2004 as a result of the increase in the direct costs per case and the increase in prevalence. Earnings losses among persons aged 18 to 64 with back problems are relatively slight, averaging $1871 among the 24.3 million individuals these ages with such problems, or the equivalent of 0.4% of GDP in 2004 terms. Earnings losses are relatively slight because a majority of persons with back problems experience temporary disability rather than permanent work loss, although it should be noted that back problems are common causes of work loss but that such loss occurs with relative infrequency when compared with the high overall prevalence of back problems in the population.

**Other Conditions**

**Ankylosing Spondylitis**

Boonen and colleagues have spearheaded efforts to characterize the burden associated with ankylosing spondylitis in a systematic review of the cost-of-illness studies in this condition and by an article describing the frequency with which various kinds of impact for the condition are experienced. In the former publication, they note that the total costs associated with ankylosing spondylitis including direct costs ranged between $7243 and $11,840, amounts comparable with the cost of RA in the prebiologic era. Woolf observed that the typical individual incurs relatively high costs for assistive devices and personnel and sustains substantial earnings losses. However, the distribution among cost categories differs among countries. In the United States, coverage for physical therapy is relatively poor and hospital admissions are rarely reimbursed for ankylosing spondylitis, so a larger proportion of total costs of the condition are attributable to wage losses than in other nations.

**Fibromyalgia**

Annemans and colleagues recently summarized the literature on the economic impacts of fibromyalgia. Fibromyalgia, a controversial diagnostic entity, is defined by persistent, widespread pain and, perhaps, heightened reactivity to painful stimuli compared with those without the condition. Symptoms related to the condition include sleep disturbance and depressive mood. Fibromyalgia disproportionately affects women, many of whom are not employed for pay outside the home and others of whom, if employed, are subject to the normal discrimination in employment that women experience more generally. As a result, direct medical care costs, averaging between $5000 and $6000 in the studies reviewed by Annemans and colleagues, dwarf indirect costs due to earnings losses, which averaged between $2000 and $3000. This occurs despite evidence that a relatively high proportion of those who are employed at onset either stop working, reduce their hours, or at the least may be subject to temporary disability. Others may change job tasks or switch jobs to accommodate the symptoms. Of note, the studies reviewed show that direct costs associated with fibromyalgia often spike before diagnosis because testing that is part of diagnosis accounts for a substantial burden on society but then may be somewhat reduced as the diagnostic phase of care passes and/or the person with fibromyalgia accommodates to the condition or responds to the combination of pain medications and antidepressives that are often prescribed.

**SUMMARY AND CONCLUSIONS**

Researchers use cost-of-illness studies to describe the impact of conditions on individuals and society (the positive function) and assist policymakers in allocating resources to redress those burdens (the normative function). Some have criticized the development of such measures of impact because it deflects attention from assigning relative values to interventions that might alleviate the impacts (i.e., from calculating the cost-effectiveness of interventions regardless of the condition for which they are to be used). In essence they argue that the fact of presenting large estimates of burden may mean that resources may be diverted from using highly cost-effective interventions on conditions of small prevalence even when no effective interventions exist for the conditions of high prevalence and impact, in effect wasting money.

However, an effective counterargument is that allocation decisions will still be made on criteria other than cost-effectiveness. For example, Verbrugge long ago noted that fatal conditions tend to garner more attention from policymakers and this puts groups that have conditions with apparent low fatality rates but which are severely disabling at a disadvantage. She observed that men tend to have higher rates of many fatal conditions, particularly cardiovascular disease, whereas women have higher rates of musculoskeletal and neurologic conditions that may have severe impacts but which are not commonly considered to cause mortality.

Certainly societies should allocate services on the basis of the return from the investment, but we must also be certain that we do not unduly discriminate against those...
with conditions that do not always garner attention proportional to their effect on people’s lives. At the least, we must acknowledge the positive aspect of cost-of-illness studies in describing how resources that are diverted by disease could be used for other purposes were the disease to be eradicated, to provide a measure of the potential returns from effective interventions as they arise.

This debate about the role of cost-of-illness studies notwithstanding, there is no debating that musculoskeletal conditions do divert substantial resources from other uses and the amount so diverted has been increasing due to the aging of the population and increases in the amount spent to treat them, especially for medications. Since the U.S. government began collecting MEPS data slightly more than a decade ago, the gross economic impact of musculoskeletal conditions has grown by the equivalent of 1.7% of GDP, to 7.3%. If one accounts only for the increment expected of conditions has grown by the equivalent of 1.7% of GDP, to a decade ago, the gross economic impact of musculoskeletal conditions do divert significant economic resources away from other uses due to the substantial expenditures for medical care and by depriving the economy of the productivity at work and in other activities of those affected by the conditions.

References

33. Badley E: The economic burden of musculoskeletal disorders in Canada is similar to that for cancer, and may be higher, J Rheumatol 22:204–206, 1995.


The references for this chapter can also be found on www.expertconsult.com.
Clinical Trial Design and Analysis

ROBERT B.M. LANDEWÉ • DÉSIRÉE M.F.M. VAN DER HEIJDE

Clinical trials are studies designed to assess the efficacy and toxicity of drugs or other interventions. Although the term clinical trial often refers to “randomized clinical trial (RCT),” every study in which patients are exposed to an intervention and in which data are systematically collected can be considered as a clinical trial. Clinical trials play an important role in new drug development, but can also serve to further explore existing drugs or interventions or to refine their use, as in determination of predictive factors for treatment efficacy, or to test treatment strategies. This chapter broadly confines itself to RCTs. It discusses methodologic principles of clinical trials, as well as RCT analysis, in the context of rheumatology, and it unveils limitations of RCTs while briefly discussing alternative solutions in design and analysis. This chapter is intended as an introduction for clinicians and researchers working in the field of rheumatology.

TRIAL DESIGN

Randomized Clinical Trials

The classical template of an RCT includes two (or more) trial arms comparing the drug or intervention of interest (e.g., the new drug) with a control intervention. The latter may include a placebo or sham intervention, or an intervention that is considered to represent standard care. By definition, the treatment arms are created through the process of randomization, which is pivotal and will be outlined later in greater detail.

To better understand differences in trial design, it is often helpful to distinguish explanatory RCTs and pragmatic RCTs.1,3 Trials of new drugs, such as those designed for drug registration, aimed at showing efficacy and short-term safety, belong to the group of explanatory RCTs. In general, all elements of trial design, such as selection of patients, sample size, choice of the comparative intervention, and duration of the trial, are chosen in such a manner that the trial can optimally demonstrate a treatment effect, that is, a difference in efficacy between the new drug and the control intervention. The methodologic robustness of a trial, which is dependent on these elements of trial design, is referred to as internal validity. Explanatory trials do not always resemble clinical practice. As an example, they often include for methodologic reasons patients with a high level of disease activity who form only a minority in clinical practice. The extent to which clinical trial results can be extrapolated to the common clinical practice is referred to as external validity. As a rule of thumb, explanatory trials have a high level of internal validity, which may, however, jeopardize external validity to some extent. Pragmatic trials more closely resemble the clinical situation. Such trials aim to optimize treatment by further exploring existing drugs or treatment strategies. Pragmatic trials incorporate fundamental principles of RCTs, such as randomization, but include a more realistic representation of patients, may have a longer duration, and may allow co-interventions. In general, pragmatic trials have a lower level of internal validity as compared with explanatory trials, but a higher level of external validity. Often, explanatory trials are initiated and sponsored by pharmaceutical industry, but most pragmatic trials are (academic) investigator-driven initiatives.

Randomization

Randomization, the process by which patients are assigned to treatment by chance, is the most important methodologic characteristic of an RCT and deserves some explanation. Randomization makes treatment arms similar for all variables except treatment, or, in other words, randomization divides all known and unknown variables that may or may not be of prognostic importance equally across treatment groups, thus reducing the probability that factors other than treatment may influence the results. It is important to realize that randomization does not completely...
preclude imbalances. Differences in variables that are of prognostic importance may simply occur by chance. However, randomization precludes intentional imbalances (e.g., dissimilarities created by physicians who consider a particular treatment more appropriate for a particular patient [selection]).

From a statistical perspective, chance differences will occur more frequently in small study samples than in larger ones, and may be of higher magnitude in small trials. It is therefore necessary to compare treatment groups at baseline with respect to important prognostic variables, and to adjust for differences in the statistical analysis in case of doubt. Usually, computer-generated randomization lists are used to randomize patients in an RCT. Technically, randomization is often performed in “blocks,” so that in every block of four or 10, there will be equal numbers of patients in the treatment and control groups. Randomizing in blocks ensures that if the sample size is less than expected, an equal proportion of patients will be included in each treatment group. Often, in multicenter trials, one center is assigned one or more blocks, ensuring that the numbers of patients receiving the new drug and the control drug are evenly distributed per center.

Some trials randomly enroll patients in strata (stratification) of equal or unequal size. Stratification (a better wording is stratified randomization) makes sense only if the variable subject to stratification represents a prognostically important feature. An appropriate example is a situation in which circumstantial evidence suggests that the efficacy of a treatment is different in males as compared with females. Stratified randomization with “males” and “females” as strata implies that randomization to treatment groups occurs after assignment of the appropriate stratum. This allows a justifiable comparison between treatment groups within each stratum because there is prognostic similarity at baseline. Appropriate stratified randomization requires a trial design and a sample size that indeed allows such a comparison (see later discussion on statistical power).

Stratified randomization should be distinguished from post hoc subgroup analysis, in which the “strata” are determined during analysis of the trial. In such post hoc comparisons, prognostic similarity cannot be assured, and statistical adjustments can account for this only rarely.

**Design Considerations**

A fundamental choice in consideration of the design of the study is the decision about a superiority design or a non-superiority design. The latter theoretically can be further categorized as a noninferiority design and an equivalence design. The basis supporting this choice is the null hypothesis underlying the study. Consequences of the choice of design are important. If a new treatment is tested against placebo, the a priori hypothesis is that this new drug is more effective than placebo, and a superiority design is a rational choice. If for a particular disease or condition, treatments are already available, it is ethically often not justifiable to subject patients to a placebo treatment for longer periods. It is not always rational to assume that a new treatment will be better than the best available treatment at that moment, and a superiority design would have a high likelihood of failure. In such situations, one can opt for a non-superiority design. These designs have the underlying hypothesis that the treatment to test is at least not worse than (or equivalent to) the comparative treatment, which can be the standard of care, or alternatively, the best currently available treatment.

In a superiority design, the question is whether the new treatment is more efficacious than the control intervention (e.g., placebo). Formally, such a study tests whether the null hypothesis of no difference between both treatment groups can be rejected. To do so, investigators agree on a minimally clinically important difference (MCID) between the intervention of interest and the control intervention such that a study should be able to demonstrate, and they design the study in such a way that this difference can be demonstrated with high likelihood (statistical power) when it really exists (see later). In a noninferiority design, the reasoning is opposite. The null hypothesis is that the new treatment is less efficacious than the control intervention. Even if the new intervention and the control intervention are truly similarly effective, a trial will almost never yield a result with a treatment effect of exactly zero (no difference). There will be variation around zero, and it is the task of investigators to decide in the design phase of the study which deviation from a treatment effect of zero they will accept to conclude that the interventions are equivalent, the noninferiority margin. Determination of the MCID in a superiority design and the noninferiority margin in a noninferiority design is a subjective decision with important consequences for the sample size. When it is important in a superiority design to be able to demonstrate very small treatment effects with a high likelihood, large sample sizes are needed; the same is true with a very narrow noninferiority margin in a non-inferiority design. Especially with a noninferiority design, considerations other than efficacy alone may give guidance to the level of the noninferiority margin. If a new drug is less toxic or less costly than existing drug(s) on the market, and as such may provide additional benefits, one could be more lenient with regard to determining the noninferiority margin. In general, noninferiority designs require (far) more patients than are required by superiority designs.

**Subject Selection**

Subjects who are entered into clinical studies should meet accepted criteria for the disease or disorder under study. Most rheumatologic conditions lack single and unequivocal diagnostic tests, and classification criteria have been developed to identify patients with similar characteristics. These classification criteria serve as eligibility criteria in an RCT. To homogenize patient populations for scientific purposes, classification criteria are designed to be highly specific. As a consequence, sensitivity may fall short, and classification criteria are often of limited use in diagnosis. The high specificity of classification criteria has implications for the makeup of the trial population. In general, patients with classic, often severe disease are overrepresented, and those with early, less typical disease are underrepresented.

In many trials in rheumatology, patients must meet certain criteria for disease activity or duration. Some trials require that the patient experience a flare after withdrawal of medication as evidence of active disease. Other studies define disease activity before withdrawal of medication as
evidence of lack of response to current treatment. Disease severity can be defined by accepted clinical criteria or by lack of response to previous treatments. For example, RA studies may be limited to patients who have not yet received methotrexate (who presumably have relatively early disease or mild disease) or to patients who have failed treatment with at least one other disease-modifying antirheumatic drug (DMARD) (greater disease severity). There are ethical and methodologic reasons for the use of such activity and/or severity criteria. Ethical arguments may proscribe that a novel intervention is first tested in patients with severe, sometimes intractable disease in which common alternatives have failed. Methodologic arguments are that a treatment effect can best be demonstrated in a population of patients prone to change. Most inflammatory rheumatic diseases have a cyclic course characterized by exacerbations and remissions. Patients with a high level of disease activity will tend to improve over time, even without an intervention—a phenomenon known as regression toward the mean; the additional effect of a new intervention in comparison with a control treatment can be demonstrated more easily in such a context.

Exclusion criteria usually include conditions such as cancer; cardiac, hepatic, or renal disease; abnormalities in hematologic parameters, medication allergies, and pregnancy. Exclusion criteria serve to decrease background noise or variability due to differences in patient characteristics. In general, inclusion and exclusion criteria will homogenize the trial patient population and contribute to an environment that is most optimal to demonstrate a treatment effect. Inclusion and exclusion criteria also prevent entry of patients in whom an adverse response is more likely to occur and those for whom the experimental treatment could be dangerous a priori. As such, inclusion criteria and exclusion criteria contribute to a high level of internal validity but jeopardize external validity. Explanatory trials usually have a comprehensive set of inclusion and exclusion criteria. Pragmatic trials are more lenient in this regard because they should better reflect the common clinical practice.

Informed Consent

Ethical considerations determine whether eligible subjects participate in a clinical trial. Governmental agencies of most countries require that institutions involved in human research have a local institutional review board (IRB). The IRB reviews all protocols before implementation and monitors ongoing studies at its institution. A crucial element in the review of a trial is the informed consent process. The consent form should explain to the study participant the purpose of the study, all potential benefits and risks (including risks to pregnant mother and fetus), alternatives to participation, and who is responsible for conducting the study. Patient confidentiality should be ensured. The consent form should clearly state that participation is completely voluntary, and that refusal to participate or withdrawal from the study will not affect future care. If compensation is provided, this must be documented in the consent form. Participants should be given contact information for questions or in case of injury and a statement about whether any medical treatment will be given if injury occurs. Investigators are responsible for ensuring that the risk to subjects is minimized and appropriate for the anticipated benefits.

Follow-up Considerations

The optimal duration of the trial represents a compromise between economical, ethical, and methodologic considerations. A trial should not be too short because an intervention needs time to exert its potentially advantageous (but also deleterious) effects; in particular, a short trial does not reflect the clinical reality of most rheumatologic conditions. Equally, a trial should not be too long because RCTs are expensive, patients should not be subjected to experimental interventions with uncertain adverse events for an excessive time period, and too much longitudinal bias should be avoided. Longitudinal bias may occur if during the trial, the treatment groups increasingly become dissimilar as the result of selective dropout, co-interventions, or other patients’ or physicians’ behavior. Selective dropout may occur if patients with a particular profile preferably withdraw from one of the treatment groups, thus creating prognostic imbalance. A common example is that of an RA trial comparing an effective drug versus placebo. Patients with relatively severe and active disease in the placebo group may preferentially discontinue trial medication and may drop out because they do not experience benefit, while less severe and less active patients remain in the trial. Co-interventions—allowed or not allowed—may similarly jeopardize prognostic similarity if they occur in an unbalanced (i.e., unequal) manner across treatment groups. A common example is a trial that tests a nonsteroidal anti-inflammatory drug (NSAID) versus placebo with respect to the relief of pain. Simple analgesics, which are preferentially used in the placebo group, may inadvertently influence pain scores, leading to incorrect conclusions. The treating physician may contribute to prognostic imbalance by prescribing co-interventions, or in general terms by treating patients differently according to their clinical response or the occurrence of adverse events. A relatively short follow-up will decrease the likelihood that unintended events will occur and as such contributes to maintaining prognostic similarity and increasing internal validity.

Explanatory trials usually have a follow-up duration that is as short as possible, and co-interventions are prohibited. The most important limitation of short-term trials in lifelong rheumatologic diseases is that they do not appropriately reflect the course of the disease encountered in clinical practice. Sometimes, RCTs, especially pragmatic trials, have a long trial duration that better reflects the clinical reality. In such a trial, internal validity is deliberately sacrificed to some extent in favor of external validity (generalizability) and the yield of long-term information.

Blinding

In double-blind studies, neither the patient nor the investigator is aware of the treatment group assignment. In single-blind studies, the investigator is aware of the treatment assignment, but the patient is not. In open-label studies, both patient and investigator are aware of the treatment assignment. The most important reason to blind treatment allocation is to avoid that any expectation about the type
of treatment provided could influence the measured outcome (expectation bias), especially (but not exclusively) if the measured outcome includes subjective components. Note that subjective refers to the patient (e.g., pain scores) as well as to the investigator (e.g., joint scores). To avoid the latter, many drug trials make use of independent (joint) assessors, who are not responsible for decisions regarding patient care and are blind for the treatment. Another common example of a blinded independent assessor in rheumatology is the reader of radiographs in imaging studies in RA, psoriatic arthritis, or ankylosing spondylitis (AS).

Regardless of any precaution taken, unblinding may inadvertently occur because of identifiable adverse reactions or minor side effects, lack of efficacy, or changes in laboratory parameters. A meaningful effect of such a type of unblinding is not easy to prove, nor can it be adjusted for in the analysis.

**Choice of Outcome Variables**

Any clinical trial has one or more outcome variables of interest. Outcome is broadly defined and refers to a clinical situation or a change in a clinical situation that is quantifiable by using assessment instruments. Outcome variables can measure real outcome that directly affects the patient (e.g., vertebral fracture in osteoporosis), or alternatively can reflect a situation that is associated with real outcome but does not (yet) affect the patient (e.g., low bone mineral density in osteoporosis). The latter type of outcome is often referred to as surrogate outcome. Reasons to use surrogate outcome measures rather than real outcome measures are that the former occur (far) earlier and more frequently and can often be assessed on a continuous scale (which is a statistical advantage), and the latter often describe an event (the presence or absence of a clinical situation) with negative implications for statistical power.

The Outcome and Measurement in Rheumatology Clinical Trials (OMERACT) initiative was created to bring unanimity to the multitude of outcome measures in rheumatology on the basis of expert consensus. Its activities were initiated in RA and were expanded to include most other rheumatologic diseases. The OMERACT framework is the so-called OMERACT filter, which describes the methodologic prerequisites that an appropriate outcome measure should fulfill to be considered valid for clinical trials. The OMERACT filter prescribes three validation requirements: An outcome measure should be truthful, discriminatory, and feasible.

Truthful refers to whether an outcome measure truly measures what it is intended to measure, and approximates the concepts of face, content, and construct validity. It means, for example, that the disease activity score (DAS) in RA should truly measure what is considered important in RA (swelling and tenderness of joints) (content validity) and is a relevant construct to describe the process of RA, for example, because the disease activity score is associated with radiographic progression and limited physical function (construct validity). Discriminatory refers to whether an outcome measure can reliably be measured (intraobserver variation and interobserver variation), whether it can distinguish between two stages of the disease (e.g., RA with high disease activity vs. RA with low disease activity), whether it should be applied to groups of patients or to an individual patient, whether the measure is sensitive to change (e.g., whether the DAS measurably decreases if the disease improves), and whether the measure can discriminate between groups of patients on effective therapy versus those on placebo or less effective therapy. Feasible refers to whether an outcome measure is easily applicable and cheap in the setting in which it is intended to be used.

Ten biannual OMERACT conferences have resulted in sets of outcome measures for almost all inflammatory rheumatologic diseases and for numerous noninflammatory disorders. These so-called core-sets have importantly improved homogeneity across different clinical trials, thus favoring comparability. In the design phase of a clinical trial, it is highly recommended to choose a primary outcome measure from these core-sets, and to measure all components of the core-set as secondary outcome measures. Reporting of all core-set measures prevents selective reporting of only positive results with respect to a few variables.

Increasingly, indices are replacing single-outcome variables in rheumatology. An index is a weighted or unweighted combination of single variables that together reflect a particular domain of outcome. A general rule is that indices perform better than single-item variables only if they consist of variables that correlate moderately with each other. If variables correlate at a too high level, there is redundancy of information. If variables do not correlate, they will reflect different domains; this complicates interpretability, and it is better to separately describe them. Important examples of useful indices in rheumatology are the already mentioned disease activity score (DAS), the ankylosing spondylitis disease activity score (ASDAS), and the American College of Rheumatology (ACR) response criteria in RA.

**Measuring Effect**

After the outcome measures are chosen, it is important to consider how the change in outcome measures is measured in the clinical trial. One could simply calculate a before-after difference in a continuous variable (a change score), but this is statistically not necessarily the best approach. The ACR has developed the ACR20 response criteria as a tool to determine in clinical trials whether one drug is more efficacious than another or placebo. The ACR20 is an index that contains several outcome measures from the World Health Organization/International League against Rheumatism (WHO/ILAR) core-set for RA, which is based on expert consensus about different measures used to assess disease activity. The ACR20 response criteria require a 20% improvement and have been thoroughly validated in several validation steps, have been shown to perform better in trials than individual core-set measures, and currently are the standard for measuring drug efficacy in clinical trials. ACR50 and ACR70 response criteria have been derived from ACR20 response criteria in that they require a higher level of improvement. These derivatives have never been appropriately validated but have proved very useful in drug research, despite an inferior discriminatory potential in comparison with ACR20 response criteria. A specific limitation of the ACR70 is that the baseline disease activity needs to occur at a certain (high) level to actually allow a 70% improvement.
The European League against Rheumatism (EULAR) has endorsed its own response criteria in RA, which are based on the DAS. The DAS is a continuous index measure that includes four core-set measures in a statistically weighted manner. The advantage of the DAS, which also underwent numerous validation steps, is that it describes a state of disease activity rather than a response, or a change. The EULAR response criteria consist of categorical response criteria (good response, moderate response, and no response) that are based entirely on the DAS and require an absolute change in the DAS, as well as a certain state level of the DAS.

Another disease in which response criteria have been developed for use in clinical trials is ankylosing spondylitis (AS). The ASsessment in AS (ASAS) International Working Group has developed the ASAS20 response criteria, which include four patient-oriented disease activity measures for use in clinical trials with AS. ASAS has also defined cutoff values for continuous change over time in the previously mentioned ASDAS, reflecting clinically important improvements that can be used as outcome measures in clinical trials.

In general, the consensus-based response criteria have contributed to better trial design, better trial conduct, and better comparability across trials. An important drawback still is that interpretability of the different indices (“What does it mean that a patient has experienced an ACR20 response?”) remains difficult.

Sample Size and Statistical Power

Statistical power is the likelihood that if a treatment effect truly exists (or drug A is truly better than drug B), the trial will indeed demonstrate this with statistical significance. So a power of 0.80 (80%) means that if there is a true difference between the treatment group and the control group, the likelihood that this RCT will indeed confirm that difference is 80%. At the basis of this definition is the reasoning that truth (difference or not) is not known, that we never will know the truth with absolute certainty, and that we can only approximate the truth by performing RCTs. Intuitively, an RCT will not always give the right answer. An RCT may conclude that there is a difference, while in truth there is not. Or alternatively, a true difference will not be supported by the results of the trial. The former is termed a type I error, and the latter is a type II error (Figure 32-1). The theoretical problem is that by definition, we do not know which RCT gives the right answer and which does not. Although we do not know the true answer, we try to design the trial in such a manner that type I errors and type II errors are minimal. The likelihood of type I and type II error is less with larger sample sizes. As a consequence, the likelihood of correctly drawing a conclusion from the experiment of one RCT is increased by increasing sample size. Apart from sample size, the likelihood that a study can detect a difference between treatment groups also depends on effect size (i.e., if the treatment has a larger effect, it is easier to detect with smaller sample sizes) and reliability of the outcome measure (if the outcome measure is more precise, or is less influenced by measurement error, it is easier to detect with smaller sample size).

The probability of a type I error (false-positive result) is referred to as alpha, and is better known as the level of statistical significance. The probability of a type II error (false-negative result) is beta. Statistical power is defined as (1 minus beta). If beta is 10%, there is a 90% likelihood of finding a difference between groups of the expected magnitude or greater when this difference truly exists. Usually, alpha, or the p value, is set at 0.05, and beta is set between 0.2 and 0.1 (power of 80% to 90%). An ethical consideration is that type I error should be avoided because it may directly and negatively affect patient care (a new drug is falsely considered to be better than an existing drug). Type II error is subtler but should be avoided for more than one reason. First, new and effective treatments may not reach the market for false reasons, but more important, trials with a high probability of type II error are truly inconclusive. Because new drugs may cause harm, a trial that a priori does not allow a firm scientific conclusion is ethically unjustifiable and costs a lot of money with no yield.

Sample size is one of the most important determinants of the power of a study to find a treatment difference. To determine the appropriate sample size for a study, the investigator needs to make an estimation of the effect size of the intervention (i.e., the difference in outcome between treatment groups, the minimum clinically important difference [MCID]), the variability of the data (e.g., standard deviation), the statistical test to be used, and the alpha (p value) and beta (i.e., false-negative rate) levels. It is important to consider nonspecific effects (placebo effect, regression toward the means) (e.g., the proportion of subjects who meet criteria for improvement in the placebo group), which have been reported as high as 20% to 40% in studies of patients with RA. The variability or standard deviation of the outcome measure can be estimated from pilot data or from other published clinical trials.

Often the sample size in each treatment arm is the same, but unequal but proportional treatment groups can also be used (2:1 ratio of treated subjects to controls). RCTs with equal-size treatment groups have greater statistical power, but unequal groups are sometimes used to maximize the number of patients who receive treatment, especially if information on safety is the most important reason for the
trial; patients sometimes are more willing to enter a trial in which they have a greater likelihood of receiving an active treatment.

Declaration of Helsinki

The Declaration of Helsinki of the World Medical Association (WMA) (1964) is a document that spells out a set of ethical guidelines for physicians and other participants in medical research. It is considered the most widely recognized source of ethical guidance for biomedical research. The fifth revision (Edinburgh, October 2000) contains 32 paragraphs that aim to find a balance between the physician's duty to promote and safeguard the health of the people (paragraph 2), implying that the well-being of the human subject should take preference over the interests of science and society (paragraph 5), and the scientific and societal appreciation that medical progress is based on research which ultimately must rest in part on experimentation involving human subjects (paragraph 4), inevitably involving risks and burden (paragraph 7). The Declaration is not a static document, and current debate focuses on the place of placebo-controlled trials (paragraph 29). Whereas this paragraph, which raised a lot of argument, justifies placebo-controlled trials only in those circumstances in which no proven prophylactic, diagnostic, or therapeutic method exists, a Note of Clarification by the WMA outlines a somewhat more liberal interpretation of this paragraph, providing circumstances in which placebo-controlled trials are allowed, even if proven therapy is available.

It is generally accepted and required by governmental institutions and institutional review boards that trial design, trial conduct, and trial report are in accordance with the stipulations of the Declaration of Helsinki.

Place of Noninferiority Designs in Rheumatology

Paragraph 29 of The Declaration, issuing the place of placebo-controlled trials, has raised interest in designing noninferiority trials in rheumatology. The recently developed biologic therapies have had a very important impact on the treatment of chronic inflammatory diseases. If paragraph 29 of the Declaration of Helsinki is interpreted conservatively, this means that placebo-controlled trials ethically are not justifiable anymore in RA, AS, and psoriatic arthritis. It immediately follows from paragraph 30 of The Declaration (every patient entering into the study should be assured of access to the best proven therapeutic methods identified by the study) that future trials investigating new treatments in these diseases should include the best available treatments in the control arm. The introduction of these ethical principles will have methodologic consequences because RCTs with a superiority design become virtually impossible as a result of the high level of efficacy in the control group.

The only acceptable alternative is the noninferiority trial, which was briefly discussed previously. An illustrative example of a noninferiority design, emerging from the controversy about the cardiovascular safety of NSAIDs and cyclooxygenase-2 (COX-2) inhibitors, is the Multinational Etoricoxib and Diclofenac Arthritis Long-term (MEDAL) study program, which we will briefly discuss here. Although this program targeted cardiovascular safety rather than efficacy, the noninferiority principles underlying the study are applicable in different settings. The primary hypothesis of the MEDAL study was that treatment with etoricoxib was noninferior to treatment with diclofenac. As a noninferiority margin, a relative risk of 1.30 was chosen. It was defined before the start of the trial that the upper limit of the 95% confidence interval for the hazard ratio (etoricoxib vs. diclofenac) should not exceed the noninferiority margin to justify the conclusion that etoricoxib was noninferior to diclofenac with respect to causing cardiovascular events. This implies that the relative risk itself obviously should be far lower than 1.30 to justify a conclusion of noninferiority. In fact, it was assumed based on previous experience that diclofenac would yield a cardiovascular event rate of 1.3%. With approximately 40,000 patient-years of exposure, it was possible to calculate that the maximum absolute event rate in the etoricoxib group that would still meet the noninferiority criterion would be 1.46% or, in other words, an excess of 1.6 cases per 1000 patient-years of treatment. This example clearly illustrates that noninferiority trials always imply a certain expense (here, a number of excess cardiovascular events). This expense may possibly be limited by narrowing the bound of noninferiority, but this is achieved at the cost of increased sample size in an exponential manner. So, the noninferiority margin is determined as the result of careful considerations about drug safety (that require a margin very close to 1) on the one hand, and the statistical appreciation that demonstration of true equivalence is not possible on the other hand.

Conclusion

An appropriate trial design is of decisive importance to optimize the likelihood that the trial will provide results that are interpretable, robust, and applicable. Usually, decisions made during the design phase are a reflection of the continuous balance between internal validity and external validity. Which of both characteristics prevails in the ultimate design depends on the aims and the context of the trial. The ideal trial design does not exist. It is the challenge of investigators to find the most appropriate design for their goal.

TRIAL ANALYSIS

Hypothesis Testing

Suppose that a scenario with a particular disease for which an effective treatment A exists, and a new treatment B is proposed. It is not known whether the new treatment B will be more effective than the established treatment A, and a clinical trial should give resolution. It is useful to realize that this trial provides only an approximation of the truth that we—by definition—do not know. If you were able to repeat this trial many times, you may find a mean treatment effect across trials that is the best possible approximation of the truth, and most trials provide results that are close to this mean treatment effect. However, a few trials give more deviant results by chance. Because you will not be able to truly carry out this trial numerous times, you will have to
interpret the results of your particular trial in the context of a high number of illusory trials with exactly similar design. To do so, a null hypothesis of the truth should be carefully formulated. The character of the null hypothesis depends on the type of trial. In a classic superiority design, in which, for example, a new drug is tested against an established drug, the null hypothesis of the truth states that the new treatment is as effective as the established (control) treatment. In a noninferiority design, however, the null hypothesis of the truth states that the new treatment is less effective than the established treatment. Statistical analysis will challenge the actually observed trial results against the null hypothesis of the truth. Essentially, the statistical test provides the probability that a treatment effect (e.g., the difference in proportions of ACR20 responders between treatment groups, or the difference in the mean decrease in DAS) can be found that is as large as what was demonstrated in this trial—or even larger—while in truth such a difference does not exist. In the paragraph about statistical power, we have referred to this scenario as type I error, or alpha, and its likelihood is called the p value. Usually, we consider a probability of type I error of less than 5% (p < 0.05) sufficiently low to safely assume that the null hypothesis of the truth is not a likely hypothesis, and that it can be rejected. We now decide that the new drug is not as effective as the established drug. Note that not as effective may imply better than as well as worse than. Statistical testing does not differentiate between the two scenarios (two-sided testing). Crude data should provide the correct interpretation.

The advantage of thinking in probabilistic terms (this trial is only one example of many possible trials with similar design) is that you immediately accept that there is always a chance that, even with very convincing results, the interpretation of this trial may be false. The lower this chance is, the more convincing is the interpretation of trial results. We have also alluded to the scenario that a trial does not show a treatment difference, when in truth there is such a difference: type II error, or beta. Such a trial lacks (statistical) power. Sample size is most often of pivotal importance; a classic example of potential type II error is the small clinical trial with a superiority design that tests a new drug against an established drug. Such a trial may give a treatment effect that is of potential clinical interest, but the p value exceeds 0.05 (not statistically significant).

It is important to make a number of reservations clear here. In analogy with type I error, one can never be certain whether type II error is truly responsible for not statistically demonstrating a potentially important difference. One can only suspect type II error as a cause and approximate the probability that it occurs (power calculation). A priori power calculations during the design phase inform you about the probability that the designed trial will statistically "help you" (rightfully reject the null hypothesis) if a particular treatment effect is demonstrated. Post hoc power calculations can be made after the completion of the trial, using the actually demonstrated treatment effect. Post hoc power calculations inform you about the power of this trial to statistically support the observed difference if this difference would have been considered clinically important before the trial. Insufficient post hoc power (e.g., 0.30) is an indication that type II error is operative, but it does not prove type II error.

**Intention to Treat**

A classical clinical trial with a superiority design is analyzed on the basis of intention to treat (ITT) by default. ITT means that every patient is analyzed as belonging to the group to which he or she was allocated by randomization, regardless of what happened with this patient afterward. In the extreme scenario, a patient who is randomized to group A may drop out even before the first drug dose is administered, and may be treated outside the clinical trial by his or her physician with the drug that is actually being administered in group B (cross-over of treatment). True ITT requires that this patient, regardless of the data that are present or missing, is analyzed in group A, even if he or she has experienced the effects of the drug in group B. ITT is considered the only means of analysis that preserves the prognostic similarity that was created at baseline. It should be noted here that many trial reports mention an ITT analysis, but that such an analysis often is not rigorously performed. For example, randomized patients who did not receive treatment are often excluded from the analysis.

Alternatively, a trial analysis may be limited to only those patients who completed the study (completers analysis), or to those who completed the study while complying with the study protocol (per-protocol analysis). Both types of analysis may introduce bias; as a consequence, prognostic similarity should not be assumed. Dropout, for example, can occur because of lack of efficacy, and completers may provide a less severe representation of the entire trial population (see later). Often, a completers analysis tends to magnify the treatment effect, and an ITT analysis is more conservative. In a noninferiority trial, however, the situation is just the reverse: The ITT analysis tends to favor noninferiority, and the completers analysis or the per-protocol analysis is more conservative in this regard.

**Problem of Incomplete Data**

One of the major threats in the interpretation of results of clinical trials is early withdrawal or dropout. Early withdrawal may occur for a variety of reasons. Some patients withdraw consent immediately after randomization, even before the first drug dose is given, because they simply changed their mind. Patients may drop out because of actually occurring or feared adverse events, because of lack of efficacy (no response, disease progression, unrealistic expectations), because of a combination of both, or because they have passed away or have relocated. Usually, dropout results in missing data, in that most patients do not come back for outcome measurement. Often, patients who formally withdraw are encouraged to continue trial assessment, but it is difficult to decide how to handle the data of these patients because they often are treated with different drugs outside the trial. The trial analysis suffers from missing data in several aspects. First, the individual response or disease course cannot be calculated anymore, but leaving out the entire patient data jeopardizes statistical power, especially if the number of withdrawn patients is high (e.g., >20%). Second, it is important to realize that early trial withdrawal is not a random process. It is easy to imagine that the most severely afflicted patients are at risk for efficacy failure; this may lead to selective dropout in a trial with high treatment
contrast, or may cause certain subgroups of patients to be at risk for particular adverse events that occur preferentially in one of the trial arms.

The net result of these inadvertent selection processes may be that treatment groups that were entirely similar at baseline lose their prognostic similarity during the trial. Indicators for selective dropout include differences in rates of dropout between treatment groups, differences in reasons for dropout between treatment groups, a high rate of dropout for a specific reason that is reducible to a specific treatment, and so forth. Usually, missing data are handled in the database by data imputation to keep the patient in the trial analysis. *Data imputation* means that a missing data point is supplied by an imaginary value. Several means of imputation are available, and no consensus has been reached about the best way to impute. Most likely, the best imputation method depends on the characteristics of the outcome measure, such as the natural course of this outcome measure. *Last observation carried forward* (LOCF) imputation is a frequently applied method of imputation in which the last truly measured value is imputed (carried forward) in subsequent (missing) data points. Other imputation rules include imputation of a mean group score of the same or the comparator group, imputation of the 95th percentile, and linear interpolation or extrapolation. More sophisticated multiple imputation techniques impute figures that are considered most likely on the basis of regression analysis of the present data. It is not easy to predict how different imputation rules affect study results. Increasingly, investigators perform sensitivity analyses with different imputation rules to challenge the robustness of their trial results. A trial result that is robust to various means of imputation has more credibility than a trial result that is dependent on the means of missing data imputation.

**Presentation of Trial Results**

An increasing number of journals require the presentation of trial results in accordance with consensus guidelines (such as the CONSORT guidelines) to increase comprehensibility and to maximize information. Such guidelines require exact presentation of the randomization process, a description of blinding, eligibility criteria (about inclusion and exclusion), and many others. An appropriate trial report should include exact information about the fate of all patients after randomization, including the major reasons for withdrawal. It is essential that the total numbers of patients per group can be reconstructed. Increasingly, such information is provided in a flow chart.

The initial part of data analysis involves examining the baseline characteristics of the trial groups, including demographics, previous and current treatments, and disease characteristics (e.g., severity scales, duration, extent of organ involvement) with descriptive statistics. Occasionally, baseline characteristics per group are statistically compared, and baseline similarity is assumed if no statistically significant differences between groups can be shown. Such an approach is rather useless and sometimes is overtly false. Groups are similar by definition because they were formed by randomization. Randomization is a probabilistic procedure, and groups may statistically differ in one or more variables at baseline just by chance. Very often, such differences are small in relation to the treatment effect, or the particular baseline characteristic is not associated with the measured outcome and therefore is not contributory. On rare occasions, baseline differences are not negligible, and the trial result should be adjusted for these differences by multivariate analysis (e.g., regression analysis, analysis of co-variance). A particular problem may occur in small trials, if even clinically important differences at baseline are not statistically significant, but their impact on the treatment effect may be substantial. A general rule is to “eyeball” baseline differences and to adjust the statistical analysis for those variables that show potentially important differences.

**Descriptive Analysis**

Simple descriptive analysis includes the presentation of means and standard deviations in cases of normal (parametric) distribution of continuous data (e.g., DAS, ASDAS), or medians and key percentiles in cases of not-normal distribution of continuous data (e.g., Sharp scores). Dichotomous data (e.g., ACR20 responses) are presented as proportions or rates, and ordinal/nominal or categorized data as percentages per category. Graphic representations (e.g., line graphs, bar graphs) are preferred because they give the clearest representation of the treatment effect. An illustrative example of visual presentation of data is the use of probability plots for radiographic data. Emphasis should be on the primary outcome variable, but all measured outcome variables should be presented. It is recommended to present data such that they can be used post hoc for systematic reviews or meta-analyses. This implies that status scores plus a measure of variability (e.g., mean DAS at baseline and at follow-up, standard deviations) and change scores plus a measure of variability (e.g., mean change score, standard deviation of change score) should be presented for primary and secondary outcome variables. Dichotomous outcome variables (e.g., ACR20 response criteria) should be accompanied by extensive information about the status values of (separate) variables. It is relevant to present not only response measures (e.g., ACR20) but also state measures (e.g., absolute DAS) because the combination of the two yields additional information and increases the interpretability of trial results. A useful extension of state measures is the concept of patient-acceptable symptom state (PASS). The PASS reflects that level of symptom severity that best discriminates an acceptable situation from an unacceptable situation from the perspective of the patient. PASS levels can be determined by different methodologic methods and are presented as proportions or rates.

**Statistical Analysis**

Because the process of randomization provides prognostic similarity at baseline, the statistical analysis of a clinical trial is in fact simple, as long as one assumes that prognostic similarity is preserved during the trial. The statistical test of choice is a test for binomial data (e.g., Chi square test) if the outcome measure of choice is dichotomous (e.g., ACR20, ASAS20 response criteria), and a test for continuous data (e.g., Student t-test, Mann-Whitney U test) if the outcome measure is continuous (e.g., DAS, ASDAS). An extension of the statistical test that provides useful
information to increase interpretability is the 95% confidence interval (95% CI) of the treatment effect. The 95% CI can be calculated for the treatment effect measured by continuous and dichotomous outcome variables. It provides the range within which the estimate of the treatment effect will lie in 95% of cases in the imaginary situation that this trial will be repeated numerous times. If the lower limit of the 95% CI of the mean treatment effect in a trial comparing drug A and drug B does not cross zero, this means that 95 out of 100 times that this trial is repeated, drug A is better than drug B, but 5 times, drug A is not better than drug B. Note that the formulation of the 95% CI closely resembles that of the p value.

**Statistical Significance and Clinical Relevance**

Clinical trials can be designed in such a way that even very small treatment effects can be statistically demonstrated. Explanatory trials in which new drugs are tested are often performed in a highly selected population of patients with strong adherence to the protocol, without co-morbidity, and with a high probability of responding. Sometimes, sample sizes are huge (“overpowered”). Such trials aim at a high level of internal validity and a low probability of type II error, but it is difficult to immediately interpret the results in the clinical situation, because the mean treatment effect can be so small that it is considered not relevant in the context of clinical practice, and patients in the trial often do not resemble the individual patients in practice (external validity). Guidance is available to evaluate the quality of a clinical trial. For example, a trial may demonstrate a small improvement in a primary outcome criterion that is statistically significant, but the effect may not have clinical importance because it is not sufficient to impact quality of life or survival, or it does not outweigh the risk or cost of treatment. A result may be statistically and clinically significant but may have little medical relevance because the benefit does not outweigh the risk or cost of treatment, or because the benefit is seen only in a very small subgroup of patients.

**Confounding**

Confounding is a special type of bias (systematic error) that can occur in a trial when the trial groups differ with respect to a particular factor other than trial treatment (prognostic dissimilarity). If this factor is also related to the outcome measure of interest, a fake treatment effect may emerge that will erroneously be attributed to the difference in treatment (confounding). As long as confounding factors are known, measurable, and indeed measured, one can adjust for them in the statistical analysis. But if such variables that are referred to as prognostically important are not measured (or even are not known), adjustment is impossible. The likelihood of confounding is far greater in observational studies than in randomized trials, and we will use an example from an observational study to clarify matters. If in an observational cohort study, the efficacy of the DMARD sulfasalazine (SSZ) in the treatment of RA is compared with that of methotrexate (MTX), and it is common practice to give SSZ treatment primarily to less severe patients (e.g., rheumatoid factor [RF]-negative) and MTX to more severe patients (RF-positive), radiographic progression may be less in patients treated with SSZ than with MTX. It is difficult to determine whether this difference is due to differences in efficacy between SSZ and MTX, or to differences in the severity of RA that may also drive radiographic progression (prognostic dissimilarity). Variables such as RF in this example may confound the relationship between treatment and outcome (radiographic progression). In this example, with a well-known confounder, the analysis may adjust for differences in RF positivity. However in theory, many unknown variables, or known but unmeasured variables, can cause prognostic dissimilarity.

We have mentioned previously that treatment groups in an RCT are only prognostically similar at baseline, and that prognostic similarity can be lost during follow-up. As a consequence, confounding is possible in RCTs, and trial results should be judged in light of this possibility.

**Interpretation of Safety Analyses**

Safety is considered extremely important in the consideration of whether a new drug or treatment should be approved. A detailed description of the process of drug approval is beyond the scope of this chapter, but there are a few methodologically important issues with regard to the interpretation of safety data in clinical trials. Usually, clinical trials aim at demonstrating efficacy of a drug or treatment. It is important to realize that many relevant adverse events occur in a relatively low frequency and/or (far) beyond the duration of the trial. Consequently, the probability that such a relevant adverse event will occur within the context and time frame of the clinical trial is low, and the interpretation of safety results of a clinical trial does not at all exclude important adverse events. Such information should be obtained through long-term observational studies or drug registries.

**Conclusions**

The descriptive and statistical analysis of clinical trial data is not an extremely challenging task and should follow the straightforward protocol imposed by the trial design. Universal guidelines are published to guide the investigator, and these guidelines are increasingly warranted by medical journals. This does not mean that interpretation of trial results is always easy and straightforward. We have mentioned a number of disturbing factors that may jeopardize the interpretation of trial results, such as missing data, dropout, and confounding. Investigators as well as the readers of medical journals should be challenged to interpret trial results in light of these potentially disturbing factors.

**GENERAL REMARKS**

Clinical trials are conducted to test the efficacy of new drugs and devices and to compare the efficacy and safety of combinations of drugs. New drug development is a long and expensive process. The choice of clinical trial design and implementation is critically important for safe, efficient, and successful drug development. Because the cost of clinical trials for new drug development has increased substantially, research leading to regulatory approval of new treatments is done primarily by the pharmaceutical industry.
in multicenter clinical trials. Such trials usually are explanatory in nature, with internal validity prevailing over external validity. It is widely recognized in medical science that investigator-initiated clinical trials are crucial in studies addressing the effects of combinations of standard drugs, or standard drugs in combination with new treatments, and in initial studies of the effects of drugs newly approved for other indications. Such trials often are more pragmatic in nature and have a higher level of external validity (generalizability). Important contributions in the field of investigator-initiated pragmatic RCTs have been published in the rheumatologic literature. It will be a considerable challenge to fund such trials in the future; this need may provide an opportunity for collaboration between academic clinical scientists and the pharmaceutical industry.

The clinical trial is not the only type of clinical research study. An important drawback of clinical trials is that the duration is short, and as a consequence, investigators have to rely on intermediate or process measures for outcomes, rather than the “hard” outcomes themselves. Longitudinal practice-based observational studies may fill in a gap in that they provide important information about the effects of a new treatment when used in diverse groups of patients, about drug toxicity, and about the long-term effects of a treatment on functional status, morbidity, and mortality. They may provide a wealth of information about the relationships between process or surrogate measures and “true” outcome measures. They also may provide useful information about prognostic factors for a certain outcome. A general judgment about the effectiveness of a particular treatment is in the end a compilation of impressions obtained from various sources, including explanatory drug registration trials, pragmatic trials better meeting the needs of clinicians, and observational studies with a focus on “true” outcomes and long-term safety.

References


The references for this chapter can also be found on www.expertconsult.com.
Assessment of Health Outcomes

DORCAS E. BEATON • MAARTEN BOERS • PETER TUGWELL

KEY POINTS

- Any single health outcome can only provide a particular view of the impact of a disease on a person.
- Core sets are minimal, but not exclusive, domains of outcomes agreed upon by professional groups as important to include in studies. They are available for several rheumatologic conditions.
- Defining the measurement need is the key to the choice of the right instrument.
- Choosing an instrument follows a step-by-step process—looking for evidence of practical aspects of using the instruments and methodological/statistical properties.
- If an instrument lacks evidence of a certain property, one can conduct a study to create the evidence rather than abandon the instrument.

In an era of rising health care costs, greater provider accountability, and an increased emphasis on decision making based on patient-reported outcomes, the capacity to discern the best outcomes and the best instruments has become a skill needed by researchers, clinicians, and funding bodies. By one definition, health outcomes refer to “all possible effects of a disease or intervention,” in our case for a disease like arthritis and the endpoints used to evaluate its treatment. In addition, there are biomarkers that “mark” a biologic process (e.g., decreased inflammation) and have some relation to health outcome. In a few instances, biomarkers can be regarded as surrogate outcome measures where the relationship with outcome (and change in outcome) is strong, and interventions that target the biomarker result in improved health outcome. Other chapters in this text refer to some of the most common instruments of health outcome and disease encountered in rheumatology such as the Disease Activity Scale (DAS, DAS28), the Health Assessment Questionnaire (HAQ), and the SF-36. Many more instruments of this type exist, but how different are they? Why do we need to choose so carefully? Using any one of these health outcome assessments is like looking out a window in a house, and the burden of arthritis is the landscape outside. From any one window there is a view of the outside world, but it is a specific view defined by the size of the window and the side of the house it is on. Another window may offer a slightly better perspective on what one would like to see. Different health outcome assessments can have a degree of overlap in their views, in which case an informed choice will need to be made between them, whereas others can hold quite distinct views. Although one outcome assessment might be useful to compare the burden of arthritis against the general population, another might be better for measuring the specific benefits of an arthritis intervention. No one instrument fulfills all requirements. To carry forward the metaphor of a window, once it is clear whether that instrument can offer a view of the target concept, one must also make sure the view is clear, precise, and consistent each time any person looks through that window—attributes that are shown by the validity and reliability of a scale. This chapter focuses on describing the different windows we have on the burden of arthritis and how they relate to each other. We then provide a framework for ensuring that a selected instrument is the right one for a given need. This chapter therefore addresses three questions: Which health outcomes assessment instruments are available, both generally and specifically, for use in rheumatology? How does one know what one needs to measure? How does one find an instrument that can meet that need?

WHICH HEALTH OUTCOMES ASSESSMENT INSTRUMENTS ARE AVAILABLE?

In reading the rheumatology literature and in monitoring the clinical care of patients, certain highly relevant outcomes will emerge. A group of these often emerges and are called “core sets” of outcomes.

Disease-Specific Instruments: Core Sets

Core sets are the minimal, but not exclusive, set of domains to be measured in a study of arthritis. Historically, they follow the Ds of outcome measurement in arthritis: disability, disease activity, damage, discomfort, dissatisfaction, and death. They are usually recommended by groups such as Outcome Measures in Rheumatology (OMERACT), the European League Against Rheumatism (EULAR), International League of Associations for Rheumatology (ILAR), and American College of Rheumatology (ACR) or by groups formed around specific diseases such as the Assessment for Ankylosing Spondylitis (ASAS) and the Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA). All have an interest in agreeing on a common set of relevant and psychometrically sound outcomes that would allow them to compare findings across studies and modalities of clinical care. Table 33-1 shows core sets for clinical trials in nine types of arthritis. It also shows what each group recommends as additional
Table 33-1  Core Sets for Six Rheumatologic Conditions and Longitudinal Observational Studies*

<table>
<thead>
<tr>
<th>Longitudinal Study Core Set of Domains</th>
<th>Rheumatoid Arthritis</th>
<th>Osteoporosis</th>
<th>Osteoarthritis</th>
<th>Systemic Lupus Erythematosus</th>
<th>Ankylosing Spondylitis</th>
<th>Psoriatic Arthritis</th>
<th>Vasculitis</th>
<th>Fibromyalgia</th>
<th>Gout</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical Trial Core Sets of Domains by Disease Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Health Status/Quality of Life</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of life Symptoms</td>
<td>R (utility)</td>
<td>R (BL)†</td>
<td>R</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>√ Pain</td>
<td>R (fatigue)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical function</td>
<td>√ Disability R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychosocial</td>
<td></td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Disease Process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregate index Biomarkers</td>
<td>√ (EULAR DAS)</td>
<td>√ Biochemical</td>
<td>O</td>
<td>√ DAI, R = severity</td>
<td>Pending</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint tenderness</td>
<td>√ 28 or 68 joints</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enthesitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint swelling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint stiffness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td></td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute-phase reactants</td>
<td></td>
<td></td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Damage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiography or imaging</td>
<td>√ &gt;1 yr</td>
<td>√ Bone mineral density</td>
<td>√ (&gt;1 yr)</td>
<td>√ Spine and hip§</td>
<td>√ Structural</td>
<td>R</td>
<td>R</td>
<td></td>
<td>O</td>
</tr>
<tr>
<td>Deformity Surgery</td>
<td></td>
<td></td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organ damage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiation therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toxicity effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dollar Costs/HCU [R]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work disability [R]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The left-hand column is the broader set of measures recommended by Wolfe and colleagues for longitudinal studies. It serves here as a broader range of outcomes and an axis for the organization of core outcomes in the other conditions (columns).

Osteoporosis: Core set depends on focus: BL, bone loss studies; †, studies aiming to reduce fracture rates.

Ankylosing spondylitis: Core set elements that vary depending on focus of study: † clinical records and symptom modifying; § disease modifying only; others = all.

†, core domain; DAI, disease activity index; EULAR DAS, European League against Rheumatism Disease Activity Scale (revised = DAS-28, 28 joint count); HCU, health care utilization; O, optional outcome; R, recommended for further research and possible inclusion in core set.
domains or as needing more research before they become core set members. Some groups such as systemic sclerosis are moving through the refinement of their core domains.\textsuperscript{35} The first column in this table represents Wolfe's broad core set for longitudinal observational studies in rheumatology\textsuperscript{20} with some minor additions. Wolfe and colleagues' list is longer than the other columns because observational studies are often looking for a broader range of outcomes than treatment trials. It is also designed to be used across different forms of arthritis. The Wolfe set therefore serves the broader list of outcomes against which we can describe what is also included in the more focused core sets. The remaining columns show that across different types of rheumatic diseases, the core sets have many common elements—most contain or recommend pain, physical function, patient and clinician global assessments, and markers of inflammation. Many also include disease activity and/or damage indices, which are often an aggregation of other clinical findings (e.g., joint count, acute phase reactants, global ratings of severity) into a score reflecting the activity of the disease at that point in time. Some core sets contain domains reflecting the unique aspects of the disease (e.g., spinal mobility in ankylosing spondylitis)\textsuperscript{26} or the unique target of the study (e.g., tophi in measuring response in gout).\textsuperscript{23} Table 33-1 focuses on the core domains that should be measured. The next step is to decide on the instrument(s) that will be able to provide that information in a reproducible, accurate manner. In some cases an instrument choice has been suggested (e.g., the HAQ for disability in rheumatoid arthritis [RA]). In other instances, several options are provided. Strand reviewed six disease activity indices in lupus and found that they gave comparable results.\textsuperscript{27} In some cases, the domains are shared but the measurement technique varies within or by disease—in RA the DAS28 uses 28 joints,\textsuperscript{8} whereas in ankylosing spondylitis 44 joints are counted.\textsuperscript{26} We briefly review some of the more commonly encountered instruments in arthritis.

**Health Status/Quality of Life**

**General Health Status.** Generic health outcomes provide information on an aspect of health across many conditions, so theoretically comparisons can be made to compare the burden of low back pain with that of arthritis or diabetes. This depends on how well an instrument captures the burden in a disease group. Generic instruments have the advantage of allowing comparisons across diseases and covering a broader range of health issues, which may otherwise be overlooked in a core set (e.g., mental health). However, generic instruments, due to their breadth, tend not to delve sufficiently into the depth of experience in any one disease. Arthritis-related fatigue, for example, is not picked up well in many generic instruments because they ask about “being tired” or “not sleeping well,” rather than the pervasive nature of exhaustion described by patients with arthritis.\textsuperscript{29} As a result, they are usually weaker in their ability to detect specific changes and their sensitivity to different levels of disease activity may be low. They should, therefore, usually be supplemented with disease-specific instruments.\textsuperscript{29}

Two of the more commonly used generic instruments are the Sickness Impact Profile (SIP)\textsuperscript{30} and the SF-36 (short form, 36 items).\textsuperscript{31} The SIP is a 136-item list of illness behaviors that provide a weighted score for the impact of a disease across 12 categories such as bodily pain, work and role functioning, and dressing,\textsuperscript{32} which lead to global scores as well (physical, psychosocial, and overall). The SIP has been shown to measure illness across a wide variety of health conditions.\textsuperscript{29} The SF-36 is a 36-item questionnaire of which 35 items are used to obtain 8 domain scores including physical functioning, mental health, role functioning, and pain. It is scored on a 0 to 100 scale (100 = better health)\textsuperscript{34} and two summary scores (mental and physical) that are scored with a normal of 50 and standard deviation of 10. The SF-36 and the briefer SF-12 are supported on the website www.qualitymetric.com and through manuals that supply age- and disease-group distributions of scores.\textsuperscript{32} Direct comparisons of generic instruments have shown differences in scores and health states attributable to the choice of instrument.\textsuperscript{33,35} Studies or clinical results may not be comparable with each other if they are using different health status scales.

**Utilities: Value of Health State.** Utility scales offer an overall score for the value of a health state, setting death at zero and full health at one. The emphasis is not on describing the state but on assigning a value, worth, or preference to that state.\textsuperscript{36,37} Utilities are necessary for economic appraisals and form the health assessment for cost per (quality-adjusted life years) QALY estimations. Utility states can be obtained by direct or indirect methods. Direct methods such as standard gamble and time trade-off involve the respondent working through exercises to elicit the value for his or her own health state against elements like time, or more/less favorable health situations.\textsuperscript{36} Indirect methods capture the state with standardized questions and then apply predetermined weights.\textsuperscript{37} Examples include the EQ-5D, which comprises five items (three response categories) combined to describe a health state. Similarly, the Health Utility Index (HUI) gathers information on six or seven dimensions of health (depending on the version) on five-item to six-item response scales to define a health state.\textsuperscript{37} Both forms, along with the increasingly popular Short-Form Six-Dimensions (SF-6D) utility index,\textsuperscript{38} then use weights determined in different populations to assign the value to these health states, hence the “indirect” weighting. The absolute values obtained across these different approaches will vary.\textsuperscript{36,39}

Generic measures of health and utility scores are broad. They often do not perform as well as the more specific measures described in the following sections because they are designed to allow comparisons across different patient groups and need to, therefore, include items that might not be relevant or amenable to change in arthritis.

**Symptoms.** Pain is usually measured using a 10-cm visual analog scale or a 0- to 10-point numeric rating scale of the intensity of the pain.\textsuperscript{40} These scales, simple instruments, have been well tested and are easily understood by patients. Fatigue is another important symptom, which many patients feel is quite distinct from being “tired.” Teams are recommending either global indices or one of several available scales that were reviewed by OMERACT attendees.\textsuperscript{42-44} Work done at OMERACT on the measurement of problems with sleep provides a recent strong example of moving through the concept
of impairment of sleep, defining it, and then focusing on the available scales that capture that concept and definition.

**Disability Scales.** Physical disability caused by RA or osteoarthritis is often measured using the Health Assessment Questionnaire–Disability Index (HAQ-DI), which covers 20 items examining different domains of daily functioning. Patients score each item on a 0–3 point scale, where 3 represents the greatest disability. Scores are obtained for each domain and then combined for a total score expressed on the same 0–3 point scale. Scores are adjusted to a worse health state (a 2/3) if a support is used to complete a task. More details on the HAQ-DI are widely available in print and on the Internet.

Other scales or subscales assess physical function such as within the Arthritis Impact Measurement Scale (AIMS) and the AIMS2, as well as measures with even more specific foci such as the Western Ontario and McMaster osteoarthritis index (WOMAC), which is commonly used in hip and knee osteoarthritis and the AUSCAN (Australian-Canadian) osteoarthritis index for hand osteoarthritis (OA).

**Disease Process (Activity, Severity)**

Core sets often include indices of disease process, which can be divided into activity (inflammatory activity) and severity (overall severity of disease) measures. There are several disease activity indices, the most commonly used being the Disease Activity Scale (DAS) and DAS28 in RA. A subset of the core outcomes (i.e., acute-phase reactants, joint counts, global ratings) was combined to form a weighted score that provides a score of 2 to 10 (DAS) or 0 to 9 (DAS28). Based on these scores, cutoffs were established to define high, moderate, and low disease states. Until recently the low disease state (DAS28 < 2.6) was considered an indicator of remission of arthritis. This is revisited later. Also recently, new criteria for remission in RA have been proposed. In these, the use of the DAS has been abandoned because it allows significant residual disease activity even at low values. Disease activity indices track the level of inflammatory activity. Other examples of disease activity indices include the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and the six available for lupus. When more than one is available, it is helpful to seek direct comparisons of instruments such as that performed by Strand to evaluate comparable information. Work is also under way to focus on an evidence-based, consensus-driven definition for a flare or worsening of arthritis to use in clinical trials to describe a worsening rather than improving situation.

**Damage Indices**

Damage indices are indicators of structural damage to joints, typically shown by joint space narrowing, erosions, subchondral cysts, or osteophytes. In RA, particular attention has been paid to this by van der Heijde, who reviews three approaches (Sharp, Larsen/Scott, and van der Heijde) that are used to assess joint damage and progression in joint damage. Progression is usually measured using the “smallest detectable change” that is determined by the threshold or error associated with the amount of variability in radiographic measurement between observers.

**Toxicity/Adverse Events (Wolfe’s “Disadvantage” Category)**

Medical and nonmedical management of many rheumatic conditions carries a risk of toxicity and adverse events, many of them unexpected. Because patients, clinicians, and policy makers want to balance benefits versus harms when considering intervention, comprehensive documentation of a range of adverse events is important in outcomes assessments separate from the treatment benefits. An OMERACT group is currently working on standardizing reporting of toxicities in rheumatologic trials.

**Death**

Arthritis is associated with increased mortality, and arthritis-specific mortality should be monitored. Death is not specifically mentioned in the disease-specific core sets for clinical trials because the latter are too short-term, but deaths would be important to monitor in observational studies. Attributing death to arthritis is challenging given its dependence on documentation at time of occurrence, which may or may not be linked to underlying arthritis in the coding of cause of death.

**Dollar Costs**

Given that resources are always limited, together with the benefits and harms, the costs (e.g., dollar costs) of a treatment are also considered an important outcome. Harmonization is important in order to get a comparable estimate of cost across studies. Several groups are working toward developing a formula for quantifying the costs associated with arthritis and with its treatment at the time of writing.

**Other Outcomes of Interest**

**Self-Efficacy/Effective Consumer**

Self-management is becoming incorporated into programs for persons with many chronic conditions. For patients with arthritis, these programs have been shown to improve levels of self-efficacy; that is, the confidence one has in one’s ability to manage pain and disease effectively. Lorig’s Self-Efficacy Scale is one of the most commonly used outcomes for this type of study. Tugwell’s group has developed a companion to this, the “effective consumer scale,” which captures the degree to which the patient is effectively managing his or her own health care decisions, interactions with the health care team, and disease monitoring. It is an instrument with demonstrated reliability, validity, and responsiveness in arthritis and has the support of consumers with arthritis.

**Work Disability: Looking beyond Absenteeism**

With the shift toward more aggressive management of earlier rheumatic disease, more people with arthritis are
working and outcomes need to shift to track work disability more comprehensively (absenteeism and at-work productivity loss). Work is difficult to measure because it depends on the job and the organization in which an individual works. Absenteeism can mean many things, and instruments should articulate how it is operationalized such as full days off work, days on insurance payments, or partial days off work. More challenging still is measuring the difficulty someone is having at work (presenteeism). In a recent review we found more than 20 instruments available. The most commonly used in arthritis is the Work Limitations Questionnaire (amount of time experiencing difficulty). Two scales developed in arthritis are promising: Gignac’s Work Activity Limitations Scale (amount of difficulty experienced) and Gilworth’s Work Instability Scale (captures risk of future work loss). The Work Productivity Scale for Rheumatoid Arthritis has been used in clinical trials and captures worker productivity on a global index, as well as nonpaid work (see next section). Direct comparisons of instruments show differences in their performance that primarily relate to the variant of the construct they are measuring.

Nonpaid Work Roles

Participation in valued nonpaid roles such as parenting, volunteer work, or leisure activities can be important aspects of the burden of disease. Outcome instruments reflecting this are necessary in order to fully capture the concept of participation as described by the International Classification of Functioning (ICF).

Patient-Specific Indices

Patient-specific scales including the MACTAR or PET in arthritis allow the patient to nominate his or her own scale content within a guided framework. Most patients report three to five items that are particularly salient to them. A surprising number of these scales have been developed. Each taps relevant content for patients, and because of this they are also responsive to change. The challenge is in the mathematics and how to analyze the numeric score that is so dependent on each individual’s own items across patients (group level mean or average score has little meaning). Analysis that focuses on individual level quantification is likely best (e.g., percent of people reaching their goal, improving in their selected activities).

Satisfaction with Health Outcomes

Satisfaction scales are often linked with the goals of a health care organization and focus on the attributes of the structure and process of care (e.g., length of wait, professionalism of staff). However, instruments developed to look at satisfaction with a specific health end point (i.e., how satisfied are you with the results of your surgery?) become a health outcome. Satisfaction with outcome is complex, and Hudak and colleagues remind us of the complex balance between experiences and ability to “live with” ongoing limitations that influence a patient’s response. However, it may be worth the effort in order to report all outcomes that are meaningful to the patients.

HOW TO DETERMINE WHAT TO MEASURE: DEFINING ONE’S MEASUREMENT NEED

Just as investigators define a research question before embarking on a clinical trial, so should users of health outcomes define a measurement need before choosing an instrument. There are three parts to this definition: what, why, and in whom?

What Is Worth Measuring?

It is important to have a sense of the concept one wants to measure before choosing an instrument. Taking this approach ensures that an appropriate instrument can be chosen and the user will be less swayed by the concept offered by an available instrument.

Defining a need is not always as easy as it might seem. What is health? What about pain? In recent years, there has been a shift in the outcomes toward a more patient-centered focus both from a health systems and political point of view, reinforcing and, indeed, legislating the importance of patient-centered care and outcomes.

With such a focus, the role of patient as partners in research and outcome definition becomes even more important. Partnerships with patients have opened many doors to improved outcome measurement. There is no better source of information on the nature of a symptom experience or the impact of arthritis on something like work or fatigue. The patient experience of disease can complement that of the researcher. The idea is that research grounded in relevant clinical need, patients’ perspectives, and patients’ priorities will enhance study design, practicality, recruitment, data interpretation, and dissemination. In addition, the reporting of the research results will most likely be more meaningful to patients because it will be done in terms that the patient understands and that are relevant to the patient. In outcome measurement, rheumatology has been in a leading position since the inclusion of patient partners (as experts in the experience of arthritis) at OMERACT conferences from 2002 onward. Perhaps the most concrete example of the impact of the patients’ perspective has been the recommendation by OMERACT to include the measurement of fatigue in the core set of measures for clinical trials in RA. Subsequently, work has focused both on other important aspects such as measurement of sleep quality and more method-oriented topics such as how to rigorously develop new patient-reported outcome measures. Groups such as EULAR or the U.S. Food and Drug Administration have made the involvement of the patient in the development of patient-reported outcomes an essential feature for new scales (e.g., as seen in the development of RAID [Rheumatoid Arthritis Impact of Disease]). Involving patients in research has several challenges that include practical issues such as overcoming access, communication barriers, establishment of a new professional relationship while continuing in a doctor-patient relationship, confidentiality issues, and training in science methodology and nomenclature. It is expected that patient or consumer involvement will become a standard feature in clinical and outcomes research and that methods of engagement will continue to evolve.
Defining key concepts or outcomes may also be facilitated through use of conceptual frameworks. Conceptual frameworks offer definitions of their key concepts and discuss how a construct relates to other variables in their model. One increasingly popular framework is the International Classification of Functioning (ICF) endorsed by the World Health Organization in 2001. The ICF framework (Figure 33-1) describes three main concepts: impairments (symptoms, structural limitations); activity limitations (difficulties while performing tasks); and participation restrictions (social role participation). Also important are environmental factors (e.g., job demands, environmental barriers, weather) and personal factors (predispositions, coping strategies). Other frameworks include the Verbrugge disablement process, which has slightly different concepts along its main pathway, and the Wilson and Cleary framework. Both of these frameworks define a main pathway from cellular findings to a broad level of disability or quality of life. Personal and environmental factors are also present in these models, but more as effect modifiers rather than as an essential part of the concept. Importantly, they also differ from the ICF in terms of the definition of “disability,” for example, reinforcing the need to be explicit about not only the concept but also the framework from which one’s definition comes. An alternative definition for physical functioning at the level of disability is described by Verbrugge. Similarly, when measuring pain, is intensity more important than frequency to patients? What about the degree to which pain interferes with daily activities? Conceptual frameworks define the realm of outcomes that should be considered and the hypothetical relationships between them. They form the basis for understanding our observations, testing hypotheses, or planning and executing an analysis. Choosing a framework that helps one to define and think about the concept one wants to measure will be helpful down the road when trying to understand the findings. Working across different frameworks or trying to fit an instrument into another framework is challenging because different instruments may vary in how they define certain aspects of health or disability. However, shifts to a new framework can prompt fruitful rethinking of concepts and how they relate to each other.

Thinking about a concept and defining it carefully and explicitly should precede reviewing any instruments or core sets. The instrument should meet the need—not the reverse.

Why Measure?
Clarity about an instrument’s intended purpose will help ensure the right one is selected. Kirshner and Guyatt describe three purposes: descriptive (measure a concept at one point in time such as the burden of illness), predictive (provide information about the future such as the HAQ predicting mortality in arthritis), and evaluative (measure change over time such as the benefit or harm from treatment). Each purpose requires evidence of certain measurement properties of the candidate instrument. In this chapter we focus on the purposes relevant to health outcome assessment: describing an end point state in a trial at one point in time (Kirshner’s descriptive purpose) and evaluating the amount of change experienced over time (Kirshner’s evaluative purpose). The purpose dictates the type of evidence to focus on when making a decision about a given instrument.

Who Comprises the Target Population?
The target population is critical but often overlooked. A given instrument may, for example, work well in severe OA of the hip but not be sensitive to the early symptoms of the disease. It is equally important to consider if one wants to measure for an individual patient or describe a group of patients as a whole, for example, in a clinical trial. The former demands much higher levels of measurement properties such as reliability coefficients greater than 0.90 as opposed to a minimum of 0.75 to 0.80 for group descriptions.

Decisions made based on these three points will help define the measurement need and provide a better foundation for critiquing candidate instruments.

**DECISION-MAKING INSTRUMENT FOR SELECTING THE OUTCOME THAT CAN MEET THE MEASUREMENT NEED**

The selection of an outcome measure depends entirely on a clear understanding of measurement need. All too often a commonly used instrument is selected rather than looking for one that matches the concept, population, and purpose. Once the measurement need is defined, the users should begin on a decision-making process that helps guide the final decision. Many guidelines that offer more detail than can be provided here, particularly for the acceptable levels of reliability and validity, are available. What we describe here is a decision-making process that can be used to assess if a given instrument fits with the articulated measurement need. This process, depicted in Figure 33-2, builds on the work of Law and the OMERACT filter. It also highlights key concepts in each area from the published guidelines.

This decision-making process has three key features that deserve mention. Once an instrument has been selected for consideration, the process begins with a clear statement of the measurement need. This is an unavoidable step that comes before any evaluation of candidate instruments. Second, there are many things to be done using what has

---

**Figure 33-1** The International Classification of Functioning conceptual framework showing the hypothesized relationships between domains of impairment, activity limitations and participation restrictions, and the direct influence of environmental and personal factors on these domains.
Figure 33-2  Algorithm showing decision-making process for the fit of a candidate measure with a target measurement need. The first three (blue boxes) can be completed by appraisal of the instrument hopefully with input from respondents. The last three (pink boxes) require data. Many instruments are weeded out as a poor fit in steps 1, 2, and 3—things that cannot be corrected. The ‘do-end’ loop denotes stages at which one can pause to create evidence if it is missing, and the instrument does not necessarily have to be abandoned.

Legend
- Boxes marked with “do it loop” are those where one can create the evidence and continue
- Blue boxes = appraisal with or without help from patients, less emphasis on statistics
- Pink boxes = data analysis is required to pass these stages

Step 1: Is It a Good Match with the Need?
Think about the concept and then decide based on the description of the candidate instrument and the nature of the items whether there is a match between the instrument’s concept and the measurement need (concept, population, purpose). An operational definition of the target concept, the applicable populations (specific patient group or general population), and intended purpose should be articulated by the developer and match the current need. If it is not there or if it is not a good match, start with another candidate instrument because this one will not work.

Step 2: Is It Feasible to Use?
Feasibility covers the practical aspects of using this scale in the intended setting. Does it take too much time? Are the licensing costs too high? Does it require special
As the decision-making process unfolds, one will see much emphasis on the truth part of the OMERACT filter. This level of “truth” describes content validity. Does it appear that the candidate instrument is covering the domains of the construct well, and do the items align in this content? This, like step 2, is usually done by the clinicians or researchers, but patients can also provide valuable insight into ensuring the content is comprehensive.

Content validity appraises the items and domains of a scale, as well as whether the authors have covered the breadth and depth of the concept. In other words, are all the important areas covered, and is there enough depth to capture the range of experience of the patients? Face validity is an appraisal of the general direction of the scale; will it hit the target? Are the response options organized in a logical direction for high and low levels of this attribute? Does the scoring make sense?

**Step 3: Does the Instrument Look Like It Has the Right Content in Order to Measure What It Is Intended to Measure (Truth 1)?**

Over the course of the decision-making process, one will make much emphasis on the truth part of the OMERACT filter. This level of “truth” describes content validity. Does it appear that the candidate instrument is covering the domains of the construct well, and do the items align in this content? This, like step 2, is usually done by the clinicians or researchers, but patients can also provide valuable insight into ensuring the content is comprehensive.

Content validity appraises the items and domains of a scale, as well as whether the authors have covered the breadth and depth of the concept. In other words, are all the important areas covered, and is there enough depth to capture the range of experience of the patients? Face validity is an appraisal of the general direction of the scale; will it hit the target? Are the response options organized in a logical direction for high and low levels of this attribute? Does the scoring make sense?

**Step 4: Do the Numeric Scores Make Sense? Are These Scores Behaving in Ways That a “Good” Measure of This Construct Would Behave (Truth 2)?**

Having convinced oneself of the content and practical feasibility of this scale, the next step begins the more data-intensive evaluation. One can abandon the item-level critique and begin to explore data to see if the numeric scores arising from the instrument make sense and behave in the same way an excellent instrument of this construct would be expected to behave. Confidence with construct validity should be established regardless of the eventual purpose (descriptive or evaluative). Sometimes it is deemphasized in evaluative instruments in favor of a more detailed evaluation of change and responsiveness; however, others (including the authors) would stress the importance of making sure one knows what is being measured before focusing solely on change scores.

Construct validity is generally measured by comparisons with other similar scales or related constructs (i.e., high and low levels of pain and function). Theoretical situations are established before analysis, the direction and magnitude of the expected relationship are declared, and then the relationship is tested. Comparisons should also be made between groups known to differ (high vs. low severity) or with scales where no relationship is found. Again, this is based on an a priori theory checking to see if the candidate instrument behaves according to the theory. These comparisons add to the evidence that the instrument is measuring what it is supposed to measure. If the evidence is not available or not available for the intended population, one has the choice of abandoning the instrument or conducting a study to create that evidence and then continuing to advance.

Construct validity can also be assessed looking at the structure of a scale. If a scale (or subscale) is set up to measure one construct, all the items should “load” onto that one factor. That is, they should be highly correlated with each other enough that together they seem to belong to an underlying trait like health, pain, or anxiety. We assess this structural validity through approaches such as factor analysis (if the instrument was designed to have multiple items aligning with one underlying trait) or item response theory, to see if the items designed to capture consecutive levels of the trait along the continuum are doing a satisfactory job.

At this stage it is also necessary to consider the precision of the measurement. The observed score produced by an instrument should be close to the true score with low error. This is estimated by the internal consistency of a multi-item scale or questionnaire using Cronbach alpha coefficients or Kuder Richardson 20 if the scale is dichotomous (yes/no). Internal consistency is a feature of a scale with many items measuring the same thing. The responses should be similar across items within the instrument. It is not a feature of a scale containing weighted sums of different attributes such as disease activity measures. The internal consistency reliability can be converted back into the scale score by calculating the precision limits (using 95% limits, the true score is somewhere within $1.96 \times \sigma \sqrt{N}$, where $\sigma$ = internal consistency and $\sigma$ = standard deviation). This identifies the range within which the true score for an individual will reside.

If more than one person will be gathering the data, inter-rater/interobserver reliability should be measured and quantified with an intraclass correlation coefficient (ICC) for continuous measures or a weighted Kappa for ordered categories. There are different types of ICC depending on the model used for the variance estimates. The type of ICC should always be named. The ICC and weighted Kappa measure the comparability of actual numeric scores and are preferred over correlation coefficients that look only for trends and not a direct match in number values. Cutoffs are always challenging, but in general reliability should be at a minimum 0.75 for group-level analyses (where analysis is always done for a group of patients, as in the mean score comparisons in a clinical trial). For a description of an individual patient, ICC should be 0.90 to 0.95.

**Step 5: Can This Instrument Evaluate Change over Time in a Group of Patients?**

After getting a general sense of the construct validity of the instrument, we sometimes want to know if this outcome can measure change over time. This is only important if one is
going to measure change. In the original OMERACT filter, this was the “discrimination” component of the information need. Can the instrument discriminate finely enough to detect the change one needs to be able to detect? If the goal is to describe an outcome such as the level of pain after a treatment, there may be enough information at step 4, or the researcher may want to check its ability to describe response at an individual level (step 6). Interobserver reliability is important if more than one measurer will be involved in the study. The hallmark of the ability to measure change in a group is twofold: First, do the scores remain the same when the target concept has not changed over time (test-retest reliability), and second, when the concept changes, does the score on the instrument change as well (responsiveness or sensitivity to change)?

**Test-Retest Reliability**

Test-retest reliability requires two administrations of the instrument over a time when no change in the target concept has occurred. As a reader of reliability studies, one should feel convinced that no change in the target (e.g., pain, function, disease activity) would have occurred in these patients in this situation. Often, people conducting studies of test-retest reliability will set up a clinical situation in which no change should have occurred or they use an external anchor (e.g., is your pain the same as last time?) to find patients who have not changed. Like interobserver reliability, the ICC is the preferred statistic for continuous scores and weighted Kappa, its equivalent, for categorical scores. The cutoffs are the same, and a coefficient can be converted into a “minimal detectable change” of \(1.96 \times s(2[1-r])^{1/2}\), where \(s = \text{standard deviation and } r = \text{test-retest reliability (ICC)}\). Ninety-five percent of people who are stable will have change scores less than this value; hence a change greater than this is not likely to occur in a stable patient, only in a changing one. It thus becomes a lower boundary of meaningful change. Anything below that boundary could be day-to-day fluctuations in scores.

Responsiveness is the accurate detection of change when it has occurred. It is best thought of as longitudinal construct validity. Like construct validity, it depends on an a priori theoretic relationship in which the attribute is changing over time (e.g., change in pain over time). Researchers all too often focus on the amount of change picked up rather than the match of the change in the instrument’s scores with the type or amount of change that has actually occurred and was expected in that testing situation. A large change is not useful if a small one was expected—it only suggests error. The amount of change expected in a study of responsiveness should be carefully described and should be a clear match with the intended application (i.e., measurement need). If the goal is to detect change in a clinical trial, then it is important to assess the instrument’s ability to detect the difference in change between treatment and control groups. If the goal is to detect change in a cohort, it might be more useful to examine change in a single group, perhaps in a treatment of known efficacy (e.g., hip replacement) or in those people who rated themselves as improved on an external anchor (e.g., global index of change). Responsiveness is summarized with statistics of signal (change) over noise (error) such as the standardized response mean (mean change/standard deviation of change), t-statistic (mean change/standard error), or effect size (mean change over standard deviation of baseline). Each can be adapted to quantify the relative change between treatment and control groups. Deyo and Centor also described the correlational approach (correlate change and another indicator of change) and the receiver operator curve approach (various change scores against external “gold standard” that the person changes), where the area under the curve serves as a summary statistic. The numeric summaries of responsiveness such as effect sizes or areas under the curve should correspond to the type of change expected (a priori theory). A large effect size or area under the curve does not mean an instrument is “responsive.” It should correspond with the change anticipated in the study, small or large. Comparisons of the effect sizes are helpful if different instruments are being compared in the same study as done by Buchbinder and Verhoeven and their colleagues, who both focused on responsiveness in early RA. Responsiveness is a highly contextualized property, and the same instrument may not be responsive in another situation (early vs. late disease, OA vs. RA).

**Step 6: Can It Define an Important Response for an Individual Patient?**

The final step, often deemed the most elusive, is the interpretability of the scores. Responder analyses that quantify outcome as the “proportion” improved, recovered, or who have responded to therapy need to have the ability to be interpreted at an individual level in order to classify an individual as responsive or not.

**Benchmarking States**

At the end of a trial, a patient’s pain is at 2/10 on a pain scale. Is this a good outcome? The meaning of different scores on an outcome assessment is used for classifying people both at the beginning of a trial and at the end (final state). To establish the meaning of the specific values of an outcome scene, comparisons are made to other known health states such as severity indices, ability to work, or self-rating as mild. Gradually, enough trends might be seen across different scenarios to gain confidence in the meaning of “good” or “mild.” In rheumatology, we see the emergence of low or minimal disease activity states (LDAS; MDA), patient acceptable symptom states (PASS), or remission criteria with the DAS28 thresholds below which people are considered to be in an acceptable state (defined as either tolerable symptoms or disease activity that does not require medication changes). At this point these thresholds are being established, and similar to change thresholds, one may find variability in the values that will need to be sorted out with methodological work and application in clinical practice.

**Changes in State**

The second type of interpretability concerns change scores. The American College of Rheumatology Response Criteria.

The American College of Rheumatology took the core set
measures and determined that if one observed an X% change in joint count and swollen joint count and in at least three other areas—erythrocyte sedimentation rate or C-reactive protein, physician global, patient global, pain, or physical disability—one had a clinical response, and the individual would be classified as a responder. The percent is usually 20%, but 50% and 70% have also been considered. The ACR20 is widely used, defining responses across a wide variety of domains, and discriminates well in clinical trials; however, it is currently being revalidated due to the changing nature of RA and its care.

Minimal Clinically Important Differences/Improvements. Defining the threshold of change above which a person has had an “important” shift in outcome is what Kirwan has described as the “elusive crock of gold at the end of the rainbow.” Nevertheless, important advances have been made. In 2000, Wells and colleagues described nine different methods for deriving minimal clinically important differences (MCIDs) from the literature. Some use distributional cutoffs (e.g., 0.5 standard deviation, effect size of 0.2 or 0.5), which have been criticized as lacking any meaningful anchor. Others depend on an external anchor, indicating that important improvement has occurred, but are sometimes challenged by their dependence on that anchor, as well as on the perspective of the person who determines the response (patient, doctor, third-party payer). MCIDs have been shown repeatedly to vary with baseline state and with improvement versus deterioration. Tubach and colleagues have changed the term to minimal clinically important improvement and only looks at improvement. MCIDs vary depending on the context of measurement. Plan on working with a range of values, make sure the measurement situation is similar to your own (e.g., severity, timing, type of intervention), and build confidence with congruence in MCIDs from across methods if you can achieve that.

Combined Approaches: Change and State

An attractive, though often overlooked, option is combining the last two. In 1996 EULAR defined clinical response as a change in DAS28 score of more than 1.2 (change), plus a final DAS28 score of less than 2.4 (final state). Jacobson and colleagues did the same in defining response to psychotherapy; change greater than error was used (minimal detectable change mentioned earlier) plus a final “normal” state. Studies from the patient’s perspective have often reflected the same thing. Treatment needs to induce a change, but perhaps it also needs to land people in a healthy state such as feeling better or being able to do their daily tasks.

The approaches described earlier focus on interpretation at the level of the individual, perhaps for use in clinical practice or in a response-type analysis of a clinical trial or economic appraisal (% responder). Verhoeven and colleagues have shown that the same instrument may not perform equally well in a responder-type analysis and for a group-level change.

At each stage of this appraisal, there is an element of judgment. Perfect evidence across all stages will never be likely. The user will need to assess the potential risk of accepting less than ideal evidence or abandoning the scale. They may also, however, create the evidence by doing it themselves. An instrument that makes it through this appraisal is likely a good fit with the measurement need. Anything short of that could lead to error and be prone to misinterpretation. By working from left to right on the figure, instruments that are not targeting the right concept or are impractical to use in the intended setting can be quickly eliminated before extensively reviewing the literature for the measurement properties.

AREAS OF GROWTH IN HEALTH OUTCOME ASSESSMENT

Item Response Theory

Traditional instruments have often been developed using either classical test theory (psychometrics) or a regression on key clinical findings (e.g., prediction rules, diagnostic or disease activity scores). However, other methods take another approach to instrument development. For example, instrument development may employ a modern measurement theory such as item response theory (IRT). Using IRT, one might choose a different set of items because this approach favors items representing the whole range of a concept (e.g., pain, disability). In contrast, classical tests tend to favor items that are highly correlated to one another, sometimes not favoring items at the extremes. Instruments developed or scored using IRT will still need to demonstrate validity and/or responsiveness, as well as reliability, but an IRT person score will be used rather than a summed score based on observed item responses. The advantage of IRT is that items are ordered from those representing low levels of the attribute to those representing high levels. Computer-based scoring can then skip items that do not need to be asked based on previous responses. This streamlined testing can allow a precise estimation of the level of that attribute using fewer items.

This type of streamlined scoring is attractive, but there are some limitations. Different patients are being assessed using different outcome questions, each targeting their level of health. It is a method that depends on technology which may not at this time be available in every setting. It may also be influenced by differential item functioning (dif), which means an item might change weight or order in certain subgroups of patients and necessitates more complex weights. An example of dif would be putting on a sweater or shirt over one’s head: It is a difficult task with shoulder pain but easy with a hand problem. This would require a different weight in each group.

The National Institutes of Health is currently funding “PROMIS” (Patient-Reported Outcomes Measurement Information System: www.nihpromis.org/) to develop a computer adaptive testing system (currently based on two-parameter graded response model IRT) for the more common chronic diseases including arthritis. Several instruments have been pooled into a large database and are being refined and rescaled at the time of writing. Well-known instruments such as the HAQ have also been used to allow for cross-calibration with the newer items. All findings and scoring algorithms will be reported on the PROMIS website.
Adaptation to an Ongoing Disease

In this chapter, we have focused on the measurement of health states and, indeed, their improvement or deterioration over time. However, people with chronic diseases will adapt to ongoing disease with behavioral strategies, or cognitive reframing of their situation.13 In some circles this is called adjustment,13 and in others it is called response shift.134 The challenge in health outcome assessment is to tell when a state is changing only because of adaptation and not the intervention. In many situations we try to induce adaptation, or cognitive reframing, and it can be constructive. It can, however, create a bias in measurement134 and a challenge to the health outcome assessor. A number of groups are researching how to incorporate adaptation into our health outcome assessments.

Changing Nature of Clinical Trials

Outcome measures provide critical yardsticks for clinical trials, yet there are many changes in the face of clinical trials. There are more pragmatic trials, comparative effectiveness studies, and nonpharmacologic trials that will demand more precise, sensitive measures of treatment benefits because they will by design have smaller relative effect (treatment vs. control arms). Further, researchers are hearing the need for different disease and outcome targets and must respond to the call for strong measures of coping, self-efficacy (including efficacy as a health consumer), and the impact of arthritis on a person’s life.41,44

CONCLUSION: ARE WE THERE YET?

There is considerable room for improvement in health outcome assessment in rheumatology, despite the work done to date. We have developed a battery of instruments, many of which are demonstrating the measurement properties described earlier and facing the challenge of a changing arthritis target (less severe, earlier disease), and we are considering several more measures for membership in core sets in order to capture a comprehensive view of the burden of arthritis. We are on the brink of deciding on the role to be played by item response theory and computer adaptive testing in widespread care settings. However, despite progress in assigning a numeric value to a complex health state, we now are in a healthy struggle with the back-translation—what does the numeric score mean in the real world of the patient-clinician decision making? It is not always a simple translation from questionnaire score to clinical meaning. Health outcome assessment is well advanced in arthritis care, and we should recognize the years of work and commitment of many professional and patient/consumer groups. Advances will continue in the use of technology, the breadth and depth of our outcomes, and the quality of measurement to keep pace with the needs of patients, clinicians, and researchers.

Acknowledgments

Dr. Tugwell holds a Canada Research Chair in Health Equity. The authors would like to thank Dr. Claire Bombardier, Dr. John Kirwan, Dr. Sarah Hewlett, Ms. Ellie Pinsker, Ms. Patricia Nedanovski, and the OMERACT executive for their help with this manuscript.

Selected References


Full references for this chapter can be found on www.expertconsult.com.
References


Biomarkers are anatomic, physiologic, biochemical, or molecular parameters associated with the presence and severity of specific diseases and are detectable by a variety of methods including physical examination, laboratory assays, and imaging. Here the focus is on biomarkers in senso stricto: markers that can be measured in patient samples such as blood, urine, synovial fluid, and synovial tissue. Such molecules often first appear in studies on disease mechanisms, their application as biomarkers being secondary. Realizing the importance and potential of biologic markers, the identification, characterization, and validation of novel biomarkers is often a primary objective in many current studies. Biomarkers are applied for various purposes including diagnosis, prognosis, monitoring of disease progression, selection of patient populations in clinical trials, assessment of the efficacy of treatment, or unraveling of the pathobiology of a disease in the clinical and in a preclinical setting. More recently, biomarker validation level-of-evidence schemes have been proposed to optimize efficient use of biomarkers in these diverse research areas.¹²

For osteoarthritis and rheumatoid arthritis (RA), the most important common feature is progressive destruction of articular tissues, resulting in impaired joint function and pain. Diagnosis is based on clinical symptoms and laboratory tests in combination with radiography, to visualize often irreversible degenerative and destructive changes in the joint. Radiologic evaluation of joints mainly images bone and is relatively insensitive: A follow-up period of at least 1 year is necessary to assess disease progression and effect of therapy. Magnetic resonance imaging (MRI) has the ability to visualize all joint tissues simultaneously; it is currently being optimized but has yet not reached its full potential due to its relatively high cost and still limited availability.⁴ Except for imaging modalities such as positron emission tomography scans, which use labeled tracers that may be used to image an ongoing biologic process, most imaging techniques provide a cumulative historical view of damage that has already occurred, rather than assessing the current rate of disease progression (Figure 34-1).

Alternative methods that detect changes in the joints in an early stage of the disease in a quantitative, reliable, and sensitive manner are necessary. The World Health Organization in its report Priority Medicine for Europe and the World states that biomarkers are essential for arthritis in general and osteoarthritis research in particular, and that the lack of adequate markers constitutes a major hurdle for osteoarthritis drug development.⁵ No osteoarthritis biomarker underwent the formal approval procedure of the U.S. Food and Drug Administration (FDA) to be allowed as qualifying end point in a phase III clinical trial.⁶ Molecular markers (i.e., markers that are used to monitor molecular events occurring during disease) are well suited for this purpose. A good marker is disease specific, reflects actual disease activity, is sensitive to change after therapy, and can predict disease outcome. Most likely, all of these requirements are not met by a single marker: Different (combinations of) markers would have to be applied for different purposes. The fact that imaging techniques often focus on one joint (e.g., knee or hip) or joint group (e.g., hands, lumbar spine), whereas urine-derived or blood-derived biomarkers are determined by all of the joints in the body, further underscores the importance of using imaging and biomarker approaches as complementary tools. The molecular markers that are described for joint diseases can be arbitrarily classified as follows:

1. Immunologic and inflammatory markers (including acute-phase proteins)
2. Markers that reflect extracellular matrix remodeling (Table 34-1)
3. Markers present in synovial tissue biopsy specimens
4. “Omics”-based biomarkers (e.g., genomics, transcriptomics, proteomics, metabolomics)
5. Genetic markers that have been shown to provide important information on risk factors for development and progression of osteoarthritis and RA; these are reviewed elsewhere and are not included in this chapter.

### KEY POINTS

- Uniform definitions of disease are essential for biomarker validation and comparison among studies.
- The validation process of a biomarker depends on the specific purpose of its use.
- For tissue homeostasis, the balance between anabolic and catabolic process is important; extracellular matrix remodeling biomarkers should reflect these different processes.
- Understanding tissue source, formation, and clearance of a biomarker is important for correctly interpreting biomarker data.
- Analyses of serial synovial biopsy specimens can potentially be used as a screening method to test new drug candidates requiring relatively small numbers of patients.
- Panels of biomarkers or biomarker profiles are potentially more powerful than single biomarkers.
- For biomarker identification, validation, and application, use of patient subclassification/stratification is essential to increase homogeneity in study and treatment groups.

**Table 34-1**

| Genetic markers that have been shown to provide important information on risk factors for development and progression of osteoarthritis and RA; these are reviewed elsewhere and are not included in this chapter. |
INFLAMMATORY BIOMARKERS AND SIGNAL MOLECULES

Extensively studied markers in inflammatory arthritis that are routinely used in clinical practice include acute-phase proteins such as C-reactive protein (CRP) and measurements such as the erythrocyte sedimentation rate (ESR), both of which provide information about the systemic inflammatory process. Although such markers of inflammation are neither disease specific nor tissue specific, they may reflect disease activity, reflect the effects of immunosuppressive and immunomodulatory therapies, and, to a certain extent, predict disease progression in RA. A more disease-specific marker for RA is the presence of anticitrullinated protein/peptide antibodies. It predicts with high probability the development of RA in patients with no clinical symptoms, distinguishes between RA and other rheumatic diseases, and is a valuable tool for prognostic prediction of joint destruction. Although synovial inflammation is often regarded as a secondary process in osteoarthritis, cytokines and other signal molecules have also been proposed as markers in this disease. In clinical studies, often highly sensitive analysis of CRP is included because it is associated with osteoarthritis of the knee and hip, but validation of this biomarker to monitor treatment efficacy is hampered by the lack of effective treatments (as is the difficulty with all osteoarthritis biomarkers).

Cytokines and Chemokines

RA patients often show elevated levels of a variety of cytokines, chemokines, and their receptors (for details see Chapter 26). Many of these molecules are subsequently used as disease biomarkers or studied as targets for intervention, analogous to the experience with tumor necrosis factor (TNF). This cytokine has been shown to play a pivotal role in arthritis and has emerged as a major therapeutic target. TNF inhibitors (neutralizing antibodies or soluble receptors) strongly reduce clinical symptoms, joint inflammation, and biomarkers of inflammation and bone destruction in RA patients. Neither the serum levels of the TNF receptors nor those of interleukin (IL)-10 were reduced, however, in response to treatment of RA patients with methotrexate and anti-TNF treatment, despite a clear improvement in clinical disease parameters and a reduction of CRP and ESR. In contrast, in synovial tissue, treatment with these drugs did reduce the local TNF level, indicating that for biomarker analysis it is important to select the right compartment. The observation that anti-TNF treatment may block joint destruction in patients who do not show a clinical response (i.e., ACR20 nonresponders) suggests that, similar to osteoarthritis, in RA uncoupling of inflammation and joint destruction may occur, which may affect the selection of biomarkers to monitor treatment effects. High levels of soluble receptors of TNF are associated with reduced physical function and worse radiologic knee osteoarthritis.

IL-6 is another cytokine that has gained interest in the past few years as a therapeutic target in RA. Anti-IL-6 receptor antibody therapy in combination with methotrexate results in decreased RA disease activity, improved function, and reduced levels of joint destruction biomarkers. Synovial fluid IL-6 levels in RA patients are correlated with infiltration of inflammatory cells in the synovial membrane and with radiographic joint destruction. In a clinical trial of cytokine blockade, baseline serum IL-6 levels predicted radiographic progression, but not the ACR response, again indicating that inflammation and joint destruction are not necessarily coupled in RA. Systemic circulating IL-6 levels are well correlated with CRP levels, and therefore the added value of this biomarker seems to be limited for monitoring RA disease activity and response to treatment.

Adipokines

Since the discovery of leptin, white adipose tissue is no longer considered as only a fat storage tissue but also as an active contributor to a wide variety of physiologic and pathologic processes including RA and osteoarthritis. It plays a critical role as an endocrine organ, producing a number of active peptides, the so-called adipokines.

The role of leptin in the pathogenesis of rheumatoid arthritis is still under debate. Increased systemic levels of leptin in RA patients compared with controls were observed, and a positive correlation among serum leptin concentration and disease activity, ESR, and the number of tender joints was demonstrated. In contrast, other studies demonstrated a link between leptin and body mass index and not with disease activity. In line with this observation is...
Table 34-1 Molecular Markers That Reflect Extra Cellular Matrix Remodeling

<table>
<thead>
<tr>
<th>Marker</th>
<th>Joint Tissue</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Synthesis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Collagen Type I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-propeptide (PINP)</td>
<td>Bone, soft tissues</td>
<td>58</td>
</tr>
<tr>
<td>C-propeptide (PICP)</td>
<td>Bone, soft tissues</td>
<td>58</td>
</tr>
<tr>
<td><strong>Collagen Type II</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-propeptide (PINP; PIILNP)</td>
<td>Cartilage</td>
<td>75, 102, 103</td>
</tr>
<tr>
<td>C-propeptide (PIILNP; chondrocalcin)</td>
<td>Cartilage</td>
<td>95, 100, 101</td>
</tr>
<tr>
<td><strong>Collagen Type III</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-propeptide (PIILNP)</td>
<td>Soft tissues</td>
<td>74, 104-106</td>
</tr>
<tr>
<td><strong>Proteoglycans and Glicosaminoglycans</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chondroitin sulfate (epitopes 846, 3-B-3, 7-D-4)</td>
<td>Cartilage</td>
<td>88, 95, 96, 112-116</td>
</tr>
<tr>
<td><strong>Miscellaneous</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone-specific alanine phosphatase</td>
<td>Bone</td>
<td>58</td>
</tr>
<tr>
<td>Osteocalcin</td>
<td>Bone</td>
<td>59</td>
</tr>
<tr>
<td>YKL-40 (CYLK-40, gp-39, chondrex)</td>
<td>Cartilage</td>
<td>58, 59</td>
</tr>
<tr>
<td>Hyaluronan</td>
<td>Cartilage, synovium</td>
<td>77, 116, 122-127</td>
</tr>
<tr>
<td><strong>Degradation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Collagen Type I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-linked N-telopeptide (NTx)</td>
<td>Bone</td>
<td>58, 59, 63, 64, 68-71</td>
</tr>
<tr>
<td>Cross-linked C-telopeptide (CTX)</td>
<td>Bone</td>
<td>59, 64, 67-71</td>
</tr>
<tr>
<td>Cross-linked C-telopeptide (ICTP)</td>
<td>Soft tissues</td>
<td>64, 66</td>
</tr>
<tr>
<td>Collagenase cleavage neoepitope (C1, 2C)</td>
<td>Bone, soft tissues</td>
<td>89, 90</td>
</tr>
<tr>
<td><strong>Collagen Type II</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-linked C-telopeptide (CTX-II; 2B4 epitope)</td>
<td>Cartilage</td>
<td>59, 73-83</td>
</tr>
<tr>
<td>Collagenase cleavage neoepitope (9A4, C2, C1, 2C)</td>
<td>Cartilage</td>
<td>86-92</td>
</tr>
<tr>
<td><strong>Pyridinolines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydroxylysylpyridinoline (HP, PYR)</td>
<td>Bone, cartilage</td>
<td>94-96</td>
</tr>
<tr>
<td>Lysylpyridinoline (LP, D-PYR, DPD)</td>
<td>Bone</td>
<td>95, 96</td>
</tr>
<tr>
<td>Glucosylgalactosyl-hydroxypropyridinoline (GGHP)</td>
<td>Synovium</td>
<td>75, 97</td>
</tr>
<tr>
<td><strong>Proteoglycans and Glicosaminoglycans</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggrecan core protein (fragments)</td>
<td>Cartilage</td>
<td>58, 118-121</td>
</tr>
<tr>
<td>Keratan sulfate (epitope 5-D-4, AN9P1)</td>
<td>Cartilage</td>
<td>58, 88, 95, 96, 107-116</td>
</tr>
<tr>
<td>Chondroitin sulfate</td>
<td>Cartilage</td>
<td>58</td>
</tr>
<tr>
<td><strong>Miscellaneous</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrix metalloproteinases</td>
<td>Cartilage, synovium</td>
<td>27, 30, 75, 138, 140-152</td>
</tr>
<tr>
<td>Aggrecanases</td>
<td>Cartilage</td>
<td>117, 153-159</td>
</tr>
<tr>
<td>Cartilage oligomeric matrix protein (COMP)</td>
<td>Cartilage, possibly synovium</td>
<td>74, 128-139</td>
</tr>
<tr>
<td>Bone sialoprotein (BSP)</td>
<td>Bone</td>
<td>57, 96</td>
</tr>
</tbody>
</table>

The fact that anti-TNF therapy did not reduce plasma leptin concentrations while inflammation decreased in the patients.41

Adiponectin is another well-known adipokine secreted by adipose tissue, and serum concentrations are inversely correlated with body mass index. In contrast to its protective role in obesity-linked diseases, adiponectin seems to fulfill a proinflammatory role in RA.42 Serum and synovial fluid adiponectin levels are significantly higher in RA patients than in healthy controls, and serum levels correlate with the severity of joint destruction.43 However, inconsistent data on the effect of anti-TNF therapy on circulating adiponectin levels exist, showing an increased level of adiponectin after therapy or no effect at all.44-48 As such, the value of circulating adiponectin for assessing anti-inflammatory therapy seems to be limited.

Other adipokines such as visfatin and resistin have been implicated in the pathogenesis of RA as well.50 Visfatin expression is not limited to adipose tissue, but synovial fibroblasts from RA patients also exhibit visfatin expression, especially in the synovial lining and at sites of invasion.50 A positive correlation between visfatin serum levels and clinical disease activity has been demonstrated; however, in another clinical study such a relation was not observed.50-51 Resistin expression is found in the sublining layer of the synovium and is more intensive in RA than in osteoarthritis.
patients. Increased circulating resistin levels in RA patients correlate with both CRP and disease activity.22,33

Although a proinflammatory role is suggested for the previously mentioned adipokines and is sustained by in vitro data, clinical data that reveal these adipokines as therapeutic targets or as biomarkers in RA are sparse and contradictory. Further research into the relation of serum/synovial fluid adipokine level and inflammatory markers such as CRP or clinical disease activity such as DAS28 is necessary to evaluate the value of these adipokines as prognostic, diagnostic, or therapeutic biomarkers in RA.

In osteoarthritis patients, serum nitrate and nitrite (reflecting nitric oxide levels) are increased compared with healthy controls, but not as much as in inflammatory conditions such as RA.34 In addition, using a genomics approach, chondrocytes from osteoarthritic cartilage have been shown to upregulate the transcription of a variety of inflammatory genes.35 Synovial tissue from osteoarthritis patients also shows signs of hyperplasia and inflammation.36 Although osteoarthritis is traditionally viewed as a noninflammatory arthropathy, an osteoarthritic joint could be considered a mildly inflamed organ.

The added value of inflammation biomarkers measured in peripheral blood in joint diseases is still undecided. Although the levels of many inflammatory molecules may change in various (stages of) diseases, such change does not mean that the molecule is directly involved in the disease process or is sensitive to change after intervention. Serum cytokine levels often do not predict clinical response. As such, it is crucial to understand the complex biologic networks and the mode of action of treatments to be able to select the right (relevant) molecules to use as biomarkers. In view of this, the ultimate application of inflammation biomarkers in osteoarthritis studies is likely to be different from RA studies.

EXTRACELLULAR MATRIX REMODELING BIOMARKERS

Because of the aforementioned issues, the focus of many biomarker studies is on markers that mirror disease-related changes in the joints, especially markers for remodeling (i.e., degradation and synthesis) of the extracellular matrix.52 Concomitant changes may occur in all joint tissues (cartilage, bone, and synovium), and for a comprehensive assessment of these changes molecular markers are necessary for each of these tissues.58,59 For the markers derived from these different tissues, there can be a substantial difference in which biologic fluid biomarker levels are determined.

The concentration of markers in body fluids reflects not only the dynamics of the disease but also the rate of clearance and the amount of remaining tissue. Cartilage-derived markers diffuse out of the tissue and enter the synovial fluid, which may vary in volume depending on the severity of ongoing inflammation. Synovial fluid urea levels may be used to correct for this “dilution.”60 Clearance from the synovial fluid predominantly occurs via lymphatic drainage, and partial degradation may occur in the lymph nodes, depending on the marker studied. Synovial clearance may depend on the severity of the ongoing synovitis and be determined by the permeability of the synovial membrane microvasculature.61 When the marker enters the systemic circulation, dilution occurs (e.g., 5 mL of synovial fluid vs. approximately 5 L of blood), and marker levels are confounded by molecules from other affected or nonaffected joints or even from nonarticular cartilage.

Bone-derived and synovium-derived markers may also enter the circulation directly. When the marker enters the systemic circulation, it is diluted, mixed with markers derived from other joints or tissues, and potentially metabolized in the liver and kidneys. Excretion in the urine depends on the marker and the previously mentioned metabolic processes. In the discussion that follows, examples are given from studies of osteoarthritis and RA to illustrate their use in assessing disease-related tissue remodeling.

Collagen Markers

The main constituent of the extracellular matrix of connective tissues is collagen, which plays an essential role in the maintenance of tissue shape, strength, and integrity. Fibrillar collagen types I, II, and III are synthesized as propeptide-containing α chains that are post-translationally modified by hydroxylation of lysyl and prolyl residues and by glycosylation of hydroxylysyl residues. Triple helical collagen molecules are secreted from the cell, the propeptides are cleaved off, and fibrils that are stabilized by intermolecular cross-links are formed. This unique sequence of events provides a variety of tools that are used to study collagen synthesis (especially propeptides) and degradation (especially cross-links) in rheumatic diseases (Figure 34-2).52

In bone, type I collagen constitutes 90% of the organic matrix and markers of bone collagen turnover have proven valuable in monitoring diseases such as osteoporosis.65 For some bone markers such as NTx (bone degradation [see later]) and osteocalcin (bone formation), assays have been approved by the FDA to monitor efficacy of antiresorptive therapies or bone formation, sometimes even as a point-of-care test. For these bone metabolites, there is a significant menopausal effect that requires properly matched control groups and careful interpretation of the data.

Several assays have been used to assess collagen type I degradation in RA and osteoarthritis. The cross-linked C-terminal telopeptide (ICTP) is released by matrix metalloproteinase (MMP) cleavage of type I collagen, and its levels reflect MMP-mediated soft tissue degradation.64 The NTx/CTX-1 assay and ICTP assay detect type I collagen degradation, but by different proteases. Cathepsin K–mediated osteoclastic bone resorption destroys ICTP antigenicity.64 Slightly elevated serum ICTP levels are found in RA compared with controls and are associated with disease activity measured by ESR, CRP levels, and swollen joint counts.65 In osteoarthritis, fourfold increased ICTP levels have been detected in patients with rapid progressive hip osteoarthritis compared with patients with slowly progressive disease.66

When the C-terminal or the N-terminal telopeptide of type I collagen (CTX or NTx) is released from bone degraded by cathepsin K, an epitope that is different from the MMP-mediated ICTP epitope is generated. In osteoarthritis patients, serum and urinary CTx levels are decreased compared with controls, which suggests decreased bone remodeling in osteoarthritis.59,67 In RA, urinary NTx and CTx...
levels are increased compared with healthy controls and are sensitive to change after treatment.68–70 Although on initial examination CTx and ICTP seem to provide identical information, closer inspection reveals that these two markers, although based on the same principle of detecting type I collagen telopeptides, may provide valuable complementary information. CTx and NTx levels are low in patients with pyknodysostosis, which is caused by a deficient activity of cathepsin K, whereas ICTP levels are elevated in this condition. It has also been shown that in postmenopausal women, anti–bone resorption therapy by hormone replacement reduced serum CTx levels, whereas ICTP levels did not change.71 In cartilage, type II collagen constitutes 80% of the dry weight of the tissue. Damage to the collagen network (collagen degradation and subsequent denaturation) is one of the first features of osteoarthritis and contributes significantly to the decreased mechanical properties of osteoarthritic cartilage.72 Using monoclonal antibodies, release of cross-linked type II collagen telopeptides (C-telopeptide, CTx-II) was shown to reflect cartilage degradation with high tissue specificity.73 Urinary CTx-II levels were significantly increased in RA and osteoarthritis patients compared with healthy controls, although the ranges overlapped considerably.73–75 In a population-based study, subjects with a CTx-II level in the highest quartile had a 4.2-fold increased risk of having radiographic osteoarthritis of the knee and hip (compared with subjects in the lowest quartile) and a 6-fold (knee) or 8.4-fold (hip) increased risk for progression of osteoarthritis.76 In osteoarthritis patients, CTx-II levels correlated with radiologic joint space narrowing and joint surface area but did not correlate with WOMAC indices of clinical status in these osteoarthritis patients.59 In patients with hip osteoarthritis, urinary CTx-II levels greater than 346 ng/mmol creatinine were associated with a twofold increase in radiographic disease progression compared with patients whose levels were less than 346 ng/mmol creatinine.77 Treatment of these patients with a candidate antiosteoarthritis drug (diacerein) seemed to modulate the CTx-II levels consistent with the effects on disease progression.77,78 In a study of patients with radiographic osteoarthritis in multiple joints, there was a significant association between the total

Figure 34-2  Collagen-based markers for joint destruction. Collagen is synthesized as propeptide-containing α chains that are post-translationally modified by hydroxylation of lysyl and prolyl residues and by glycosylation of hydroxylysyl residues. These modifications cease when three α chains entwine to form a collagen triple helix. Triple helical collagen molecules are secreted from the cell, and the propeptides are cleaved off extracellularly. Subsequently, collagen molecules spontaneously assemble into fibrils with quarter-staggered overlap of the individual triple helices. Finally, the fibrils are stabilized by formation of intermolecular pyridinoline cross-links. CTx, C-terminal telopeptide (formally: carboxy-terminal collagen crosslink); HP, hydroxylysylpyridinoline; HP/LP, ratio between HP and LP; ICTP, cross-linked carboxyterminal telopeptide of type I collagen; LP, lysylpyridinoline; NTx, N-terminal telopeptide (formally: amino-terminal collagen crosslink).
radiographic OA score and urinary CTx-II levels. Subsequent multivariate analysis showed that the joint site-specific ROA score at all joint sites except for spinal disk degeneration contributed independently to this association. In RA patients, increased baseline CTx-II levels were associated with progression of joint damage, which was independent of baseline damage, treatment, and disease activity. Other studies also reported that urinary CTx-II is a strong, independent predictor of progressive joint destruction in RA.

In patients with active RA, the decrease in urinary CTx-II levels that was observed 3 months after initiation of treatment predicted the radiographic disease progression over 5 years, suggesting that urinary CTx-II levels may be used as an early marker of treatment efficacy. Anti-TNF treatment of patients with active RA for more than 1 year resulted in a significant reduction of CTx-II levels in the progressive group but not in the nonprogressive group, although the levels were not significantly different from baseline. Additional data on the influence of intervention therapy on CTx-II levels are required to judge whether this biomarker is valuable for assessing efficacy of treatment to halt further joint destruction.

Apart from telopeptide fragments, neoepitopes resulting from the cleavage of type II collagen by collagenases (MMP-1, MMP-8, and MMP-13) have been used to monitor cartilage collagen damage. Urinary excretion of collagen fragments containing the C-terminal neoepitope (TIINE assay, using antibody 9A4) was increased in osteoarthritic patients compared with age-matched healthy controls. In synovial fluid, levels of a similar C-terminal neoepitope (C2C epitope, using the Co2-3/Clong antibody) were significantly higher in osteoarthritis than in RA. In RA patients, serum levels of these biomarkers (C2C and C1,2C, a similar epitope present in types I and II collagen) were associated with progression of radiographic joint damage. In a randomized, double-blind, placebo-controlled glucocorticoid discontinuation trial of 137 subjects with knee osteoarthritis, neither C2C level nor C1,2C level, or their ratio, was affected by the treatment.

Urinary levels of the 622-632 peptide of type II collagen (also known as HELIX-II) were increased in patients with primary knee osteoarthritis (281 ng/mmol creatinine) and RA patients (409 ng/mmol creatinine) compared with healthy controls (180 ng/mmol creatinine). However, this and other reports applying the HELIX-II assay have to be interpreted with care because due to an apparent database error in the sequence, the epitope used to generate the HELIX-II antibody does not occur in cartilage type II collagen. Consequently, based on collagen sequences and HELIX-II epitope properties, type III collagen is one of several candidate sources of the cross-reacting signal in body fluids, but not type II collagen. This example illustrates an intrinsic potential deficit of all enzyme-linked immunosorbent assays (ELISAs), whether in competitive or sandwich format. These assays measure all immune reactivity but do not characterize which peptide binds to the ELISA antibody. Results are therefore potentially confounded by the contribution of additional peptides containing that same epitope but derived from other proteins. Liquid chromatography–tandem mass spectrometry (LC–MS/MS) assays are ideal to quantify unique peptides from a mixture of peptides. Therefore immune affinity LC–MS/MS assays that capture relevant peptides using an antibody and quantify the peptide of interest using mass spectroscopy combine the best of both assays.

**Collagen Cross-Links**

Degradation of fibrillar collagen types I, II, and III results in the quantitative excretion of the cross-links hydroxylsylpyridinoline (HP), derived from bone and soft tissue including cartilage, and llysylpyridinoline (LP), derived from bone, in the urine. In RA patients who received combination therapy (sulfasalazine, methotrexate, and prednisolone in the COBRA trial), time-integrated urinary HP levels correlated with the progression of radiographic joint damage (Sharpe/van der Heijde score) and bone mineral density. In osteoarthritic patients, urinary HP and LP levels were increased compared with age-matched controls. In the same patient population, followed for 1 year, only the cluster of baseline bone metabolism markers (comprising HP, LP, and serum bone sialoprotein)—out of the 14 molecular markers measured—significantly correlated with baseline clinical scores for pain, stiffness, and disease activity. None of the baseline cartilage or bone metabolism markers correlated with disease progression after 1 year of follow-up. In these patients, HP that is not derived from bone (extraosseous HP) showed a highly significant correlation with the acute-phase response, suggesting that in osteoarthritis, cartilage degradation is related to the degree of inflammation.

The glycosylated analogue of HP, glucosyl-galactosylpyridinoline (GGHP), is present in human synovial tissue and is released during its degradation in vitro. GGHP is virtually absent in bone, whereas low levels have been detected in muscle and liver and intermediate levels have been detected in cartilage. Methodological problems such as the stability of HP during the alkaline hydrolysis that is necessary to liberate glycosylated molecules from tissues have so far prevented solid data on tissue distribution. Urinary GGHP levels were increased in early RA patients versus controls, and baseline levels correlated with disease progression, albeit weakly. Pending definite information on its tissue distribution, urinary GGHP might prove to be a marker for synovial tissue degradation.

Overall, collagen cross-links seem useful in measuring bone and cartilage catabolism and contribute to understanding of the pathophysiology of osteoarthritis and RA. In this respect, age-related changes in articular cartilage may significantly affect its susceptibility to proteinase-mediated degradation; this underscores the use of proper age-matched control groups. Results obtained in studies with young animals cannot be extrapolated to the adult human situation.

**Collagen Synthesis**

In an attempt to repair tissue damage, collagen synthesis is increased in osteoarthritis cartilage, leading to increased tissue levels of the C-terminal propeptide of type II collagen (PIICP, chondrocalcin; also referred to as CPII). The rate of PIICP release is proportional to the rate of current collagen synthesis because the propeptide has a half-life of
approximately 18 hours. Synovial fluid PIICP levels are also increased and correlate with osteoarthritis severity and body mass index. PIICP levels were lower in serum of osteoarthritis patients than in that of healthy controls. Similarly, the serum N-propeptide of collagen type IIA (PIANP, an alternative splice variant that is expressed in embryonic and osteoarthritic cartilage) is lowered in osteoarthritis serum versus controls. Preliminary studies indicate that the balance between cartilage synthesis (PIANP) and cartilage degradation (urinary CTx-II) could be used to discriminate between patients with rapid compared with slow progression. More recently, a competition ELISA was developed for the N-propeptide of the type II collagen isoform that is expressed by mature adult chondrocytes (PIINP). Urine levels of PIINP were twofold higher than plasma levels in persons without radiographic signs of osteoarthritis. In patients with radiographic knee and/or hip osteoarthritis, plasma PIINP levels were fivefold lower than in the reference group without osteoarthritis. A comparable fivefold decrease was observed in RA patients, using a different ELISA. Absolute values of nanograms PIINP per milliliter plasma were comparable (=18 ng/mL) for both assays and control groups.

In inflamed synovial tissue, the synthesis of type III collagen is upregulated, resulting in the production of its N-terminal propeptide. In osteoarthritis patients and RA patients with knee involvement, serum PIINP levels are increased compared with age-matched controls. In RA patients, prednisolone treatment that resulted in clinical improvement also reduced serum PIINP levels by 25% and levels remained suppressed until treatment was withdrawn. Thus far, collagen type III–specific degradation markers have not been described, and because type III collagen has a broad distribution in soft tissues and blood vessels, its potential as a specific biomarker seems limited.

Proteoglycan Markers

The main noncollagenous constituent of articular cartilage is aggrecan, a large proteoglycan composed of a core protein to which glycosaminoglycan chains (e.g., keratan sulfate, chondroitin sulfate) are attached. Researchers have described various assays to measure aggrecan metabolism, but the available information is not always consistent. Depending on the antibodies used, serum keratan sulfate levels were reported to be either increased (antibody 5D4), decreased (antibody AN9P1), or similar to those in osteoarthritis patients compared with controls. Additionally, previous work has suggested that serum 5D4-reactive keratan sulfate levels are either similar or higher in osteoarthritis than in RA. In one study, the serum keratan sulfate levels (antibody AN9P1) were 30% lower in osteoarthritis patients than in age-matched controls. For RA patients, a negative correlation between serum keratan sulfate levels and inflammation has been found.

The aggrecan epitope 846, which reflects the synthesis of proteoglycans in an attempt to repair, was increased in cartilage of osteoarthritis patients, and synovial fluid levels correlated with other markers such as cartilage oligomeric matrix protein (COMP), PIICP, tissue inhibitor of metalloproteinasises 1 (TIMP-1), MMP-1, and MMP-3, as well as with the degree of radiologic damage. The epitope 846 levels in serum were lower, however, in osteoarthritis patients than in healthy controls and RA patients. In the RA patients, elevated levels of epitope 846 could predict a benign course of the disease. Taken together, none of the aggrecan-derived markers has shown sufficient power to discriminate between patients and controls or to provide consistent information that can be used in clinical studies. This effect may be partly caused by diurnal variation of these biomarkers, which may obscure relevant differences between study groups, especially when serum or urine sampling is not standardized. Similarly, increased motility of patients after initiation of successful treatment may affect circulating proteoglycan biomarker levels.

The identification of two members of the ADAM-TS family of proteases (ADAM-TS4 and ADAM-TS5) as aggrecanases supplied new tools to develop aggrecan-based markers for cartilage destruction. Antibodies directed at aggrecanase and MMP-generated neoepitopes in aggrecan core protein have been produced and applied in a variety of in vitro studies aimed at unraveling mechanisms of cartilage destruction in osteoarthritis. In synovial fluid from patients with a variety of joint diseases, MMP and aggrecanase-released aggrecan fragments have been detected. These studies have also shown that these two groups of proteases may have distinct roles in articular cartilage catabolism.

Hyaluronan

The glucuronic acid chain hyaluronan (hyaluronic acid) is a constituent of cartilage and synovium and is synthesized by many cell types. It functions as the anchor for proteoglycans such as aggrecan, allowing the formation of the large aggregates that are responsible for the resilience of cartilage. In the synovial membrane, hyaluronan is synthesized by synovial fibroblasts and secreted into the synovial fluid to provide lubrication of the joint and to facilitate joint movement. Synovial fluid hyaluronan levels are decreased in osteoarthritis patients and may partly explain impaired joint function and pain. This provides the rationale for visco-supplementation therapy in osteoarthritis patients, consisting of intra-articular injections with hyaluronan derivatives.

Elevated blood hyaluronan levels have been reported for osteoarthritis patients and RA patients. In RA, some studies failed to show a relationship between plasma hyaluronan levels and measures of disease activity, whereas others showed significant correlations between serum hyaluronan levels and a variety of measures of inflammation and destruction (e.g., CRP, ESR, Ritchie index, radiologic damage). In osteoarthritis patients, elevated serum hyaluronan levels correlated weakly with the degree of cartilage degeneration. In addition, baseline hyaluronan levels could predict progression of osteoarthritis and serum levels were shown to increase with disease severity. These results suggest that an increase in circulating systemic hyaluronan levels could reflect synovial inflammation rather than cartilage destruction, prompting care in use of hyaluronan as a joint destruction marker. Also, the observation that diet and increased physical activity can influence its serum levels should be considered when using hyaluronan as a biomarker in joint diseases.
Cartilage Oligomeric Matrix Protein

Since its discovery in the early 1990s, COMP has received much attention as a putative cartilage destruction marker. Although its exact function is unclear, COMP has been implicated in collagen fibrillogenesis. Increased COMP levels have been detected in the synovial fluid of osteoarthritis patients. In addition, COMP levels are increased in the serum of osteoarthritis patients compared with healthy controls and are associated with progression of radiographic signs of osteoarthritis. In osteoarthritis patients, the serum COMP levels were even higher than in RA patients. In early RA patients, increased serum COMP levels were identified as a strong predictor of early large-joint destruction. These data generated interest in the use of measurement of COMP levels as a selective cartilage destruction marker.

The expression of COMP in joint tissues other than cartilage including synovium, tendons, ligaments, and menisci raised concerns about its tissue specificity. High expression of COMP messenger RNA (mRNA) has been shown in murine osteoarthritis, indicating that synovial fluid COMP levels may reflect not only tissue degradation, but also the rate of new synthesis. The concerns about the use of COMP measurement as a specific cartilage degradation marker are fueled further by a study showing that the extent of synovial inflammation is one of the factors determining the serum COMP. In concordance, in early RA a positive correlation was found between serum COMP and the inflammatory marker CRP. Circulating COMP levels in this study were statistically higher in patients showing bone erosions on magnetic resonance imaging (MRI) than those without bone erosions. Other investigators did not observe a relationship between markers of inflammation and serum COMP levels (using a polyclonal antiserum recognizing all COMP forms) in RA patients. In this same study, COMP levels did not have any prognostic value with respect to progression of joint damage. Similar to the CTx and ICTP assays for collagen degradation, the use of antibodies or antiserum recognizing different epitopes within (fragments of) the COMP molecule might explain the apparent inconsistent results in the literature.

Metalloproteinases

In addition to cartilage breakdown products, metalloproteinases (MMPs and aggrecanases) and their endogenous inhibitors (TIMPs) that are involved in the pathologic degradation of joint tissues could serve as useful markers. Data on MMP levels as a predictor for the progression of joint erosion in early RA are rapidly accumulating. Serum and synovial fluid MMP-3 (stromelysin) levels are increased in RA patients compared with controls. MMP-3 levels correlate with inflammation markers and disease activity in patients with untreated active RA and in early RA patients who received nonsteroidal anti-inflammatory drugs only. In these studies, serum MMP-3 levels were not related to radiographic progression. Another study revealed an association between serum proMMP-3 concentrations at disease onset and progression of joint destruction, which was independent of known risk factors such as the presence of the shared epitope and serum levels of rheumatoid factor and CRP. Serum MMP-3 also correlated significantly with radiographic progression at entry and longitudinally in a study with early RA patients. In addition, it was demonstrated in a prospective study that increased serum MMP-3 is a predictive marker for the progression in disability. Several investigators showed that MMP-3 levels respond to therapeutic intervention with disease-modifying antirheumatic drugs (DMARDs), resulting in decreased MMP-3 level, which is associated with clinical improvement. Furthermore, baseline serum MMP-3 levels contribute to reaching normal physical function after 2 years of infliximab (anti-TNF) treatment.

MMP-1 (collagenase) levels have also been shown to indicate joint erosion independent of inflammation. In early RA patients, there was a positive correlation between the area under the curve measurements of MMP-1 serum levels (but not the area under the curve of CRP levels) and the number of new joint erosions after 18 months of follow-up. Arthritic patients treated with anti-TNF antibodies (infliximab) for 14 weeks did not show reduced MMP-2 and MMP-9 levels (assessed by zymography, which does not reliably detect the other MMPs) despite clear clinical improvement. These data suggest that MMP levels, although they may be correlated with parameters of inflammation, do not reflect exactly the same pathways as do acute-phase reactants and could provide valuable additional information on joint destruction in RA.

Extrapolated to osteoarthritis, in which secondary inflammation is usually mild, these data suggest that MMP levels may provide valuable predictive markers. In a cross-sectional study in osteoarthritis patients, MMP-3 serum levels were similar, however, to levels in healthy controls. In a subset (120 patients) of 431 patients participating in a randomized, placebo-controlled trial evaluating the effects of doxycycline treatment on unilateral knee osteoarthritis, baseline plasma MMP-3 levels predicted joint space narrowing. In another study, the serum levels of TIMP-1 did not differ significantly between patients with unilateral or bilateral hip osteoarthritis and healthy controls. Within the osteoarthritis group, patients with rapid disease progression (joint space narrowing > 0.6 mm/yr; 1-year follow-up) had significantly lower serum TIMP-1 levels than patients with slow progression (joint space narrowing < 0.6 mm/yr). Apart from these, comprehensive studies for predictive MMP markers in osteoarthritis are not yet available and information about the use of MMP levels in osteoarthritis is still incomplete.

In recent years, more data on the role of aggrecanase in degradation of the major proteoglycan of cartilage (aggrecan) became available. Its pivotal role in tissue destruction is now widely accepted. Because of the difficulties in measuring aggrecanase activity in biologic samples, however, their value as a biomarker to monitor joint destruction is not yet fully understood. It is hoped that the more recent development of several aggrecanase assays, especially the immunoaffinity-based liquid chromatography–tandem mass spectrometry (LC-MS/MS) method, which detects cleavage at the ARGS site and the AGEG site, will facilitate its use as a biomarker. The development of preclinical models and approaches to develop and evaluate selective aggrecanase inhibitors may contribute to the
development of antiosteoarthritis drugs, which concomitantly are tools to validate the biomarkers.\textsuperscript{40,158,159}

**Biologic Markers in Synovial Tissue**

Because many inflammatory arthropathies including RA primarily involve the synovial tissue, there has been increased interest in investigations of the pathologic changes in synovial biopsy specimens. This development has been stimulated further by technical advances such as the advent of new methods to obtain synovial tissue specimens from actively inflamed joints and clinically quiescent joints under local anesthesia and because of the development of immunohistologic methods, in situ hybridization, quantitative polymerase chain reaction, and microarray technology. Previous work has shown the relationship between the features of rheumatoid synovial tissue on the one hand and arthritis activity\textsuperscript{160} and joint destruction\textsuperscript{161} on the other (see also Chapter 53). The importance of evaluation of synovial tissue samples has been underscored by the observation that clinical signs of arthritis activity are associated with histologic signs of synovitis after treatment of RA patients with the monoclonal antibody alem-tuzumab (Campath-1H), despite profound depletion of circulating lymphocytes.\textsuperscript{162} Similarly, rituximab treatment leads to a rapid and significant decrease in synovial B cell numbers in only a subset of RA patients, whereas circulating B cells are completely depleted in nearly all patients (Figure 34-3).\textsuperscript{163-165}

Several methodological questions needed to be answered before serial synovial biopsy could be used to screen for potentially relevant effects after antirheumatic treatment.\textsuperscript{166} It has been shown in cross-sectional studies that biopsy samples can be acquired by blind needle technique and by miniarthroscopy.\textsuperscript{167} There are limitations and disadvantages, however, of the use of serial blind needle biopsy in the evaluation of treatment. It is usually restricted to larger joints such as the knee joint; the operator is not able to select the tissue visually, causing potential sampling error; and it is not always possible to obtain adequate tissue samples. This is especially true when clinically quiescent joints are investigated (e.g., after successful therapy). Comparison of the features of synovial inflammation in biopsy samples from inflamed knee joints and paired inflamed small joints of RA patients revealed that inflammation in one inflamed joint is generally representative of the inflammation in other inflamed joints.\textsuperscript{168} It is possible to use serial samples from the same joint, selecting either large or small joints, for the evaluation of antirheumatic therapy. Sampling error can be reduced by selecting at least six biopsy specimens from multiple regions, resulting in variance of less than 10%.\textsuperscript{22,169} When the biopsy samples are taken from an actively inflamed joint, there is on average no clear-cut difference in the features of synovial inflammation or the expression of mediators of inflammation and destruction at the pannus-cartilage junction compared with other regions away from the pannus-cartilage junction.\textsuperscript{170-172}

Ultrasound-guided biopsy is a newer technique, which can be performed in both small and large joints, bursae, and tendon sheaths under local anaesthesia.\textsuperscript{173} Although this method is appealing, further validation is necessary. A recommendation for standardization of various synovial biopsy techniques to be used in clinical trials has recently been published.\textsuperscript{174}

An extensive quality control system is required to allow reliable analysis by immunohistochemistry, tissue enzyme-linked immunosorbent assay, quantitative polymerase chain reaction, or microarray analysis.\textsuperscript{175} Finally, sophisticated computer-assisted image analysis systems allow reliable and efficient evaluation of the synovial cell infiltrate and the expression of adhesion molecules, cytokines, and MMPs in innovative clinical trials.\textsuperscript{176}

Using this approach, successful treatment with disease-modifying antirheumatic drugs such as gold,\textsuperscript{177} methotrexate,\textsuperscript{178,179} and leflunomide\textsuperscript{180} was shown to be associated with decreased mononuclear cell infiltration. Similarly, successful treatment of RA patients with infliximab,\textsuperscript{23,180,181} rituximab,\textsuperscript{164,182,183} and abatacept\textsuperscript{184} results in reduced

![Figure 34-3](image-url) Variable tissue response to treatment with the anti-CD20 antibody rituximab in patients with rheumatoid arthritis. Arthroscopic samples were obtained before (A and B) and 4 weeks after (C and D) initiation of treatment. Rheumatoid synovial tissue from paired biopsy samples stained for CD22+ B cells. Single staining, peroxidase technique, counterstaining with Mayer’s Hämalaunlösung. Original magnification 200×.
synovial inflammation. The number of macrophages in the
synovium was decreased already 24 to 48 hours after initia-
tion of infliximab treatment. More specifically, high-dose intra-
venous methylprednisolone reduced expression of TNF in
synovial biopsy samples 24 hours after treatment, a result
that correlated with a clinical response to, and subsequent
relapse after, methylprednisolone therapy.

A randomized trial has formally addressed the question
of which feature in RA synovial tissue samples could be used
as a biomarker for clinical efficacy in relatively small studies
of short duration. Patients received either prednisolone
according to the COBRA regimen or placebo for 2 weeks.
This study identified sublining macrophages as an important
immunohistologic biomarker associated with the clinical
response to corticosteroids. The utility of macrophages in
the synovial sublining as a candidate biomarker across dis-
crete interventions and kinetics has also been observed, with
a correlation between the mean change in disease activity score (Δ DAS28) and the mean change in the
number of sublining macrophages. When patients from
several actively treated studies were grouped, the standard-
ized response mean, a measure of sensitivity to change, was
1.16 for the change in DAS28 and 0.83 for the change in
sublining macrophages. For the patients from the placebo
groups, the standardized response mean was –0.23 (for
DAS28) and 0.30 (for macrophages), consistent with the
notion that the biomarker is less susceptible to placebo
effects or expectation bias than clinical evaluation, which
includes subjective measures of disease activity. In
addition to its role as a possible marker of clinical response
the change in numbers of sublining macrophages could
potentially help to distinguish effective from ineffective
treatment. Taken together, these studies suggest that analy-
ses of serial biopsy samples can be used as a screening
method to test new drug candidates requiring relatively
small numbers of subjects. The absence of changes after
treatment would suggest that the therapy is probably not
effective. The demonstration of biologic changes at the site
of inflammation could provide the rationale for larger,
placebo-controlled trials, however. Most of the biopsy
studies have been performed in RA patients, but it appears
likely that the same approach can be used for the evaluation
of novel therapies in patients with other rheumatologic
disease such as spondyloarthritis.

As an alternative to immunohistologic and in situ
hybridization methods, quantitative polymerase chain reac-
tion on small synovial biopsy specimens can be employed
to evaluate drug effects in clinical trials. Using this
approach, prednisolone was shown to reduce expression
of IL-8 and MMP-1 in synovial biopsy specimens of RA
patients after 2 weeks of treatment. Because the biopsy
specimens contain various cell types, it is important to
realize that a change in gene expression level may also
reflect a change in cellular composition of the biopsy, not
in the expression level within a certain cell type.

Biomarker Panels

Almost none of the markers for osteoarthritis and RA that
are currently used can distinguish successfully between
patients and controls on an individual basis, although
average marker levels differ among groups. Principal com-
ponent analysis of 14 biochemical markers revealed that
the markers segregate into five clusters: inflammation (IL-6,
CRP, TNF receptor I, TNF receptor II, and eosinophil cat-
tonic protein); bone (HP, LP; and BSP); cartilage synthesis
(CPII, epitope 846, and hyaluronan); cartilage degradation
(COMP and keratan sulfate); and transforming growth
factor-B1 (which is independent of all other markers). The
combination of these markers (TNF receptor II, COMP, and epitope 846) from the independent clusters
inflammation, cartilage degradation, and cartilage synthesis
could discriminate correctly between osteoarthritis patients
and controls in approximately 90% of the cases. In a study
of 376 patients with hip osteoarthritis, a similar approach
resulted in five different clusters with similar makeup: a
cartilage and bone cluster (PINP, CTx-I, and CTx-II); a
putative synovitis cluster (COMP, PIINP, and HA); a puta-
tive systemic inflammation cluster (CRP and YKL-40); and
MMP-1 and MMP-3 as two independent factors. Similar-
ly, in RA, a combination of seven clinical scores and
molecular markers provided a clinical prediction model that
could discriminate, at the first visit, among three forms
of arthritis—self-limiting arthritis, persistent nonerosive
arthritis, and persistent erosive arthritis. Responsiveness
of peripheral blood cells to various stimuli using a
10-cytokine profile resulted in an immunologic signature
that performed well in distinguishing early RA patients
from controls and also correlated with several markers of
disease severity in late RA. In contrast, the same 10-cytokine
profile assessed in serum was far less effective in discrimi-
inating the groups.

Different clusters of biomarkers may relate to osteoarthri-
tis at different joint sites, suggesting that pathophysiologic
processes may be different among those sites. These studies
support the notion that panels of biomarkers may provide a
valuable additional tool in the monitoring of osteoarthritis
patients and RA patients and in helping to understand
disease processes.

The progressive destruction of the articular cartilage is
considered a major determinant of disability in patients
with joint disease. A report showed that the balance
between cartilage synthesis and degradation discriminates
between osteoarthritis patients with rapid versus slow pro-
gression, as assessed by the change in joint space width and
arthroscopically scored chondropathy. These studies
support the hypothesis that the “omics” approaches, com-
bining even more markers than the few used previously, may
be successful in the identification of disease-specific finger-
prints and may also provide tools to monitor tissue-specific
degradation.

“OMICS” BASE BIOMARKERS

The current technical progress in genomics, transcrip-
tomics, proteomics, and metabolomics, in combination
with advanced bioinformatics, makes it possible to analyze
numerous markers in one sample, which could be urine,
serum, synovial fluid, or (synovial) tissue. The resulting
profile combines the levels of a variety of markers to create
a “disease fingerprint” that could serve as a powerful marker
by itself. In addition to specific markers for early diagnosis,
markers that specifically reflect cartilage degradation are necessary. These technologies can arbitrarily be separated into three primary levels: **genomics**, which deals with variations in DNA composition (e.g., single nucleotide polymorphisms) and expression levels (differences in mRNA levels, also known as transcriptomics); **proteomics**, which analyzes the proteome, or total of proteins in a sample; and **metabolomics**, which focuses on metabolites. Variations on these themes include technologies such as lipidomics (profiling the lipids and free fatty acids in a cell or biologic fluid), degradomics (the study of protein degradation products), and toponomics (the study of the localization of molecules within a cell).

**Genomics**

Of all the “-omics” technologies, genomics was the first to evolve in the footsteps of the human genome project, and various aspects of genomics approaches to study joint diseases have been extensively reviewed. Comparisons in gene expression levels among controls, osteoarthritis, and RA have been made for a variety of tissues such as chondrocytes, blood-derived cells, and synovial tissue. Within a group of RA patients, complementary DNA (cDNA) microarray analysis with a focus on immune-related genes could separate classes of patients with potentially different pathogenicity on the basis of expression of genes involved in the adaptive immune response versus genes involved in tissue remodeling. In osteoarthritis, chondrocytes from cartilage have been shown to upregulate the transcription of a variety of inflammatory genes. In another study, 3543 genes were differentially expressed by blood cells of patients with mild knee osteoarthritis compared with healthy controls. Logistic regression indicated that nine of these genes were discriminatory between subjects with mild osteoarthritis and controls, with a sensitivity of 86% and specificity of 83% in a training set of 78 samples. The optimal biomarker combinations were evaluated using a blind test set (67 subjects), which showed 72% sensitivity and 66% specificity for the diagnosis of osteoarthritis. These data underscore that the combination of biomarkers (in this case the expression of nine genes) may be useful in differential diagnosis.

Also in other rheumatic diseases, expression profiling has contributed to the understanding of disease pathways. In patients with systemic lupus erythematosus, several studies have shown that interferon-regulated genes are highly upregulated in peripheral blood cells and in kidney glomeruli.

One of the ultimate uses of the identification of differentially expressed genes in a disease is illustrated by the antitumor drug trastuzumab (a recombinant monoclonal antibody against the human epidermal growth factor receptor 2 [HER2] protein), which would not have reached the marker if not for the accompanying prognostic test. Normal cells express low levels of HER2 protein on their plasma membrane. In approximately 25% of breast cancer patients, HER2 is overexpressed, changing the growth control of these cells. The prognostic test measures the expression levels of HER2, assisting in the selection of patients who would benefit from trastuzumab treatment.

**Proteomics and Lipidomics**

Similar to the genomics revolution, which was partly driven by technology that allowed the production of gene-chip and high-throughput DNA sequencing methods, the proteomics field was boosted by the development of better two-dimensional electrophoresis technologies, protein and antibody arrays, and the rapid improvements in the area of mass spectroscopy, all of which facilitated the reproducible analysis of a panel of proteins within a sample. Two-dimensional gel electrophoresis has been used to identify proteins secreted into the culture medium of normal and osteoarthritic cartilage samples but also to analyze the protein composition of the mitochondria of healthy human chondrocytes. Using surface-enhanced laser desorption/ionization time-of-flight mass spectroscopy, 103 serum samples of RA patients; osteoarthritis patients; patients with non-RA inflammatory conditions (psoriatic arthritis, asthma, Crohn’s disease); and controls were analyzed. This approach yielded several signals in the mass spectrum that contributed to the separation between RA patients and controls. A different approach was taken by Xiang and coworkers, who first separated human chondrocyte proteins by two-dimensional electrophoresis, then blotted the proteins to a membrane, and finally incubated these membranes with serum of osteoarthritis patients, RA patients, and controls to identify which chondrocyte-derived autoantigens are present in these patients. This approach yielded triosephosphate isomerase as a potential osteoarthritis-specific biomarker. Using a protein microarray platform that simultaneously evaluates the presence and abundance of 169 proteins relevant to inflammation, cell growth, activation, and metabolism, 16 serum proteins were found different between OA cases compared with controls. Yet another approach focuses on the panels of autoantibodies present in patients with various autoimmune diseases to act as biomarkers. Using an array of 30 antigens known to be expressed in the glomeruli, researchers studied the clusters of autoantibodies that occur in the serum of lupus patients and found they were related to the patients’ disease activity. All biomarkers that are identified by the previously described examples naturally need further validation to establish their true usefulness for diagnostic, prognostic, or disease-monitoring application in patients with joint disease.

From recent advances in analytic methods to simultaneously analyze multiple lipid species emerges the field of lipidomics, which can be divided into two biochemical areas of equal significance: membrane functional-lipidomics and mediator functional-lipidomics. Osteoarthritis has long been considered a disease in which irreversible degradation of articular cartilage was the pivotal pathophysiologic process. Recent studies that indicate the development of osteoarthritis may be related to the coexistence of disordered glucose and lipid metabolism have triggered the evaluation of the lipidome in osteoarthritic patients. Using ultra-performance liquid chromatography coupled to time-of-flight mass spectroscopy (UPLC-ToF-MS), clear differences in serum lipid composition were observed between subjects with no, mild, or moderate osteoarthritis.
**Metabolomics**

Biologic fluids such as urine, blood, and synovial fluid contain numerous metabolites that may provide valuable information on the metabolism of an organism and about its health status. Metabolic profiling, also referred to as metabolomics, metabonomics, or related terms, is defined as the quantitative and qualitative analysis of the whole complement of small molecules in a sample (e.g., cell, tissue, body fluids). The technology has emerged from approaches used to profile body fluids that were developed many decades ago for the study of inborn errors of metabolism and the effects of nutrition. A wide array of analytic methods is used to analyze the various metabolites. Gas chromatography—mass spectroscopy and nuclear magnetic resonance (NMR) can be employed for a global insight into a broad range of metabolites such as (phosphorylated) sugars, amino acids, fatty acids, nucleobases and nucleosides, amines, higher alcohols, and bile acids. Liquid chromatography—mass spectroscopy methods can be used to not only zoom in on free fatty acids and lipids but also analyze amino acids, peptides, sugars, aminosugars, hormones, and steroids. All of these analytic measures need to be combined with data preprocessing to obtain clean data. This is necessary to analyze these large metabolite profiles reliably and to relate relevant changes in metabolites to biologic processes, using multivariate statistics.

The application of metabolomics in the area of joint diseases is relatively recent. 1H-NMR (500 MHz) has been used to compare the effects of unilateral knee joint denervation on the biochemical profiles of synovial fluid in a bilateral canine anterior cruciate ligament transection model of osteoarthritis. Increases in glycerol, hydroxybutyrate, glutamine/glutamate, creatinine/creatinine, acetate, and N-acetyl-glycoprotein concentrations were observed in synovial fluids from denervated osteoarthritis knees compared with normally innervated osteoarthritis knees. These metabolite profiles of denervated osteoarthritis knees support the idea of neurogenic acceleration of osteoarthritis in that the observed differences in metabolite concentrations found in the denervated knee fluids seem to correlate with metabolic changes resulting from aggravation of the osteoarthritis process caused by joint denervation. Using 1H-NMR (300 MHz) and multivariate data analysis, a metabolite profile was detected, which was strongly associated with osteoarthritis in 10- and 12-month-old Hartley guinea pigs that spontaneously develop osteoarthritis. 1H-NMR also revealed a urinary metabolite profile that could distinguish osteoarthritis patients from healthy individuals. The human urine profile largely resembles the Dunkin Hartley guinea pig profile; the presence of hydroxybutyrate, pyruvate, creatine/creatinine, and glycerol in the metabolite profile could point to an enhanced use of fat and altered energy use, consistent with the canine synovial fluid composition.

**SYSTEMS BIOLOGY**

Although each is already tremendously powerful on its own, the combination of genomics (transcriptomics), proteomics, and metabolomics theoretically could deliver a full picture of a living system (cell, organ, or organism). Such a systems biology approach would provide insight into which disturbances of a healthy system cause it to shift toward a pathologic phenotype, which mechanisms are employed by the organism to maintain its equilibrium, and which factors indicate a point of no return, followed by failure of the intrinsic balancing mechanisms and disease initiation. As such, a systems biology approach would help to identify the most promising molecules that describe this shift and can act as biomarkers, while concomitantly key molecules can be detected that, when normalized, could rebalance the system, acting as therapeutic targets. The first steps in this area are being made for RA, but steps for osteoarthritis projects also have been initiated.

**BIOMARKER VALIDATION AND APPLICATION**

Following or parallel to the crucial investigations to identify relevant biomarkers for disease, validation studies need to be performed that focus on analytic aspects and clinical validity: Does the biomarker reflect the disease process, and how does it change with endogenous or drug-induced changes in pathophysiologic processes? The actual application of such biomarkers in preclinical or clinical studies requires an in-depth analytic validation. The obvious reason for this is that to draw conclusions on the basis of biomarker data, it is essential to be able to trust that a measured value is reliable and reproducible. The HELIX-II example (see earlier) illustrates the importance of characterizing the analyte of interest and ensuring that the chosen assay detects that specific analyte.

Some of the essential steps in biomarker validation are described next. Details on validation procedures and requirements can be found elsewhere (www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM267449.pdf).

The fundamental technical parameters to show that a given biomarker can be quantitatively measured in a given biologic matrix (e.g., serum, urine, synovial fluid, saliva) include the following:

1. **Accuracy**—how close is the mean measured concentration of at least five replicates to the true value of the analyte?
2. **Precision**—what is the variation between individual measurements of one sample? Typically, at each test concentration, the precision should not exceed the 15% of the coefficient of variation.
3. **Selectivity or specificity**—how well does the analytic method distinguish between the analyte of interest and other components of the samples?
4. **Sensitivity**—what is the smallest amount (or the largest amount) of analyte that can be reliably detected?
5. **Reproducibility**—how well can the measurements be repeated on a different day, by different operators, or by using different equipment and still result in the same measured values?
6. **Stability**—how stable is the analyte of interest in a certain matrix, tube, or storage condition? This also includes studies of the stability of the analyte on repeated freeze-thaw cycles and short-term and long-term storage at various temperatures.
Other factors that are important in validating biomarker assays are the availability of references or calibrators that are used as external standards to quantify the results. In addition, the development of standard operating procedures and documentation is essential for optimal assay performance.

Many biomarkers (e.g., urinary CTx-II) are measured in body fluids such as urine, serum, or plasma and related to disease end points in one joint (e.g., radiologic knee osteoarthritis), thereby ignoring the contribution of other joints and tissues to the systemic biomarker levels. This commonly used approach is one of the underlying reasons for the inconsistent performance of various osteoarthritis biomarkers. Proper biomarker validation would include relating the systemic biomarker level to whole body burden of disease, which is hampered by the inability to fully phenotype the burden of osteoarthritis in a patient. Osteoarthritis presence and severity in single or sometimes multiple joints are currently documented by radiographs and/or magnetic resonance imaging. Using this approach on a routine basis for whole body burden of disease is impractical due to radiation exposure, time, and cost. Therefore proper biomarker validation would have the concomitant advantage of providing a cost-effective alternative to whole body imaging. An initial study suggests that indeed systemic joint tissue concentrations of several biomarkers can be quantitative indicators of specific subtypes of osteoarthritis and of total body burden of disease.6

The applicability of biomarkers does require more than validation. Sharing reagents, standards, and procedures among laboratories enables the comparison of study results and therefore the advancement of the field. The BIPEDS classification (which stands for Burden of Disease, Investigative, Prognostic, Efficacy of Intervention, Diagnosis of Disease, and Safety of Intervention) for osteoarthritis biomarkers was developed by the National Institutes of Health–funded osteoarthritis Biomarkers Network to improve the ability to develop and analyze osteoarthritis biomarkers, as well as to communicate these advances within a common framework.227 A review of the status of commercially available biochemical markers for primary knee and/or hip osteoarthritis according to the BIPED classification revealed an uneven distribution of scores on biochemical marker performance, and heterogeneity among the 84 evaluated publications complicated direct comparison of individual biochemical markers.227 This underscores the need for international standardization of future investigations to obtain more high-quality, homogeneous data on the full spectrum of biochemical markers.

In the field of RA the Outcome Measures in Rheumatology (OMERACT) initiative is an informal international network that strives to improve end point outcome measurement through a data-driven, interactive consensus process. The initiative has established an OMERACT filter to assess whether a measure is applicable. The filter can be summarized in three words: (1) truth (does it measure what it intends to measure?), (2) discrimination (does the measure discriminate between situations of interest?), and (3) feasibility (can the measure be applied easily, given constraints of time, money, and interpretability?). At the OMERACT 8 meeting a draft set of criteria for the validation of soluble biomarkers reflecting damage end points was proposed, and the Soluble Biomarker Group revised it at OMERACT 9.228,229 The set of criteria (OMERACT 9 v2 criteria) focuses on the performance characteristics of biomarker assays, the importance of addressing potential confounders, and the requirement for clinical validation studies. Well-designed prospective studies adhering to the guidelines formulated by the Soluble Biomarker group should be performed to establish if a (panel) of soluble biomarkers can reliably reflect the disease processes in RA and replace the measurement of joint destruction in RA.230

Given the current developments in the field of RA with recommendations to treat early in the disease with DMARDs for the best outcome on disease progression, it becomes increasingly important to identify RA patients at an early stage. To this end, experts developed a new set of criteria for classification that consists of clinical and inflammatory markers (i.e., CRP, RF, ACPA, and ESR) and has to prove its value in future studies.231,232 On the other hand, a prediction rule has been developed. It includes several of the classical soluble biomarkers (CRP, RF, and ACPA) to classify the patients and has subsequently been validated in several other studies.234,237

Future challenge also lies in the prediction of response to a certain treatment. For instance, about one third of the RA patients does not respond to anti-TNF treatment. Patients would greatly benefit from developments that predict their responsiveness to a treatment. It would not only avoid exposing patients to potential risks known to come with the use of certain biologics, but also positively influence the economic burden of treating RA patients. In literature, the focus has predominantly been on predicting the response to anti-TNF therapies because this therapy is widely applied once the response to methotrexate (first line of treatment) is not sufficient. Several approaches have been used by investigators to predict the response to anti-TNF therapies including the use of single biomarkers, panels of biomarkers, expression profiles, and the presence of lymphocyte aggregates in the synovium.238,243 All of these prediction models should be validated in independent large cohort studies to confirm their value for distinguishing responders from nonresponders to anti-TNF therapies.

CONCLUSION

Over the years, many reports have been published that employ molecular markers in body fluids to assess inflammation and tissue destruction in joint diseases. Starting from single markers with limited tissue or disease specificity, panels of biomarkers are increasingly used and these become more and more tissue specific. The novel markers include collagen-based markers (telopeptides, neoepitopes, and cross-links), COMP, and MMPs, which are often included in clinical trials. Apart from well-known markers such as CRP, ESR, rheumatoid factor, anticitrullinated protein/peptide antibodies, and a few other autoantibodies, none of the markers has made it to the clinic for routine evaluation of patient disease status. The requirements for such a clinically useful marker are high because it should be superior to existing markers in being able to predict disease progression or monitor therapy efficacy in individual patients. In this respect, biomarkers for osteoarthritis provide the greatest challenge: Joint destruction often proceeds without signs of
inflammation. Validation of osteoarthritis biomarkers is hampered by the absence of a generally accepted, effective treatment for the disease. In early RA patients, suppression of inflammation often, although not always, coincides with decreased joint destruction, and monitoring inflammation may fulfill its role as surrogate destruction marker. In later stages of the disease, when inflammation and destruction seem more uncoupled, specific destruction markers also are necessary for RA patients. In general, the combination of multiple markers holds most promise to meet these needs to increase disease specificity or tissue specificity, or both, and to reduce the extensive overlap in marker levels that exists between patients and controls.

Analysis of molecular markers in synovial tissue is increasingly used, especially in clinical trials on targeted therapies. Tissue specificity is not a problem, and examination of serial biopsy samples may be used to monitor the response in individual patients and screen for interesting biologic effects at the site of inflammation. This approach is generally well tolerated by patients, but it requires a more demanding setup. It can be anticipated that future development will include the use of more extensive markers of joint degradation—in addition to the available markers of inflammation—and the use of panels of biomarkers in synovial tissue samples.

As illustrated by studies described in this chapter, many investigators measure different sets of biomarkers and may use different definitions of disease (or progression or both). In addition, common terminology to describe a biomarker is lacking and investigators may have a historical bias in favor of (or against) certain biomarkers. In combination, these issues may slow down the urgently needed progress in the development of clinically applicable biomarkers for joint diseases. To solve this, further collaboration between researchers of various disciplines and the execution of large, unbiased studies incorporating a wide panel of available (or newly developed) biomarkers and complementary methods including imaging and patient assessments are necessary.

### Selected References


Full references for this chapter can be found on www.expertconsult.com.


129. Clark AG, Jordan JM, Vilim V, et al: Serum cartilage oligomeric matrix protein reflects osteoarthritis presence and severity: the John- 
son County Osteoarthritis Project, Arthritis Rheum 42(11):2356– 
2364, 1999.

130. Vilim V, Olejarova M, Machacek S, et al: Serum levels of cartilage 
oligomeric matrix protein (COMP) correlate with radiographic 
progression of knee osteoarthritis, Osteoarthr Cartil 10(9):707–713, 
2002.

knee joints of patients with rheumatoid arthritis and osteoarthritis. 
Involvement of YKL-40 in the joint pathology, Osteoarthr Cartil 

levels of cartilage oligomeric matrix protein in a subgroup of patients 

oligomeric matrix protein (COMP) and bone markers in 
knee joint synovial fluid in rheumatoid arthritis: relation to develop-
ment of knee or hip joint destruction, Ann Rheum Dis 56(2):91–96, 
1997.

134. Wollheim FA, Eberhardt KB, Johnson U, Saxne T: HLA DRB1* 
5101/5102 and cartilage oligomeric matrix protein (COMP) as predictors 
of joint destruction in recent-onset rheumatoid arthritis, Br J Rheu-

oligomeric matrix protein at the onset of articular cartilage degenera-
tion in a transgenic mouse model of osteoarthritis, Arthritis Rheum 

matrix protein reflects the presence of clinically diagnosed synovitis in 
knee joints with osteoarthritis, Osteoarthr Cartil 9(7):612–618, 

oligomeric matrix protein determines the subset of patients with 
early-stage rheumatoid arthritis with high serum C-reactive protein, 
matrix metalloproteinase-3, and MRI-proven bone erosion, J Rheu-

proteinases, their inhibitors and cartilage oligomeric matrix protein: 
relationship to radiological progression and inflammation in early 
rheumatoid arthritis. A prospective 5-year study, Rheumatology 

antibodies recognizing different fragments of cartilage oligomeric 
matrix protein in human body fluids, Arth Biochem Biophys 341(1):8– 

serum levels of matrix metalloproteinase 1, matrix metalloproteinase 
3, and tissue inhibitor of metalloproteinases 1 in rheumatoid arthri-

serum levels are correlated with disease activity and predict clinical 

3, -8, -9 as markers of disease activity and joint damage progression 
in early rheumatoid arthritis, Ann Rheum Dis 62(11):1094–1099, 
2003.

progression of knee osteoarthritis, Arthritis Rheum 54(7):902– 

may predict progression of knee osteoarthritis, Osteoarthr Cartil 

145. Stanton H, Rogerson FM, East CJ, et al: ADAMT5 is the major 
aggrecanase in mouse cartilage in vivo and in vitro, Nature 

ADAMTS5 prevents cartilage degradation in a murine model of 

147. Pratta MA, Su JL, Leesnitzer MA, et al: Development and charac-
terization of a highly specific and sensitive sandwich ELISA for detec-
tion of aggrecanase-generated aggrecan fragments, Osteoarthr Cartil 

the screening of ADAMTS-4 inhibitors, Matrix Biol 25(4):261–267, 
2006.

149. Dufield DR, Nemirovsky OV, Jennings MG, et al: An immunooas-
infinity liquid chromatography-tandem mass spectrometry assay for detec-
tion of endogenous aggrecan fragments in biological fluids: use as a 
biomarker for aggrecanase activity and cartilage degradation, Anal 

codynamic model for monitoring aggrecanase activity: injection of 
monosodium iodoacetate (MIA) in rats and assessment of aggrecan 
neoptipe release in synovial fluid using novel ELISAs, Osteoarthr Cartil 

151. Li WW, Nemirovsky O, Fountain S, et al: Clinical validation of an 
immunosaffinity LC-MS/MS assay for the quantification of a collagen 
type II neoptipe peptide: a biomarker of matrix metalloproteinase 

infiltrate in early rheumatoid synovial tissue in relation to local 

ase gene expression and joint erosion in early rheumatoid arthritis, 

response to treatment with Campath-1H, Arthritis Rheum 38(2):254– 

on the synovial cell infiltrate in patients with rheumatoid arthritis, 

immunomodulatory synovial effects (ARISE trial). 1: clinical and 

in rheumatoid arthritis synovium following 8 weeks of rituximab 

158. Gerlag D, Tak PP: Synovial biopsy, Best Pract Res Clin Rheumatol 

analysis of inflammation in rheumatoid arthritis synovial membrane 
samples selected at arthroscopy compared with samples obtained 

from the knee joints and the small joints of rheumatoid arthritis 
patients: implications for pathogenesis and evaluation of treatment, 


Occupational and Recreational Musculoskeletal Disorders

RICHARD S. PANUSH

KEY POINTS

Some occupational and recreational activities have been linked with musculoskeletal syndromes or disorders. These include certain syndromes manifested by neck pain; shoulder, elbow, hand, or wrist pain or tendinitis; carpal tunnel syndrome; and hand-arm vibration syndrome.

The concepts of so-called cumulative trauma disorders and repetitive strain disorders, though intuitive, are poorly supported by the literature. Causal relationships between most occupations or activities and these “syndromes” have not been well established.

Some activities and mechanical stresses have been associated with osteoarthritis at certain sites—for example, the hips of farmers, the knees of workers whose jobs involve frequent knee bending, and the hands of workers doing repetitive tasks with their hands. Seamstresses, diamond workers, and textile workers are among the latter.

Certain rheumatic disorders have been related to environmental or occupational risks such as Raynaud’s phenomenon with vibration and polyvinyl chloride; autoimmune disease with schoolteachers; scleroderma with chlorinated hydrocarbons, organic solvents, and silica; scleroderma-like syndromes with rapeseed oil and L-tryptophan; lupus syndromes with canavanine, hydrazine, mercury, pesticides, silica, mercury, paints, dyes, nail polish, and solvents; vasculitis with farming, silica, solvents, and allergies; granulomatous vasculitis with mercury and lead; lupus, scleroderma, and Paget’s disease with pet ownership; rheumatoid arthritis with silica (Caplan’s syndrome); and saturnine gout with lead exposure.

Putting a normal joint through its physiologic range of motion is not necessarily harmful for an otherwise healthy individual; however, if the joint, motion, stress, or biomechanics are not normal, there may be a risk of joint harm.

Most normal individuals comfortably engaging in reasonable recreational activities can do so without evidence of lasting soft tissue or articular damage; runners have been best studied. Conversely, individuals who exercise with pain, effusions, underlying joint abnormalities (e.g., ligamentous or meniscal damage), or abnormal or unusual biomechanics or as professional or elite athletes (e.g., boxers, American football or soccer players) seem to be at increased risk of joint injury.

Performing artists, vocalists, dancers, and musicians have a risk of soft tissue and joint injury analogous to that of athletes.

“...The diseases of persons incident to this craft arise from three causes: first constant sitting, second the perpetual motion of the hand in the same manner, and thirdly the attention and application of the mind. … Constant writing also considerably fatigues the hand and whole arm on account of the continual and almost tense tension of the muscles and tendons. I knew a man who, by perpetual writing, began first to complain of an excessive weariness of his whole right arm, which could be removed by no medicines, and which was at last succeeded by a perfect palsy of the whole arm.”

—Ramazzini, 1713

“...When job demands … repeatedly exceed the biomechanical capacity of the worker, the activities become trauma-inducing. Hence, traumatogens are workplace sources of biomechanical strain that contribute to the onset of injuries affecting the musculoskeletal system.”

This chapter discusses the possible association of certain occupational and recreational activities with musculoskeletal disorders. It has been conventional wisdom that “wear and tear” from at least some activities leads to reversible or irreversible damage to the musculoskeletal system. Despite the apparent logic that work or recreational activities might cause rheumatic and musculoskeletal disorders or soft tissue syndromes, this putative association is controversial and perhaps seriously flawed. There are confounding aspects to many of the available data including imprecise diagnostic labels, subjectivity of complaints, anecdotal and survey data, inadequate controls, differing definitions of disease and disability, limited duration of follow-up observations, inadequate epidemiology, inferential observations, difficulty quantifying activities and defining health effects, assumptions of the validity of claims data, variable quality of reported observations, psychologic factors influencing symptoms, and conflicting data.

OCCUPATION-RELATED MUSCULOSKELETAL DISORDERS

Many presumptive work-related musculoskeletal disorders have been described and are summarized in Table 35-1. These have been reported as sprains, strains, inflammations, dislocations, and irritations. Work-related musculoskeletal injuries comprise at least 50% of nonfatal injury cases resulting in days away from work. The cost of work-related disability from musculoskeletal disorders has been equivalent to approximately 1% of the United States’ gross national product, making these entities of considerable

Table 35-1

<table>
<thead>
<tr>
<th>Occupation-Related Musculoskeletal Disorders</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sprains, Strains, Inflammations, Dislocations, and Irritations</td>
<td>Commonly reported injuries resulting from repetitive motion or trauma at work.</td>
</tr>
<tr>
<td>Cumulative Trauma Disorders</td>
<td>Described as repetitive strain injuries, these disorders involve chronic pain and inflammation.</td>
</tr>
<tr>
<td>Repetitive Motion Injuries</td>
<td>Associated with occupations requiring repetitive movements, such as assembly line workers.</td>
</tr>
<tr>
<td>Vibration-Associated Disorders</td>
<td>Include carpal tunnel syndrome and other conditions linked to prolonged exposure to vibrations.</td>
</tr>
<tr>
<td>Occupational Arthritis</td>
<td>Involves degenerative joint disease associated with specific work environments.</td>
</tr>
</tbody>
</table>

RICHARD S. PANUSH

KARINA D. TORRALBA
Table 35-1  Occupation-Related Musculoskeletal Syndromes

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Gamekeeper’s thumb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cherry pitter’s thumb</td>
<td>Espresso maker’s wrist</td>
</tr>
<tr>
<td>Staple gun carpal tunnel syndrome</td>
<td>Espresso elbow</td>
</tr>
<tr>
<td>Bricklayer’s shoulder</td>
<td>Pizza maker’s palsy</td>
</tr>
<tr>
<td>Carpenter’s elbow</td>
<td>Poster presenter’s thumb</td>
</tr>
<tr>
<td>Janitor’s elbow</td>
<td>Rope maker’s claw hand</td>
</tr>
<tr>
<td>Stitcher’s wrist</td>
<td>Telegraphist’s cramp</td>
</tr>
<tr>
<td>Cotton twister’s hand</td>
<td>Waiter’s shoulder</td>
</tr>
<tr>
<td>Writer’s cramp</td>
<td>Ladder shins</td>
</tr>
<tr>
<td>Bowler’s thumb</td>
<td>Tobacco primer’s wrist</td>
</tr>
<tr>
<td>Jeweler’s thumb</td>
<td>Carpet layer’s knee</td>
</tr>
</tbody>
</table>


One review summarized that “most believe scientific data are insufficient to establish a definite causal relationship of these so-called cumulative trauma disorders to the worker’s lifting.” Other risk factors for work-related musculoskeletal disorders involving the back included awkward posture, high static muscle load, high-force exertion at the hands and wrists, sudden applications of force, work with short cycle times, little task variety, frequent tight deadlines, inadequate rest or recovery periods, high cognitive demands, little control over work, cold work environment, localized mechanical stresses to tissues, and poor spinal support.

The development of recommended treatment methods (rehabilitation) for these so-called occupational musculoskeletal disorders has included collaboration by workers, employers, insurers, and health professionals. The process has been divided into three phases: protection from and resolution of symptoms, restoration of strength and dynamic stability, and return to work. This process included symptomatic therapies, physical therapy, and ergonomic evaluation.

Until recently, the prevailing view was that many musculoskeletal disorders were consistently and predictably work related. That understanding has now come under considerable scrutiny and criticism. Despite the quantity of published information (see Table 35-2), the previously cited literature about occupational musculoskeletal disorders is now considered flawed; its quality was uneven and perhaps poor in some instances. Definitions of musculoskeletal disorders were imprecise; diagnoses, by rheumatologic standards, were infrequent; studies were usually not prospective, and there were selection biases; psychologic influences and secondary gain were often ignored; questionnaires were often used without validation of subjective complaints; and quantification of putative causative factors was difficult. Indeed, a review of this literature concluded that none of the published studies satisfactorily established a causal relationship between work and distinct medical entities.

In fact, certain experiences argued forcefully against the notion of work-related musculoskeletal disorders. In Lithuania, for example, where insurance was limited and disability was not a societal expectation or entitlement, “whiplash” from auto accidents did not exist. In Australia, when legislation for compensability was made more stringent, an epidemic of whiplash and repetitive-strain injuries abated.

In the United States, too, expressed symptoms correlated closely with the likelihood of obtaining compensation. In other cases, ergonomic interventions had no effect on alleged work-related symptoms and close analysis of epidemics of work-related musculoskeletal disorders revealed serious inconsistencies. These concerns led the American Society for Surgery of the Hand to editorialize that “the current medical literature does not provide the information necessary to establish a causal relationship between specific work activities and the development of well-recognized disease entities. Until scientifically valid studies are conducted, the society urges the government to exercise restraint in considering regulations designed to reduce the incidence of these conditions because premature regulations could have far-reaching legal and economic effects, as well as an adverse impact on the care of workers.”

One review summarized that “most believe scientific data are insufficient to establish a definite causal relationship of these so-called cumulative trauma disorders to the worker’s lifting.” Other risk factors for work-related musculoskeletal disorders involving the back included awkward posture, high static muscle load, high-force exertion at the hands and wrists, sudden applications of force, work with short cycle times, little task variety, frequent tight deadlines, inadequate rest or recovery periods, high cognitive demands, little control over work, cold work environment, localized mechanical stresses to tissues, and poor spinal support.

The development of recommended treatment methods (rehabilitation) for these so-called occupational musculoskeletal disorders has included collaboration by workers, employers, insurers, and health professionals. The process has been divided into three phases: protection from and resolution of symptoms, restoration of strength and dynamic stability, and return to work. This process included symptomatic therapies, physical therapy, and ergonomic evaluation.

Until recently, the prevailing view was that many musculoskeletal disorders were consistently and predictably work related. That understanding has now come under considerable scrutiny and criticism. Despite the quantity of published information (see Table 35-2), the previously cited literature about occupational musculoskeletal disorders is now considered flawed; its quality was uneven and perhaps poor in some instances. Definitions of musculoskeletal disorders were imprecise; diagnoses, by rheumatologic standards, were infrequent; studies were usually not prospective, and there were selection biases; psychologic influences and secondary gain were often ignored; questionnaires were often used without validation of subjective complaints; and quantification of putative causative factors was difficult. Indeed, a review of this literature concluded that none of the published studies satisfactorily established a causal relationship between work and distinct medical entities.

In fact, certain experiences argued forcefully against the notion of work-related musculoskeletal disorders. In Lithuania, for example, where insurance was limited and disability was not a societal expectation or entitlement, “whiplash” from auto accidents did not exist. In Australia, when legislation for compensability was made more stringent, an epidemic of whiplash and repetitive-strain injuries abated.

In the United States, too, expressed symptoms correlated closely with the likelihood of obtaining compensation. In other cases, ergonomic interventions had no effect on alleged work-related symptoms and close analysis of epidemics of work-related musculoskeletal disorders revealed serious inconsistencies. These concerns led the American Society for Surgery of the Hand to editorialize that “the current medical literature does not provide the information necessary to establish a causal relationship between specific work activities and the development of well-recognized disease entities. Until scientifically valid studies are conducted, the society urges the government to exercise restraint in considering regulations designed to reduce the incidence of these conditions because premature regulations could have far-reaching legal and economic effects, as well as an adverse impact on the care of workers.”

One review summarized that “most believe scientific data are insufficient to establish a definite causal relationship of these so-called cumulative trauma disorders to the worker’s
CHAPTER 35 | OCCUPATIONAL AND RECREATIONAL MUSCULOSKELETAL DISORDERS

OCCUPATION-RELATED RHEUMATIC DISEASES

Work-related rheumatic diseases have not been consistently well studied, but associations between occupations and well-defined rheumatic disorders are clearer than those involving musculoskeletal disorders. This topic also recapitu-
lates the simplistic notion that joints deteriorate with use. However, this perception is neither necessarily logical nor correct.

Osteoarthritis

Is osteoarthritis (OA) caused, at least in part, by mechanical stress? One analytic approach to determining a possible relationship between activity and joint disease is to consider the epidemiologic evidence that degenerative arthritis may follow repetitive trauma. Most discussions of the pathogenesis of OA include a role for “stress.” Studies of skeletons of several populations have sug-gested that age at onset, frequency, and location of osteoarthritic changes were directly related to the nature and degree of physical activities. However, not all these studies adhered to contemporary standards, nor have they been confirmed. One report, for example, failed to find an increased incidence of OA in pneumatic drill users and criticized inadequate sample sizes, lack of statistical analyzes, and omission of appropriate control populations in previous reports. The investigators further commented that earlier work was “frequently misinterpreted” and that their studies suggested that “impact, without injury or preceding abnormality of either joint contour or ligaments, is unlikely to produce osteoarthritis.”

Do epidemiologic studies of OA implicate physical or mechanical factors related to disease predisposition or develop-ment? The first national Health and Nutrition Examination Survey of 1971 to 1975 (HANES I) and the Framingham studies explored cross-sectional associations between radiographic OA of the knee and possible risk factors. Strong associations were noted between knee OA and obesity and those occupations involving the stress of knee bending, but not all habitual physical activities and leisure-time physical activities (running, walking, team sports, racquet sports, and others) were linked with knee OA. (See Chapter 98 for more information concern-ing the pathogenesis of OA.)

Other Occupational Rheumatologic Disorders

Certain rheumatic diseases other than repetitive strain or cumulative trauma disorders have been associated with occupational risks. These included reports of reflex sympathetic dystrophy after trauma; Raynaud’s phenomenon with vibration or exposure to chemicals (polyvinyl chloride); autoimmune disease from teaching school, farming and occupations with exposure to animals and pesticides; mining, textile machine and decorating operations.
Increased trauma to the hand, knee, hip, shoulder, elbow, wrist, metacarpophalangeal (MCP) joint, and other areas have been associated with osteoarthritis (OA) in various studies.

### Table 35-3: Occupational Physical Activity and Possible Associations with Osteoarthritis

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Involved Joints</th>
<th>Risk of OA</th>
<th>References*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumatic driller</td>
<td>Shoulder, elbow, wrist, MCP joint</td>
<td>Increased/none</td>
<td>Jurmain (1977) (cited in ref 47), Burke et al48 (1977)</td>
</tr>
<tr>
<td>Dockworker</td>
<td>Intervertebral disc, DIP joint, elbow, knee</td>
<td>Increased</td>
<td>Lawrence39 (1955)</td>
</tr>
<tr>
<td>Cotton mill worker</td>
<td>Hand</td>
<td>Increased</td>
<td>Lawrence41 (1961)</td>
</tr>
<tr>
<td>Diamond worker</td>
<td>Hand</td>
<td>Increased</td>
<td>Kellgren and Lawrence45 (1957), Tempelaar and Van Breeman46 (1932)</td>
</tr>
<tr>
<td>Shipyard laborer</td>
<td>Knee</td>
<td>Increased</td>
<td>Goldberg and Montgomery47 (1987) (cited in Felson48,49)</td>
</tr>
<tr>
<td>Foundry worker</td>
<td>Lumbar spine</td>
<td>Increased</td>
<td>Lawrence et al46 (cited in Felson48,49)</td>
</tr>
<tr>
<td>Textile worker</td>
<td>Hand</td>
<td>Increased</td>
<td>Tempelaar and Van Breeman48 (1932)</td>
</tr>
<tr>
<td>Manual laborer</td>
<td>MCP joint</td>
<td>Increased</td>
<td>Hadler et al50 (1978)</td>
</tr>
<tr>
<td>Occupations requiring knee bending</td>
<td>Knee</td>
<td>Increased</td>
<td>Williams et al49 (1987) (cited in Felson48,50)</td>
</tr>
</tbody>
</table>


DIP, distal interphalangeal; MCP, metacarpophalangeal; OA, osteoarthritis.

### RECREATION- AND SPORTS-RELATED MUSCULOSKELETAL DISORDERS

Do recreational or sports-related activities lead to musculoskeletal disorders? It has been suggested that the risk of joint degeneration is increased by participation in sports that have high impact levels with torsional loading. The presence of prior joint injury, surgery, arthritis, joint instability and/or malalignment, neuromuscular disturbances, and muscle weakness also predisposed to higher risks of joint damage during sports participation. Patients with sports injuries (such as from downhill skiing and football) to the anterior cruciate and medial collateral ligaments frequently developed the chondromalacia patellae and radiographic abnormalities of OA (20% to 52%). Retrospective studies suggest that the development of OA may be associated with varus deformity, previous meniscectomy, and relative body weight. Both partial and total meniscectomies have been associated with degenerative changes. Early joint stabilization and direct meniscus repair surgery may decrease the incidence of premature OA. These observations supported the concept that abnormal biomechanical forces, either congenital or secondary to joint injury, are important factors in the development of exercise-related OA. Other factors considered important in the development of sports-related OA included certain physical characteristics of the participant, biomechanical and biochemical factors, age, gender, hormonal influences, nutrition, characteristics of the playing surface, unique features of particular sports, and duration and intensity of exercise participation, as has been reviewed extensively elsewhere.

### Table 35-4: Other Occupation-Related Rheumatic Diseases

<table>
<thead>
<tr>
<th>Disease or Syndrome</th>
<th>Occupation or Risk Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflex sympathetic dystrophy</td>
<td>Trauma</td>
</tr>
<tr>
<td>Raynaud’s phenomenon</td>
<td>Vibration</td>
</tr>
<tr>
<td>Autoimmune disease</td>
<td>Chemicals (polyvinyl chloride)</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>Teaching school</td>
</tr>
<tr>
<td>Scleroderma-like syndromes</td>
<td>Chlorinated hydrocarbons</td>
</tr>
<tr>
<td>Systemic lupus erythematous</td>
<td>Organic solvents</td>
</tr>
<tr>
<td>Lupus, scleroderma, and Paget’s disease</td>
<td>Silica</td>
</tr>
<tr>
<td>Rheumatoid arthritis (Caplan’s syndrome)</td>
<td>Pet ownership</td>
</tr>
<tr>
<td>Gout (saturnine)</td>
<td>Rabbits</td>
</tr>
</tbody>
</table>

Rabbits with experimentally induced arthritis in one hind limb did not develop progressive OA when exercised on treadmills, but sheep in normal health walking on concrete did develop OA. Other studies found that beagle dogs running 4 to 20 km a day did not develop OA. Although...
these observations were not entirely consistent, they suggested that physical activities in some circumstances might predispose to degenerative joint disease.

There have been some pertinent observations in human studies (Table 35-5). Wrestlers were reported to have an increased incidence of OA of the lumbar spine, cervical spine, and knee; boxers, of the carpometacarpal joints; parachutists, of knees, ankles, and spine, which was not confirmed; cyclists, of the patella; cricketers, of the fingers; basketball and volleyball, of the knee; athletes involved in sports requiring repetitive overhead throwing such as baseball, tennis, volleyball players, and swimmers, of early glenohumeral arthritis; meniscal and anterior cruciate ligament injuries incurred in youth-related sports, of knee osteoarthritis; soccer players of talar joint, ankle, cervical spine, knee, and hip OA.33-35,96 Studies of American football players have suggested that they are susceptible to OA of the knees, particularly those who sustained knee injuries while playing football.55 Among football players (average age, 23 years) competing for a place on a professional team, 90% had radiographic abnormalities of the foot or ankle, compared with 4% of an age-matched control population; linemen had more changes than did ball carriers or linebackers, who in turn had more changes than did flankers or defensive backs. All those who had played football for 9 years or longer had abnormal findings on radiography.33-37

Most of these studies suffered in several respects: criteria for OA (or “osteoarthrosis,” “degenerative joint disease,” or these observations were not entirely consistent, they suggested that physical activities in some circumstances might predispose to degenerative joint disease.

There have been some pertinent observations in human studies (Table 35-5). Wrestlers were reported to have an increased incidence of OA of the lumbar spine, cervical spine, and knee; boxers, of the carpometacarpal joints; parachutists, of knees, ankles, and spine, which was not confirmed; cyclists, of the patella; cricketers, of the fingers; basketball and volleyball, of the knee; athletes involved in sports requiring repetitive overhead throwing such as baseball, tennis, volleyball players, and swimmers, of early glenohumeral arthritis; meniscal and anterior cruciate ligament injuries incurred in youth-related sports, of knee osteoarthritis; soccer players of talar joint, ankle, cervical spine, knee, and hip OA.33-35,96 Studies of American football players have suggested that they are susceptible to OA of the knees, particularly those who sustained knee injuries while playing football.55 Among football players (average age, 23 years) competing for a place on a professional team, 90% had radiographic abnormalities of the foot or ankle, compared with 4% of an age-matched control population; linemen had more changes than did ball carriers or linebackers, who in turn had more changes than did flankers or defensive backs. All those who had played football for 9 years or longer had abnormal findings on radiography.33-37

Most of these studies suffered in several respects: criteria for OA (or “osteoarthrosis,” “degenerative joint disease,” or

### Table 35-5  Sports Participation and Alleged Associations with Osteoarthritis

<table>
<thead>
<tr>
<th>Sport</th>
<th>Site (Joint)</th>
<th>References*</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ballet</td>
<td>Ankle</td>
<td>Ottani and Betti (1953), Coste et al (1960), Brodelius (1961), Miller et al (1975)</td>
<td>Probably increased</td>
</tr>
<tr>
<td></td>
<td>Hip</td>
<td>Washington (1978)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Knee</td>
<td>Washington (1978)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metatarsophalangeal</td>
<td>Washington (1978)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shoulder</td>
<td>Bennett (1941)</td>
<td></td>
</tr>
<tr>
<td>Boxing</td>
<td>Hand (carpometacarpal joints)</td>
<td>Iselin (1960)</td>
<td></td>
</tr>
<tr>
<td>Cricket</td>
<td>Finger</td>
<td>Vere Hodge (1971)</td>
<td></td>
</tr>
<tr>
<td>Cycling</td>
<td>Finger</td>
<td>Bagneres (1967)</td>
<td></td>
</tr>
<tr>
<td>American football</td>
<td>Ankle</td>
<td>Vincelette et al (1972)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Foot</td>
<td>Rall et al (1964)</td>
<td></td>
</tr>
<tr>
<td>Gymnastics</td>
<td>Elbow</td>
<td>Bozdech (1971)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shoulder</td>
<td>Murray and Duncan (1971)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrist</td>
<td>Murray and Duncan (1971)</td>
<td></td>
</tr>
<tr>
<td>Lacrosse</td>
<td>Ankle</td>
<td>Thomas (1971)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Knee</td>
<td>Murray and Duncan (1971)</td>
<td></td>
</tr>
<tr>
<td>Martial arts</td>
<td>Spine</td>
<td>Rubens-Duval et al (1960)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ankle</td>
<td>Murray and Duncan (1971)</td>
<td></td>
</tr>
<tr>
<td>Parachuting</td>
<td>Spine</td>
<td>Murray-Leslie et al (1977a)</td>
<td></td>
</tr>
<tr>
<td>Rugby</td>
<td>Knee</td>
<td>Slocum (1960)</td>
<td></td>
</tr>
<tr>
<td>Weightlifting</td>
<td>Spine</td>
<td>Brodelius (1961), Solonen (1966)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Talofibular</td>
<td>Burel et al (1960)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Elbow</td>
<td>Layani et al (1960)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Knee</td>
<td>Layani et al (1960)</td>
<td></td>
</tr>
</tbody>
</table>

“abnormality”) were not always clear, specified, or consistent; duration of follow-up was often not indicated or was inadequate to determine the risk of musculoskeletal problems at a later age; intensity and duration of physical activity were variable and difficult to quantify; selection bias toward individuals exercising or participating versus those not exercising or participating was not weighted; other possible risk factors and predispositions to musculoskeletal disorders were rarely considered; studies were not always properly controlled, and examinations were not always “blind”; little information regarding nonprofessional, recreational athletes was available; and little clinical information about functional status was provided.

Several studies have examined a possible relationship between running and OA. Uncontrolled observations generally suggested that runners without underlying biomechanical problems of the lower extremity joints did not develop arthritis at a different rate from a normal population of nonrunners. However, those individuals who had underlying articular biomechanical abnormalities from a previously injured joint (and perhaps elite athletes, particularly women) did appear to be at greater risk for the subsequent development of OA. Early studies showed that groups of long-duration, high-mileage runners and nonrunning control subjects had a comparable (and low) prevalence of OA and suggested that recreational running need not lead inevitably to OA. These observations have generally now been confirmed by others (Table 35-6). Eight- and 9-year follow-up observations were supportive; most of the original runners were still running, with a prevalence of OA that was comparable with that of the control subjects. Perhaps even more significant was the growing evidence

### Table 35-6  Studies of Running and Risk of Developing Osteoarthritis

<table>
<thead>
<tr>
<th>References</th>
<th>No. of Runners</th>
<th>Mean Age (yr)</th>
<th>Mean No. of Years Running</th>
<th>Miles/Wk</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor et al (1989) (cited in refs 33-35)</td>
<td>319</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>OA noted more frequently in former runners (with underlying anatomic “tilt” abnormality—epiphysiolysis) than in nonathletes</td>
</tr>
<tr>
<td>Puranen et al (1975)</td>
<td>74</td>
<td>56</td>
<td>21</td>
<td>NA</td>
<td>Champion distance runners had no more hip OA than did nonrunners in their sixth decade</td>
</tr>
<tr>
<td>De Carvalho and Langfeldt (1977)</td>
<td>32</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>X-ray findings of runners’ hips and knees were similar to those of control subjects</td>
</tr>
<tr>
<td>Marti et al (1990)</td>
<td>20</td>
<td>35</td>
<td>13</td>
<td>48</td>
<td>OA occurred in runners with underlying anatomic (biomechanical) abnormality</td>
</tr>
<tr>
<td>Sohn and Micheli (1985)</td>
<td>504</td>
<td>57</td>
<td>9-15</td>
<td>18-19</td>
<td>No association between moderate long-distance running and future development of OA (of hip and knees)</td>
</tr>
<tr>
<td>Panush et al (1986)</td>
<td>17</td>
<td>53</td>
<td>12</td>
<td>28</td>
<td>Comparable low prevalence of lower-extremity OA in runners and nonrunners</td>
</tr>
<tr>
<td>Lane et al (1986)</td>
<td>41</td>
<td>58</td>
<td>9</td>
<td>(5 hr/wk)</td>
<td>No differences between runners and control subjects in cartilage loss, crepitus, joint stability, or symptoms</td>
</tr>
<tr>
<td>Lane et al (1987) (cited in refs 33-35)</td>
<td>498</td>
<td>59</td>
<td>12</td>
<td>27</td>
<td>No differences between groups in conditions thought to predispose to OA and musculoskeletal disability</td>
</tr>
<tr>
<td>Marti et al (1989, 1990)</td>
<td>27</td>
<td>42</td>
<td>NA</td>
<td>61 (in reference years)</td>
<td>More radiographic changes of hip OA in former Swiss national team long-distance runners than in bobsledders and control subjects; few runners had clinical symptoms of OA; no difference in ankle joints</td>
</tr>
<tr>
<td>Konradsen et al (1990)</td>
<td>30</td>
<td>58</td>
<td>40</td>
<td>12-24</td>
<td>No clinical or radiographic differences in hips, knees, and ankles between runners and nonrunners</td>
</tr>
<tr>
<td>Vingard et al (1995)</td>
<td>114</td>
<td>50-80</td>
<td>NA</td>
<td>NA</td>
<td>Unvalidated questionnaire reported threefold increase of hip arthrosis in former athletes</td>
</tr>
<tr>
<td>Kujala et al (1994)</td>
<td>342</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>More former athletes hospitalized with hip OA than expected</td>
</tr>
<tr>
<td>Kujala et al (1995)</td>
<td>28</td>
<td>60</td>
<td>32</td>
<td>NA</td>
<td>Women soccer players and weightlifters, nonrunners were at risk of premature OA</td>
</tr>
<tr>
<td>Panush et al (1995)</td>
<td>16</td>
<td>63</td>
<td>22</td>
<td>22</td>
<td>8-yr follow-up of original observations made in 1986 still found no differences between runners and nonrunners</td>
</tr>
<tr>
<td>Lane et al (1998)</td>
<td>35</td>
<td>60</td>
<td>10-13</td>
<td>23-28</td>
<td>Running did not appear to influence the development of radiographic OA (with possible exception of spur formation in women)</td>
</tr>
</tbody>
</table>

NA, not available; OA, osteoarthritis.
that running and other aerobic exercise protected against the development of disability and early mortality. Former college varsity long-distance runners were compared with former college swimmers in another study; there was no association between moderate levels of running or number of years running and the development of symptomatic OA. Other authors have concluded that running alone does not cause OA; rather, prior injuries and anatomic variances were directly responsible for some of the changes. Prospective studies have found that runners were not at risk for the development of premature OA of the knees.

Studies examining hip OA in former athletes noted that former champion distance runners had no more clinical or radiographic evidence of OA than did nonrunners. However, another study found more radiographic changes due to degenerative hip disease in former national team long-distance runners than in bobsled competitors and control subjects. In all the subjects studied, age and mileage run in 1973 were strong predictors of radiographic hip OA; for runners, running pace in 1973 was the strongest predictor of subsequent radiographic hip OA in 1988. These authors concluded that high-intensity, high-mileage running should not be dismissed as a risk factor for premature OA of the hip. Other reports found that former top-level soccer players and weightlifters, but not runners, were at risk for the development of knee OA, but it was suggested elsewhere that former athletes seemed to be disproportionately represented in hospital admissions for OA of hip, knee, or ankle. A questionnaire of former elite and track-and-field athletes noted they had increased hip OA. Similarly, radiographic OA of the hip and knee was reported in women who were formerly runners and tennis players.

Cross-sectional studies on the effect of weight-bearing exercise on the development of OA of the hip, knee, or ankle and foot must be interpreted with caution, however. The radiographic scoring methods used by each group of investigators differ, and their reliability has not been adequately tested. This information is important when the major end points in the studies are radiographic features of OA.

**PERFORMING ARTS–RELATED MUSCULOSKELETAL DISORDERS**

Musculoskeletal problems are common among performing artists. Performing artists—particularly musicians and dancers—have unique medical and musculoskeletal problems that deserve special consideration. Injuries that might be trivial to others may be catastrophic to such artists. These injuries are usually associated with overuse—the consequences of tissues stressed beyond anatomic or normal physical limits. Understanding the technical requirements and biomechanics required in the performance of a craft, as well as the lifestyle required to pursue a successful career in these fields should help physicians appreciate the causative factors that lead to these injuries.

Important principles in approaching such patients follow: (1) Musculoskeletal problems comprise the bulk of health issues for these individuals. (2) Performing artists are usually wary of consulting with physicians (skeptical of their expertise). (3) An appropriate evaluation should be carried out by someone knowledgeable about the technical and biomechanical requirements of the patient’s craft(s). It should consider instrument(s), instrument usage, travel with instruments, shoes, performance surface and setting, practice and performance routines, repertoire, coaches and training/trainers, and lifestyle and psychological factors, as appropriate. (4) Evaluation should include attention to joint laxity and other physical features of the artist, as well as to their relationship to performance, considering those entities encountered as listed in Table 35–7. It should assess muscle tension and fatigue. Patients should demonstrate how they use an instrument while both the actively moving body parts and the relatively immobilized parts are examined. (5) There should be inquiry about all prescription and nonprescription therapies, nutritional and exercise practices, and nonmainstream treatments. (6) There must be understanding and sympathy for the unique expectations of these performers and expertise in assessing their medical problems and developing treatment plans. (7) Prevention should be emphasized—assuring performance ability, promoting endurance and conditioning, facilitating good posture, protecting joints, maintaining proper ergonomics, and establishing appropriate exercise regimens. (8) Therapeutic interventions will usually be conservative.

**Instrumentalists**

The frequency of musculoskeletal problems in musicians rivals the frequency of disability in athletes. Up to 82% of orchestral musicians have experienced medical problems related to their occupation, mainly musculoskeletal. Up to 76% of musicians have reported a musculoskeletal issue that is grave enough to influence their ability to perform. Woodwind players and female instrumentalists seemed to be affected more compared with other types of other instrumentalists and male artists, respectively. Muscle-tendon overuse or repetitive stress injuries, nerve entrapment problems, and focal dystonias were most common (see Table 35–7).

The causes of, mechanisms of, and therapies for these musculoskeletal problems are unclear. Overuse, tendinitis, cumulative trauma disorder, repetitive motion disorder, occupational cervicobrachial disorder, and regional pain syndrome may be critical risk factors in the development of joint laxity in musicians. Joint laxity declined with age and was associated with gender, starting earlier in men but persisting in women through their mid-40s. The presence or absence of hypermobility at certain sites was associated with musicians’ complaints of associated symptoms. Hypermobility in musicians might produce advantages or disadvantages, depending on the site of the laxity and the instrument played. Paganini, with his long fingers and reported hyperextensibility, had a wider finger reach on the violin than his contemporaries, but he may have had a predisposition to OA because of this. Of interest and seemingly unexplained was the high frequency of symptoms among women (68% to 84%); perhaps this is related to their higher incidence of hypermobility.

Stress is a factor in all performance fields and contributes to motor function problems such as occupational cramps; dealing with this problem often requires the best efforts of a team of physicians and therapists.
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Affliction (Common Name)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Epicondylitis</td>
<td>Fry (1986b), Hiner et al (1987)</td>
</tr>
<tr>
<td></td>
<td>Rotator cuff tears</td>
<td>Fry (1986b), Newmark and Hochberg (1987)</td>
</tr>
<tr>
<td></td>
<td>Garrod’s pads</td>
<td>Bird (1987)</td>
</tr>
<tr>
<td></td>
<td>Nerve entrapment</td>
<td>Knishkowy and Lederman (1986)</td>
</tr>
<tr>
<td></td>
<td>Interosseous</td>
<td>Maffulli and Maffulli (1991)</td>
</tr>
<tr>
<td></td>
<td>Ulnar</td>
<td></td>
</tr>
<tr>
<td>Cello</td>
<td>Myalgias</td>
<td>Fry (1986b)</td>
</tr>
<tr>
<td></td>
<td>Tendinitis</td>
<td>Caldron et al (1986), Fry (1986b)</td>
</tr>
<tr>
<td></td>
<td>Epicondylitis</td>
<td>Fry (1986b)</td>
</tr>
<tr>
<td></td>
<td>Low back pain</td>
<td>Fry (1986b)</td>
</tr>
<tr>
<td>Bass</td>
<td>Low back pain</td>
<td>Schott (1983)</td>
</tr>
<tr>
<td></td>
<td>Myalgias</td>
<td>Fry (1986b)</td>
</tr>
<tr>
<td></td>
<td>Motor palsies</td>
<td>Caldron et al (1986)</td>
</tr>
<tr>
<td>Harp</td>
<td>Tendinitis</td>
<td>Caldron et al (1986)</td>
</tr>
<tr>
<td>Woodwinds</td>
<td>First web space muscle strain</td>
<td>Fry (1986b), Newmark and Hochberg (1987)</td>
</tr>
<tr>
<td>Clarinet and obo</td>
<td>Motor palsies</td>
<td>Jankovic and Shale (1989)</td>
</tr>
<tr>
<td>Flute</td>
<td>Myalgias</td>
<td>Fry (1986b)</td>
</tr>
<tr>
<td></td>
<td>Spine pain</td>
<td>Fry (1986b)</td>
</tr>
<tr>
<td></td>
<td>Temporomandibular joint syndrome</td>
<td>La France (1985)</td>
</tr>
<tr>
<td></td>
<td>Tendinitis</td>
<td>Patrone et al (1988)</td>
</tr>
<tr>
<td></td>
<td>Nerve entrapment</td>
<td>Cynamon (1981)</td>
</tr>
<tr>
<td></td>
<td>Digital</td>
<td>Charness et al (1985)</td>
</tr>
<tr>
<td></td>
<td>Posterior interosseous</td>
<td></td>
</tr>
<tr>
<td>English horn</td>
<td>de Quervain’s tenosynovitis</td>
<td>Studman and Milberg (1982)</td>
</tr>
<tr>
<td>Percussion</td>
<td>Thoracic outlet syndrome</td>
<td>Lederman (1987)</td>
</tr>
<tr>
<td>Saxophone</td>
<td>Osteoarthritis</td>
<td>Caldron et al (1986)</td>
</tr>
<tr>
<td>Cymbals</td>
<td>Nerve entrapment</td>
<td>Fry (1986b)</td>
</tr>
<tr>
<td></td>
<td>Bicipital tenosynovitis (cymbal player’s shoulder)</td>
<td>Makin and Brown (1985)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Huddleston and Pratt (1983)</td>
</tr>
</tbody>
</table>
Table 35-7  Musculoskeletal and Rheumatic Disorders Associated with Overuse in Performing Artists—cont’d

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Affliction (Common Name)</th>
<th>References*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miscellaneous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guitar, strings</td>
<td>Tendinitis</td>
<td>Newmark and Hochberg (1987)</td>
</tr>
<tr>
<td></td>
<td>Synovitis</td>
<td>Mortonroth (1978), Bird and Wright (1981)</td>
</tr>
</tbody>
</table>


Vocal Artists

Musculoskeletal problems among singers have not been addressed extensively. In a report from the Royal Theater in Copenhagen, the frequency of musculoskeletal problems was the same in both instrumentalists and opera singers. However, singers had more hip, knee, and foot joint complaints, perhaps reflecting the effects of prolonged standing.119

Dancers

Dance has always been viewed as a demanding art form, but only recently have the athletic rigors of this discipline become widely appreciated. Classic ballet ranked first in activities generating physical and mental stress, followed by professional football and professional hockey. The dancer and athlete have much in common, but there are important differences in training and performance technique that influence the nature of their injuries. Other important sociocultural differences affect their care. Professional dancers (as well as musicians and vocalists) have traditionally been unaccustomed to what physicians know how to effectively approach the unique issues of dance (and music). Injured dancers seeking care have often been told that the treatment is to stop dancing. Others, seeking assistance with weight control, have been told to gain weight. Dancers frequently underreport their injuries and seek care from nonmedical therapists.

The incidence of dance-related injuries ranges from 17% to 95%.121 The majority of injuries involved the foot, ankle, and knee. It is difficult to generalize about dance injuries because dance is not a monolithic effort. It is a broad-based hierarchic endeavor in which thousands of local school-based and private amateur dance classes supply a much smaller number of university-based dance programs, which lead finally to relatively few professional dance companies. This system of training encompasses many forms of dance that are highly divergent, ranging from classic ballet to break dancing. Fortunately, most injuries are from overuse and are rarely catastrophic, regardless of the dance style or setting. As with other overuse injuries in sports, they are influenced by a variety of factors that may be classified as intrinsic such as biomechanical and anatomic variations or extrinsic such as those related to occupation or equipment.117

The distribution of injuries is strongly influenced by the type and style of dance and the age and sex of the population.118,122 A better understanding of the technical and aesthetic requirements of a dance, as well as the biomechanics involved to perform these requirements, are necessary in order to appreciate the kind of injuries that can be sustained by dancers. For example, ballet dancers in companies whose choreography emphasizes bravura technique with big jumps and balances are more likely to develop Achilles tendinitis than are those in companies that do not. Men are more likely to have back injuries because of the requisite jumping and lifting, whereas women who dance on pointe are more prone to toe, foot, and ankle problems. Also in ballet, the most important physical feature is proper turnout of the hip, which requires maximal external rotation of the lower extremity that can result in hyperlordosis of the lumbar spine, valgus heel with forefoot pronation, and external rotation of the knee.121,122

Tendinitis of the flexor hallucis longus tendon, commonly known as dancer’s tendinitis, may be confused with posterior tibial tendinitis due to the location of pain at the posteromedial ankle.124 Other dancer- and environment-related factors that increase the risk of dance-related injuries include nutritional status, improper support from foot-wear and floors, and their rehearsal and performance schedules.119,123 Most dance shoes rarely have a shock-absorbing sole, and some dances may be done barefoot.123 Traditionally constructed with paper, glue, and satin or canvas or leather, ballet pointe shoes tend to soften once broken in, thus contributing to ankle injury. Intensive rehearsals before and during the opening months of a performance season, pressures to return to work quickly after an injury, and the “show must go on” mentality must also be considered in the care of dancers.117,123 Touring companies may encounter nonflexible surfaces including concrete, predisposing to shin splints and stress fractures. Stress fractures may be associated with the pressure to maintain a certain weight, which may result in amenorrhea, disordered eating, and low bone density. Physicians caring for dancers, particularly ballet dancers at any level, must be aware of the aesthetic pressures for extreme leanness and the potential consequences. Unfortunately, the dance world is not lacking in other serious medical problems including mental illness, drug abuse, and human immunodeficiency virus (HIV) infection.117

References


The references for this chapter can also be found on www.expertconsult.com.
For nearly half a century, excess rates of cardiovascular disease (CVD) have been reported among patients with inflammatory rheumatic diseases. More recently, the discovery of the inflammatory and immune mechanisms underlying atherosclerosis has spurred renewed interest in the association between CV risk and the rheumatic diseases. In this chapter we review the risks of cardiovascular comorbidity in the rheumatic diseases, focusing on rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). We also discuss the contribution of traditional and nontraditional cardiovascular risk factors to the observed excess CVD risk.

**CARDIOVASCULAR MORTALITY**

**Rheumatoid Arthritis**

The mortality of patients with established RA is known to be higher than that of the general population. Approximately 50% of all deaths in RA subjects are attributable to cardiovascular causes including ischemic heart disease (IHD) and stroke, and CVD appears to occur earlier in individuals with RA. The latter observation is consistent with the recent hypothesis of accelerated aging in RA in general. More than 50% of premature deaths in RA are due to CVD. A meta-analysis of 24 mortality studies in RA, published between 1970 and 2005, reported a weighted combined all-cause standardized mortality ratio (met-SMR) of 1.50 (95% confidence interval [CI],) with similar increases for IHD (met-SMR, 1.59; 95% CI, 1.46 to 1.73) and stroke (met-SMR, 1.52; 95% CI, 1.40 to 1.67); and for men (met-SMR, 1.45) and women (met-SMR, 1.58). Moreover, patients with RA frequently experience “silent” IHD and/or silent myocardial infarction (MI), showing no symptoms at all before a sudden cardiac death. Sudden cardiac deaths are almost twice as common in RA patients as in the general population (hazard ratio [HR], 1.99; 95% CI, 1.06 to 3.55).

The excess CV mortality in RA may be confined to, or at least substantially higher in, subjects who are rheumatoid factor (RF) positive. The link may be even stronger with anticitrullinated protein antibody (ACPA) positivity. As might be expected, the relative risk of CV mortality is highest in younger age groups (those younger than 55 in whom controls have lower absolute risk and therefore in whom the distinction from those with RA is likely to be exacerbated) and in women, while the attributable risk is highest in the oldest age groups and in men.

Controversy persists regarding how soon after symptom onset the excess CV mortality risk becomes apparent and/or whether there is a secular trend toward improving CV mortality in RA (as is seen in the general population). This may be partially explained by differences in the period of follow-up (i.e., follow-up starting from the time of symptom onset, from a physician’s diagnosis of RA, from the date of fulfillment of American College of Rheumatology (ACR) criteria, or other diagnostic criteria). The latter may not occur until some years after the first symptoms. In the Norfolk Arthritis Register (NOAR) the excess CV mortality is detectable beginning around 7 years after symptom onset. In a Dutch inception cohort of 1049 RA patients recruited between 1985 and 2007, excess mortality became apparent around 10 years after diagnosis (with all subjects having <1-year symptom duration). Unfortunately, there appears to be little evidence that CV mortality has improved significantly. A meta-analysis that included 17 studies (91,916 patients) reported CV mortality risk between 1976 and 2007, showing no trend toward improving CV SMR with time.

**Systemic Lupus Erythematosus**

As first described by Urowitz and colleagues in 1976, mortality in SLE appears to follow a bimodal pattern, with an early peak in mortality (within 1 year of diagnosis) and a...
later peak occurring >5 years after diagnosis). Reported survival in the first 5 years of SLE has improved considerably from about 50% in the 1950s to more than 90% in the 1990s. However, there is some question as to whether this may be due, at least in part, to earlier diagnosis and improved ascertainment of mild cases. In the Toronto lupus cohort of 1241 patients recruited between 1970 and 2005, the SMR improved from 13.84 (range, 9.78 to 19.76) during 1970 to 1978 for those who entered the cohort in that decade to 3.81 (range, 1.98 to 7.32) during 1997 to 2005 in those who entered the cohort in that time period. The SMR during 1997 to 2005 was similar for patients regardless of their disease duration, ranging from 3.23 for those who had entered the cohort in 1970 to 1978 to 3.93 for those who had entered the cohort in 1988 to 1996. Similarly, evidence from Olmsted County, Minnesota showed an SMR of 2.70 with significant improvement in survival in recent decades. In a study of 434 female lupus patients from Seoul, Korea followed from 1992 to 2002, the SMR was 3.02 (95% CI, 1.45 to 5.55). No study has been large enough to permit study of cause-specific SMR.

CARDIOVASCULAR COMORBIDITY

Rheumatoid Arthritis

Ischemic Heart Disease

Patients with RA are at increased risk of IHD. Data from the Rochester Epidemiology Project have shown that, in the 2-year period immediately preceding the fulfillment of the ACR 1987 criteria, RA patients were more likely to experience hospitalization for MI (odds ratio [OR], 3.17; 95% CI, 1.16 to 8.68) and unrecognized (“silent”) MI (OR, 5.86; 95% CI, 1.29 to 26.64) than age- and sex-matched controls. The increased risk of unrecognized MI persisted after the diagnosis of RA (HR, 2.13; 95% CI, 1.13 to 4.03). Holmqvist and colleagues failed to demonstrate a statistically significant elevated increase in MI, angina, or heart failure before the onset of symptoms in two large Swedish cohorts, although trends toward such elevation were reported. As in studies of mortality, these results suggest that accelerated atherosclerosis begins at the onset of RA symptoms, or even earlier, and not at the time of diagnosis or later in the disease course.

Patterns of clinical care and outcome after MI may vary in persons with RA when compared with the general population. Some evidence suggests that although RA patients receive similar MI care to non-RA patients, they experience higher rates of heart failure and death after MI. However, others have recently reported that RA patients who experience acute MI receive acute reperfusion and secondary prevention medications (including β-blockers and lipid-lowering agents) less frequently than controls. Moreover, another group recently reported that among patients with MI, those with RA were more likely to undergo thrombolysis and percutaneous coronary intervention (PCI) but less likely to receive medical therapy or coronary artery bypass grafting, or both. That study also suggested that patients with RA may have an in-hospital survival advantage, particularly those undergoing medical therapy and PCI, though potential confounding could not be ruled out.

Heart Failure

Patients with RA are also at increased risk of developing heart failure compared with the general population. In the Rochester RA cohort, the cumulative incidence of congestive heart failure (CHF; defined according to the Framingham criteria) at 30-year follow-up was 34% compared with 25% in the non-RA cohort (Figure 36-1; P < 0.001). Even after adjustment for demographics, CV risk factors, and IHD, patients with RA had almost twice the risk of developing CHF as non-RA subjects (HR, 1.87; 95% CI, 1.47 to 2.39). This increased risk of CHF appeared to be predominantly in the subgroup of RA patients who were RF positive (HR in RF-positive patients, 2.59; 95% CI, 1.95 to 3.43 vs. 1.28 in RF-negative patients; 95% CI, 0.93 to 1.78).

The clinical presentation of CHF in patients with RA differs from CHF in non-RA patients. RA patients with CHF are less likely to be obese, be hypertensive, or have clinical IHD. Moreover, RA patients with CHF are less likely to have typical signs and symptoms. Importantly, the proportion of CHF patients with preserved ejection fraction (>50%) is significantly higher among RA compared with non-RA patients. It has also been shown that RA patients with CHF tend to be investigated and managed less aggressively. Finally, RA patients with CHF also appeared to have poorer outcomes, experiencing approximately twice the risk for death in the period immediately after detection of heart failure compared with non-RA patients.

Systemic Lupus Erythematosus

Ischemic Heart Disease

Accelerated atherosclerosis is an established complication of SLE. The prevalence of atherosclerotic vascular events varies from 1.8% in early disease to more than 27%
later in the course of SLE. The majority of studies reported increased risk of MI ranging from 2- to greater than 10-fold in various SLE patient groups compared with the general population. This increased relative risk of MI is particularly apparent in younger SLE patients. The most striking example comes from the University of Pittsburgh lupus cohort, where women with SLE aged 35 to 44 were more than 50 times as likely to have an MI compared with women without SLE in the Framingham Offspring study (relative risk [RR], 52.43; 95% CI, 21.6 to 98.5). Furthermore, the majority (67%) of women with SLE were younger than 55 years of age at the time of their first cardiac event. In addition, a greater than twofold increased risk of hospitalizations for MI has been reported in young women with SLE between 18 and 44 years of age compared with those without SLE (OR, 2.27; 95% CI, 1.08 to 3.46).

A recent study of patients undergoing coronary revascularization procedures found no significant differences in the mean percent of coronary stenosis and total occlusion in SLE versus non-SLE subjects. Except for the increased likelihood of lesions confined to the left anterior descending artery in SLE versus non-SLE subjects (42.3% vs. 19.3%; P = 0.003), the pattern of coronary involvement including artery dominance and prevalence of multivessel disease appeared similar in SLE versus non-SLE subjects. However, the study reported significantly worse cardiovascular outcomes at 1 year following PCI in SLE versus non-SLE subjects including higher risk of MI (15.6% vs. 4.8%; P = 0.01), and repeat PCI (31.3% vs. 11.8%; P = 0.009) in SLE, even after adjustment for important co-variates. Given increased vulnerability of atherosclerotic plaque in SLE, which is associated with the risk of occlusive events irrespective of size of the plaque, these findings suggest an increased risk of unfavorable cardiovascular events in SLE patients versus non-SLE subjects with a seemingly similar pattern of coronary involvement.

In a large population-based study of patients hospitalized in California with acute MI from 1996 to 2000, in-hospital mortality and length of stay were essentially similar in patients with SLE compared with those who did not have SLE adjusting for age, race, ethnicity, type of medical insurance, and Charlson Index. In contrast, data from the 1993 to 2002 U.S. Nationwide Inpatient Sample showed significantly increased rates of in-hospital mortality (RR, 1.46; 95% CI, 1.31 to 1.61) and prolonged hospitalization (RR, 1.68; 95% CI, 1.43 to 2.04) for acute MI in SLE compared with controls, adjusting for age, sex, race/ethnicity, income, and CHF. Some differences in methodology (i.e., smaller sample size and older age of SLE patients and control subjects in the earlier study and shorter observation period in the later study) may at least in part explain these contrasting results. Considering the presence of myocardial involvement, chronic systemic inflammation, vasculitis, and hyperviscosity syndrome in SLE, worse outcomes following acute coronary events and associated interventions can reasonably be expected. However, this requires further study.

Heart Failure

The risk of CHF and related hospitalization in SLE appears to be substantially increased. In particular, young women with SLE between 18 and 44 years of age have a greater than 2.5-fold increased risk of hospitalization for CHF versus those who did not have SLE, even after adjustment for age, race, insurance status, hospital characteristics, and the presence of hypertension, diabetes mellitus, and chronic renal failure. Consequently, CHF accounted for a substantially higher percent of hospitalizations in women with SLE versus those without SLE within this age group (1.32% vs. 0.35%, respectively; P < 0.0001). The nature of CHF in SLE is likely multifactorial, only partly attributable to atherosclerosis. The presentation of CHF in SLE may vary from severe overt CHF to insidious myocardial involvement. Finally, mortality in SLE patients with CHF is significantly higher than in those without CHF (17.9% vs. 5.8%; P < 0.001) and approximates 3.5-fold as compared with the general population.

TRADITIONAL RISK FACTORS FOR CARDIOVASCULAR DISEASE

The role of traditional risk factors in the development of CVD in persons with RA and SLE is an area of active research. One possible explanation for the increase in CVD seen in RA and SLE could be that the traditional CVD risk factors are more common in these diseases or that they are as common but are more deleterious. Alternatively, the excess CVD risk may be explained by the adverse impact of the inflammation and immune changes of RA and SLE on the vessel wall. In fact, the polarization of these possible explanations is oversimplistic because traditional risk factors and inflammation are intimately interconnected and may act synergistically.

Population studies have elucidated the role of a number of “traditional” risk factors including increasing age, male gender, smoking, hypertension, hypercholesterolemia, and diabetes in the development of CVD in persons with RA and SLE. These factors have been combined into a number of scores for estimating the risk of future CVD events in the general population. The most well-known of these is the Framingham risk score, which provides an estimate of the 10-year risk of a future CVD event for an individual subject. More recently, body composition (in particular, visceral adiposity) and physical inactivity have also been identified as traditional risk factors for CVD for use in the general population.

Traditional Cardiovascular Risk Factors

Smoking in Rheumatoid Arthritis

Smoking is a known risk factor for the development of RA, in particular RF and ACPA-positive RA. Thus as expected, there is a higher prevalence of current smokers and ex-smokers among subjects with RA than among the general population. In a meta-analysis of four case-control studies (1415 RA patients) of traditional CVD risk factors in RA, the prevalence of smoking was found to be significantly higher than in controls (OR, 1.56; 95% CI, 1.35 to 1.80). Smokers with RA also appear to have a worse prognosis in terms of RF titers, disability, radiographic damage, and treatment response. There is a known interaction among
smoking, human leukocyte antigen (HLA) DR1 shared epitope (SE) alleles, and the production of ACPA and among smoking, ACPA, and the SE in premature CVD mortality in RA.

**Smoking in Systemic Lupus Erythematosus**

In a case-control study involving 250 women with SLE, smoking was no more common than in the general population (RR, 0.86; 95% CI, 0.59 to 1.24). Nevertheless, smoking was a risk factor for vascular events in the cohort overall and in the inception subcohort.

**Hypertension in Rheumatoid Arthritis**

Hypertension is common in patients with RA, but it remains unclear whether it is more common than in the general population. A recent meta-analysis of seven case-control studies (1053 RA patients) found the prevalence of hypertension to be the same in RA patients as in controls (OR, 1.09; 95% CI, 0.91 to 1.31). There is, however, some evidence for underdiagnosis and undertreatment of hypertension in RA patients. Multiple other factors may influence blood pressure control in persons with RA including physical inactivity, obesity, specific genetic polymorphisms, and several antirheumatic medications including nonsteroidal anti-inflammatory drugs (NSAIDs), corticosteroids, leflunomide, and cyclosporine.

**Hypertension in Systemic Lupus Erythematosus**

Hypertension is substantially more common in women with lupus than in the general population. A case-cohort study found the relative risk of hypertension in women to be 2.59 (95% CI, 1.79 to 3.75). Hypertension has been noted to be a predictor of mortality and vascular events in lupus in a number of studies. Hypertension was found to be a predictor of vascular events in the total Toronto cohort of 1067 patients but not in the 561 patients in the inception cohort.

**Dyslipidemia in Rheumatoid Arthritis**

Although findings have been somewhat inconsistent, hyperlipidemia appears to have a paradoxical relationship with CV risk in persons with RA, namely decreased lipid levels associated with increased CV risk. Serum levels of total cholesterol and low-density lipoprotein (LDL) cholesterol decline precipitously during the 3- to 5-year period before RA incidence, and lower total and LDL cholesterol levels have been shown to be associated with higher CV risk. Suppression of total and LDL cholesterol levels during acute or chronic high-grade inflammation is well described, as is a proportionately greater suppression of high-density lipoprotein (HDL) cholesterol, resulting in a disadvantageous atherogenic index (total-to-HDL cholesterol ratio). This may explain the fact that hyperlipidemia (high total or LDL cholesterol) appears to be less common in RA compared with non-RA subjects. Dyslipidemia (alterations of individual lipid components and their ratios as defined by specific criteria) may affect up to half of all RA patients in hospital care. A recent meta-analysis showed that RA is associated with an abnormal lipid pattern, principally low levels of HDL cholesterol. Lipid alterations appear to predate the diagnosis of RA. Serum levels of total cholesterol and LDL cholesterol decline precipitously during the 3- to 5-year period before RA diagnosis. In vitro animal model and human in vivo studies in subjects without RA clearly demonstrate that the interplay between inflammation and lipid components is far more complex than simple alterations of their serum levels (Figure 36-2). For example, acute phase proteins such as serum amyloid A and phospholipid A2 can alter HDL composition and function, whereas inflammation may have profound effects on enzymes fundamental to the metabolism of HDL (e.g., hepatic lipase) or indeed the enzymatic content of HDL itself (e.g., reduced paraoxonase); this may increase susceptibility to oxidation and convert HDL to a more pro-oxidant, pro-atherogenic complex. Such inflammation-induced alterations of structure and function are not confined to HDL but also involve triglycerides and LDL; they require further study, specifically in RA and in the context of disease control through nonbiologic and biologic disease-modifying antirheumatic drugs (DMARDs).

To date, several studies suggest antirheumatic therapy mediates effects on lipid levels including glucocorticoids, hydroxychloroquine, gold, cyclosporine, and the biologics anti–tumor necrosis factor (TNF), rituximab, and tocilizumab: these are generally short-term studies of small numbers of patients addressing predominantly serum levels rather than other modifications or mechanisms. Multiple other factors are involved in lipid regulation and function including physical activity, adiposity, diet, alcohol intake, and smoking. However, their effects have not been assessed in any detail in persons with RA. Similarly, the importance of genetic regulation of lipid metabolism,
particularly in the context of gene-environment interactions, has not been addressed in the RA population. This may be particularly important because lipid alterations appear to predate the diagnosis of RA.

It is well known that HDL cholesterol is normally protective against atherosclerosis. However, in some circumstances such as after infection or postoperatively, HDL may become “proinflammatory” by virtue of altered molecular composition. McMahon and colleagues reported that around 20% of women with RA had proinflammatory HDL (piHDL) compared with only 3.1% in the normal population. Unlike normal HDL, piHDL does not protect LDL from oxidation. Lipoprotein (a) levels are also reported to be increased in RA subjects and associated with increased CVD risk. In the Apolipoprotein-related MOrtality RISK (AMORIS) study the risk of MI was 60% higher among the 1779 subjects who had RA compared with the 478,627 subjects without RA. However, although total cholesterol and triglyceride levels were associated with the development of acute MI in subjects without RA, this was not the case in those with RA.

**Dyslipidemia in Systemic Lupus Erythematosus**

As in RA, inflammation may have a deleterious effect on the atherogenic index in SLE. However, although hypercholesterolemia is no more common in SLE than in the general population, when present it is associated with an increased risk of vascular events. Hypercholesterolemia was a risk factor for future cardiovascular events in the total Toronto cohort but not in the inception cohort. Again, this may be a reflection of aggressive management of hyperlipidemia in the inception cohort. In the United Kingdom General Practice Research Database the combination of SLE and hypercholesterolemia was associated with an 18-fold increased risk of MI compared with the general population. As in RA, patients with SLE may have increased levels of proinflammatory HDL.

**Diabetes in Rheumatoid Arthritis**

A recent meta-analysis of seven case-control studies (1230 RA patients) reported that the prevalence of diabetes was increased in comparison with controls (OR, 1.74; 95% CI, 1.22 to 2.50). Abdominal obesity, antihypertensive medication, disease activity, and use of corticosteroids all affect glucose metabolism in RA. On the other hand, use of hydroxychloroquine has been shown to reduce the risk of developing diabetes in RA by 77%.

**Diabetes in Systemic Lupus Erythematosus**

The incidence of diabetes mellitus is substantially increased in women with SLE (RR, 6.00; 95% CI, 1.36 to 26.53). However, diabetes was not a predictor of vascular events in either the entire Toronto cohort or the inception subcohort.52

**Body Composition**

Escalante and colleagues reported a “paradoxical effect of BMI on survival in persons with RA” demonstrating that, as BMI declined, so did survival probability among study subjects with RA. Among persons without RA, low BMI is not associated with increased risk of CV death. However, among RA patients, low BMI, which may indicate uncontrolled active systemic inflammation, is associated with a threefold increased risk of CV death even after adjustment for cardiac history, smoking, diabetes mellitus, hypertension, and malignancy.

Obesity is associated with an increased frequency of traditional CV risk factors in patients with RA as in the general population. In particular, abdominal fat is associated with insulin resistance and inflammatory load in patients with RA and there is new evidence that in such patients, abdominal fat is distributed differently between the visceral and subcutaneous compartments, with visceral fat more strongly associated with cardiometabolic risk. Adipose tissue is metabolically active and, through a network of adipocytokines, regulates not only energy intake and expenditure but also inflammation. Interventions to reverse rheumatoid cachexia, control obesity, and regulate insulin resistance including comprehensive physical exercise programs in RA have been little studied. A sedentary lifestyle is common in patients with SLE.

**Impact of Traditional Cardiovascular Risk Factors in Rheumatoid Arthritis**

The absolute CV risk in RA subjects has been estimated to be similar to that in non-RA subjects who were approximately 10 years older. Further studies suggest that CV disease and CV death in RA are of similar magnitude to that seen in patients with type II diabetes mellitus (Figure 36-3). This information, together with the findings that traditional CV risk factors behave differently in RA subjects, suggest that risk scores based on traditional CV risk factors alone are likely to inaccurately estimate CV risk in RA. Indeed, recent studies have reported that such risk scores (e.g., Framingham) can underestimate CV risk fivefold in some RA patients. All this clearly highlights the need for RA-specific risk prediction tools. To this end, the European League against Rheumatism (EULAR) has recently proposed the application of a ×1.5 multiplier to the risk calculated on the basis of standard algorithms. This approach, although appealing in its pragmatism, requires validation.

**Impact of Traditional Cardiovascular Risk Factors in Systemic Lupus Erythematosus**

On average, persons with SLE and CVD have one less classic risk factor than persons in the general population with CVD. Esailde and colleagues calculated the baseline 10-year CHD and stroke risk using the Framingham score for all lupus patients attending a lupus clinic in Montreal. Even after adjusting for the Framingham score, patients still had a 7.5- to 17-fold increased risk of cardiovascular events. Although lupus patients with vascular events were found to have a higher number of traditional risk factors than those without events, they did not have a higher Framingham risk score. This suggests that the relative importance of the individual risk factor differs between lupus patients and the general population.
Impact of Rheumatoid Arthritis Disease Activity and Severity on Cardiovascular Comorbidity

A number of studies have linked the risk of CV comorbidity with markers of RA disease activity such as baseline C-reactive protein, fast recorded erythrocyte sedimentation rate (ESR), and an ESR greater than or equal to 60 mm/hr on three or more occasions. In a study of 231 male veterans with RA, a baseline disease activity score (DAS28) of 5.1 or greater predicted CV events (HR, 1.3; 95% CI, 1.1 to 1.6). Markers of disease severity such as RF, ACPA, physical disability, destructive changes on joint radiographs, rheumatoid nodules, vasculitis, rheumatoid lung disease, and corticosteroid use are statistically significantly associated with increased risk of CV events and/or death even after adjustment for traditional CV risk factors. Analyses of ESR levels in 172 RA cases with heart failure demonstrated that the proportion with significantly elevated ESR (≥40 mm/hr) was highest during the 6-month period before CHF diagnosis as compared with any other time over the entire follow-up period. RF and antinuclear antibody are risk factors for MI, CHF, and/or vascular disease even in subjects without RA, suggesting that immune dysregulation may promote CV risk not only in persons with rheumatic disease but also in the general population.

Impact of Systemic Lupus Erythematosus Disease Activity and Severity on Cardiovascular Comorbidity

Various markers of inflammation and organ damage have been found to be associated with CV events in SLE including renal, neuropsychiatric, and vasculitic disease. In addition, SLE may be associated with thrombotic tendency and this, too, may contribute to the burden of CV disease.

Medications as Cardiovascular Risk Factors

Medications used for the treatment of rheumatic diseases may also affect CV risk. Because of the common use of NSAIDs and concerns surrounding CV risk with their use, this has been extensively studied. Although some evidence indicates that NSAID use is not associated with increased CV risk in RA, a recent network meta-analysis of 31 trials in 116,429 patients concluded that there is little evidence to suggest that any of the investigated drugs (i.e., naproxen, diclofenac, celecoxib, etoricoxib, rofecoxib, or lumiracoxib) are safe. It is important to note, however, that the study populations in these trials consisted largely of patients with osteoarthritis and other musculoskeletal conditions rather than RA. In contrast, use of DMARDs (methotrexate in particular) and/or biologic agents has been suggested to decrease CV risk in patients with RA. This is believed to be due to effective long-term control of systemic inflammation. Although these findings are intriguing, they cannot be considered definitive evidence that these agents reduce CV risk due to confounding by indication and/or contraindication.

Statins have established effectiveness in the primary prevention of CV events in the general population and at-risk subpopulations (e.g., patients with diabetes mellitus). These agents have both lipid-modifying and anti-inflammatory effects. However, their role in RA is only more recently being explored; thus utilization of statins is low in patients with RA, even in those at high risk. A number of studies are ongoing to resolve these questions such as TRACE-RA (Trial of Atorvastatin in the Primary Prevention of Cardiovascular Endpoints in Rheumatoid Arthritis), a multicenter, placebo-controlled study, aiming to enroll more than 3000 patients with RA without overt CV disease.

Cardiovascular Risk in Other Inflammatory Rheumatic Diseases

Although the evidence for an increased risk of CV events is strongest for RA and SLE, emerging literature suggests that psoriatic arthritis and ankylosing spondylitis may also be associated with excess CV risk. The literature is conflicting regarding whether or not psoriasis is a risk factor for CVD. A large Danish nationwide study demonstrated a small but significant excess CV risk associated with psoriasis, whereas a large U.S. cohort study failed to demonstrate such a risk even with severe psoriasis. Psoriatic arthritis has been shown to be associated with subclinical atherosclerosis after adjusting for traditional CV risk factors. There are also reports suggesting that hyperuricemia, commonly associated with psoriatic arthritis, may be correlated with subclinical atherosclerosis. These findings are further confounded by the complex risk factor profile among psoriasis patients who are more likely to be smokers and diabetics and have metabolic syndrome and atherogenic dyslipidemia. Further research is necessary to disentangle these factors in order to determine the independent impact of psoriasis and psoriatic arthritis on CV risk.

Elevated rates of cardiovascular morbidity and mortality have also been observed in ankylosing spondylitis (AS), although less than that observed with RA. In addition,
AS appears to have a similar abnormal lipid profile to that seen in RA.\textsuperscript{12,13} As in RA, there is growing evidence indicating subclinical vascular disease and endothelial dysfunction in AS.\textsuperscript{14,15} Thus although the body of evidence is far smaller in AS than in RA, there appear to be substantial similarities in the findings of these two conditions.

There is also limited evidence describing excess CV risk with other, rarer rheumatic conditions. Nonetheless, taken together, these findings support the association between chronic inflammatory rheumatic diseases and an excess risk of CV morbidity and mortality. Additional research is necessary to better elucidate the relative contribution of disease factors, treatments, and underlying risk factors on CVD in persons with rheumatic diseases.

**Connection to the Clinic**
Physicians who care for patients with rheumatic diseases should manage these individuals as a high cardiovascular risk subgroup, regardless of their traditional risk factor parameters.

**Future Directions**
1. Large multicenter randomized clinical trials are necessary to establish the cardiovascular (CV) risks and/or benefits associated with antirheumatic drugs.
2. Disease-specific risk assessment tools are necessary to accurately determine an individual patient’s CV risk in the clinical setting.

**References**
33. Holmqvist ME, Wedren S, Jacobsson LT, et al: No increased occurrence of ischemic heart disease prior to the onset of rheumatoid...


97. Bruce IN: “Not only ... but also”: factors that contribute to accelerated atherosclerosis and premature coronary heart disease in systemic lupus erythematosus, Rheumatology 44(12):1492–1502, 2005.


The references for this chapter can also be found on www.expertconsult.com.
Systemic rheumatic diseases have been associated with an increased risk of development of malignancy. This increased risk is the result of fundamental underlying immunologic effects of autoimmunity on cancer risk and on the risk of cancers associated with drug treatments for rheumatic diseases. Accelerated growth of cancer cells in immunodeficient mice and increased risk of cancer in heavily immunosuppressed transplant patients have shaped the perception of the immune system as a potent barrier against neoplasms.\textsuperscript{1,2} It might be expected that immunosuppressive treatment would inevitably result in effects favoring malignant cell growth. However, emerging evidence supports the seemingly paradoxical notion perhaps formulated first by Rudolph Virchow in 1863 that inflammation is a critical component of cancer initiation and progression, and that reduction of systemic inflammation may reduce cancer risk in these conditions.\textsuperscript{3}

Assessment of cancer risk in rheumatic diseases must be weighed against the lifetime risk of developing cancer, which is approximately 20% in Western Europe and North America, with 5% of the general population having current cancer or a history of cancer.\textsuperscript{4} Approximately 1 in 10 women will develop breast cancer, and as many as 1 in 8 men will develop prostate cancer, 1 in 25 colorectal cancer, 1 in 40 lung cancer, and approximately 1 in 100 lymphoma or other lymphoproliferative malignancy.\textsuperscript{5}

The combination of increased risk for some and decreased risk for other types of cancers in different rheumatic diseases may result in a neutral effect for malignancies in general, emphasizing why, from a clinical standpoint, it is important to identify risks pertaining to specific cancers, which may be uncommon. The statistical approach for capturing differences in sparse event data, particularly when malignancy is not a prespecified study outcome, and assumptions of proportional hazards models and stable frequencies of events over time for a nonlinear risk such as cancer can lead to major errors in interpretation.\textsuperscript{6}

**MALIGNANCY IN AUTOIMMUNE RHEUMATIC DISEASES**

Several of the rheumatic diseases, particularly lymphoproliferative disorders, appear to be associated with increased risk of malignancy. A list of rheumatic diseases that have been associated with malignancy is provided in Table 37-1. In a global assessment of susceptibility to Hodgkin's lymphoma using population-based linked registry data from Sweden and Denmark, 32 autoimmune and related conditions were identified from hospital diagnoses in 7476 case subjects with Hodgkin's lymphoma, 18,573 matched control subjects, and more than 86,000 first-degree relatives of case and control subjects. Significantly increased risks of Hodgkin's lymphoma were associated with a personal history of autoimmune disorders, including rheumatoid arthritis (odds ratio [OR], 2.7; 95% confidence interval [CI], 1.9 to 4.0), systemic lupus erythematosus (OR, 5.8; 95% CI, 2.2 to 15.1), sarcoidosis (OR, 14.1; 95% CI, 5.4 to 36.8), and immune thrombocytopenic purpura (OR, ∞; P = 0.022). A statistically significant increase in the risk of Hodgkin's lymphoma was associated with family histories of sarcoidosis (OR, 1.8; 95% CI, 1.01 to 3.1) and ulcerative colitis (OR, 1.6; 95% CI, 1.02 to 2.6).\textsuperscript{6} Shared susceptibility for autoimmune disease and lymphoma is suggested by the association between personal and family history for some of these disorders, particularly sarcoidosis, and a statistically significantly increased risk of Hodgkin's lymphoma, but was not confirmed in a large registry-based study of patients with early rheumatoid arthritis from Sweden.\textsuperscript{6}

The occurrence of cancer has a profound effect on the already compromised quality of life of patients with rheumatic diseases and may affect survivorship. A population-based study of cancer survival in patients with inflammatory arthritis from Great Britain suggests decreased survival compared with the general population.\textsuperscript{5} Survivorship related to the occurrence of cancer in this population was unrelated to disease-modifying antirheumatic drug (DMARD) exposure.\textsuperscript{8}

### KEY POINTS

- Risk of malignancy, especially lymphoproliferative malignancy, is increased in autoimmune rheumatic diseases.
- The occurrence of cancer in patients with rheumatic diseases adversely affects quality of life and life expectancy compared with the general population.
- This risk is related to the pathobiology of the underlying rheumatic disease, including the inflammatory burden, immunologic defects such as overexpression of Bcl-2 oncogenes, traditional risk factors such as smoking, and, in some cases, associated viral infection.
- Several of the immunomodulatory treatments used in the management of autoimmune disease, especially chemotherapeutic agents, are associated with an increased risk of cancer.
- The decision to use immunomodulating therapies in patients with rheumatic disease must take into account host and environmental risk factors for cancer. Effective screening and monitoring strategies can markedly reduce the risk of cancer in these patients.
Rheumatoid Arthritis

**Table 37-1  Rheumatic Diseases Associated with Malignancy**

<table>
<thead>
<tr>
<th>Connective Tissue Disease</th>
<th>Malignancy</th>
<th>Associated Factors</th>
<th>Clinical Alert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sjögren’s syndrome</td>
<td>Lymphoproliferative disorders</td>
<td>Glandular features: lymphadenopathy, parotid or salivary enlargement</td>
<td>Clues to progression from pseudolymphoma to lymphoma include worsening of clinical features, disappearance of rheumatoid factor, and decline of IgM</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Lymphoproliferative disorders</td>
<td>Extraludular features: purpura, vasculitis, splenomegaly, lymphopenia, low C4 cryoglobulins</td>
<td>Non-Hodgkin’s lymphoma should be considered in SLE patients who develop adenopathy or masses; lymphoma of the spleen is another cause of splenic enlargement in SLE</td>
</tr>
<tr>
<td>Systemic lupus erythematosus (SLE)</td>
<td>Lymphoproliferative disorders</td>
<td>Presence of paraproteinemias, greater disease severity, longer disease duration, immunosuppression, Felty’s syndrome</td>
<td>Annual chest radiograph after fibrosis is detected</td>
</tr>
<tr>
<td>Systemic sclerosis (scleroderma)</td>
<td>Alveolar cell carcinoma</td>
<td>Pulmonary fibrosis, interstitial lung disease</td>
<td>Change in skin features or poorly healing lesions should be evaluated</td>
</tr>
<tr>
<td></td>
<td>Nonmelanoma skin cancer</td>
<td>Areas of scleroderma and fibrosis in the skin</td>
<td>If indicated, esophagoscopy and biopsy of distal esophageal constricting lesions</td>
</tr>
<tr>
<td>Dermatomyositis</td>
<td>Ovarian, lung, and gastric cancers in Western populations; nasopharyngeal carcinoma in Asian populations</td>
<td>Older, normal creatinine kinase levels, presence of cutaneous vasculitis; less likely in setting of myositis-specific antibodies</td>
<td>Malignancy evaluation needs to be tailored to individual patient’s age, symptoms, and signs</td>
</tr>
</tbody>
</table>

A link between lymphoma and rheumatoid arthritis was first reported from a medical record linkage study in 1978. Subsequently, a considerable body of evidence emerged supporting rheumatoid arthritis as a pathogenic factor in the development of lymphoma. A standardized incidence ratio (SIR) of 2.4 for lymphoma was described in a population of more than 20,000 Danish patients, as was an increased risk of 1.9 in 1852 U.S. patients.

Using a different approach, a pooled SIR of 3.9 for lymphoma was found using a random effects model in a meta-analysis. Studies using other methods have found odds ratios of 1.3 to 1.5 for lymphoma in case-control studies.

Patients with rheumatoid arthritis may be at higher risk for non-Hodgkin’s lymphoma, especially diffuse large B cell type. Large B cell lymphomas represent up to two-thirds of non-Hodgkin’s lymphomas in patients with rheumatoid arthritis—about twice the rate of diffuse large B cell lymphoma as a proportion of overall non-Hodgkin’s lymphoma in the general population. However, other studies have suggested that the immunophenotype, grade, and histology of lymphomas in patients with rheumatoid arthritis are not different from the general population.

In many studies, the risk of cancer is particularly increased early in the disease course, and cancer risk appears to be greater in patients who have persistently high disease activity, high cumulative disease activity, and more severe disease, and in those who have positive rheumatoid factor (SIR, 3.6; 95% CI, 1.3 to 7.8). The unadjusted OR for average disease activity comparing highest versus lowest quartile was 71.3 (95% CI, 24.1 to 211.4), and the OR for cumulative disease activity of the 10th decile versus the 1st decile was 61.6 (95% CI, 21.0 to 181.0) in a case-control registry study from Sweden.

Extra-articular disease of rheumatoid arthritis, particularly Felty’s syndrome and Sjögren’s syndrome, confers a further increased risk of non-Hodgkin’s lymphoma; one study of 906 men with rheumatoid arthritis revealed a twofold increase in total cancer incidence among patients with rheumatoid arthritis who have Felty’s syndrome. Large granular T cell leukemia (T-LGL) may rarely occur in association with rheumatoid arthritis. T-LGL in rheumatoid arthritis usually is chronic and rarely becomes aggressive.

Whether patients with rheumatoid arthritis are at higher risk for lymphoproliferative disorders other than lymphoma is unclear. At least one study from Canada noted an increased risk of leukemia with an SIR of 2.47 among patients with rheumatoid arthritis. It is interesting to note that an increased risk of lymphoma was not noted in this patient population. A U.S. Veterans Affairs study of 906 men with rheumatoid arthritis reported a twofold increased risk of overall cancer, although no single form of cancer...
stood out, other than non-Hodgkin’s lymphoma, for which a 12-fold increased risk was reported.\(^\text{19}\)

A study that used statewide discharge records from California linking rheumatoid arthritis to Cancer Registry data for 1991 to 2002 revealed 5533 incident cancers among 84,075 rheumatoid arthritis patients observed for approximately 400,000 person-years.\(^\text{20}\) As in other studies, the risk of developing lymphoproliferative cancers was significantly higher among both women and men with rheumatoid arthritis. Men had significantly higher risks for lung, liver, and esophageal cancers, although a lower risk for prostate cancer was noted. Females were at decreased risk for several cancers, including cancers of the breast, ovary, uterus, cervix, and melanoma, and risk reduction ranged from 15% to 57%, compared with the general population. In this study, Hispanic patients had increased risk of leukemia and vaginal/vulva, lung, and liver cancers.\(^\text{20}\)

The risk of premature death from leukemia and lymphoma in patients with rheumatoid arthritis has been reported to be similar or moderately increased compared with patients without rheumatoid arthritis.\(^\text{21,22}\) The rate per 100 patient deaths for leukemia/lymphoma is 1.78, including patients who have been treated with methotrexate and azathioprine.\(^\text{22}\)

A meta-analysis of 21 publications from 1990 to 2007 summarized the risk of malignancy in patients with rheumatoid arthritis.\(^\text{23}\) The risk of lymphoma was increased approximately twofold (SIR, 2.08; 95% CI, 1.8 to 2.39), with greater risks of Hodgkin’s and non-Hodgkin’s lymphoma. The risk of lung cancer was increased with an SIR of 1.63 (95% CI, 1.43 to 1.87). The risk of colorectal cancer was decreased (SIR, 0.77; 95% CI, 0.65 to 0.90), as was the risk of breast cancer (SIR, 0.84; 95% CI, 0.79 to 0.90). The overall SIR for malignancy was slightly increased at 1.05 (95% CI, 1.01 to 1.09). The overall increased risk of cancer in patients with rheumatoid arthritis was largely driven by increased risks of lymphoproliferative cancers.

The risk of lung cancer was also higher among patients with rheumatoid arthritis in the U.S. veteran population. Patients with rheumatoid arthritis were 43% more likely (OR, 1.43) to develop lung cancer compared with patients without rheumatoid arthritis after adjustments for age, gender, race, and tobacco and asbestos exposure.\(^\text{24}\) No effect of lung cancer on life expectancy was found in a retrospective cohort study; adenocarcinoma was the major histologic form of cancer found in patients with rheumatoid arthritis and in controls. Rheumatoid arthritis appeared to have no influence on the lung cancer stage.\(^\text{25}\)

A study of 42,262 patients hospitalized with rheumatoid arthritis from 1980 to 2004 indexed to the Swedish national cancer registry found that SIRs for Hodgkin’s and non-Hodgkin’s lymphoma were increased for upper digestive tract cancers and for squamous cell skin cancer, and detected an excess of nonthyroid endocrine tumors in patients with rheumatoid arthritis.\(^\text{26}\) Colon and rectal and endometrial cancers were less frequent in patients with rheumatoid arthritis. The decreased risk of colorectal cancer may be attributable to the use of long-term nonsteroidal anti-inflammatory agents in patients with rheumatoid arthritis.\(^\text{27}\)

In summary, cancer is frequent in the general population and is at least as common among patients with rheumatoid arthritis. Following a diagnosis of rheumatoid arthritis at the typical age of 55 years, one in five patients will be diagnosed with cancer; however, in the great majority of patients, the cancer cannot be linked to rheumatoid arthritis or to its treatment but rather reflects the background cancer risk. Although a history of previous lymphoma does not appear to be a risk factor for developing rheumatoid arthritis, risks of certain cancers, particularly hematopoietic malignancies, are increased and are related to the disease itself.\(^\text{7}\)

**Systemic Lupus Erythematosus**

### KEY POINTS

- The risk of lymphoma is at least twofold increased in systemic lupus erythematosus (SLE).
- Risks of solid malignancy, including lung, thyroid, and kidney cancers, and of skin cancer are increased overall in SLE, although rates of cervix and prostate cancers appear to be somewhat lower. Breast cancer risk has been reported to be increased in some studies and decreased in others.

The risk of at least certain malignancies appears to be increased in patients with SLE. Overall SIR ranges from 1.1 to 2.6,\(^\text{28}\) with the most increased risk being that of lymphomas (SIR of 3.57 for non-Hodgkin’s lymphoma and 2.35 for Hodgkin’s disease).\(^\text{28,29}\) The risk is especially high for diffuse large B cell lymphoma, often of aggressive subtypes.\(^\text{30,31}\)

A large multicenter (23 centers) international cohort of 9547 patients with an average follow-up of 8 years confirmed an increased risk of cancer in patients with SLE. For all cancers combined, the SIR estimate was 1.15 (95% CI, 1.05 to 1.27); for all hematologic malignancies, it was 2.75 (95% CI, 2.13 to 3.49); and for non-Hodgkin’s lymphoma, it was 3.65 (95% CI, 2.63 to 4.93). The data also suggest increased risks of lung cancer (SIR, 1.37; 95% CI, 1.05 to 1.76) and hepatobiliary cancer (SIR, 2.60; 95% CI, 1.25 to 4.78).\(^\text{29}\)

Patients with SLE in a California statewide patient hospital discharge database from 1991 to 2001 were followed using Cancer Registry data to compare observed versus expected numbers of cancers based on age, sex, and specific incidence rates in the California population.\(^\text{32}\) A total of 30,478 SLE patients were observed for 157,969 person-years. There were a total of 1,273 cancers for an incidence rate of 8.13 per 100 person-years. Patients with SLE had increased risks of lung cancer (SIR, 1.37; 95% CI, 1.05 to 1.76) and for lymphomas (SIR, 2.19; 95% CI, 1.74 to 2.68). Patients with SLE had higher risks of Hodgkin’s lymphoma (SIR of 4.65 for Hodgkin’s lymphoma and 3.57 for non-Hodgkin’s lymphoma).\(^\text{32}\)

Patients with SLE in a California statewide patient hospital discharge database from 1991 to 2001 were followed using Cancer Registry data to compare observed versus expected numbers of cancers based on age, sex, and specific incidence rates in the California population.\(^\text{32}\)

Other studies have reported possibly increased risk of malignancies other than lymphoproliferative cancers in SLE. Patients may be at slightly higher risk of thyroid cancer.\(^\text{33}\) An increased risk of squamous cell skin cancer was found in 238 patients with SLE.\(^\text{34}\) The risk of breast cancer may be increased by about 1.5-fold to twofold compared with the general population, even after consideration of age,
purity, family history, and exogenous estrogens. The risk of abnormal Pap smears and cervical dysplasia appears to be higher in women with SLE than in those without SLE, although the risk of invasive cervical cancer is not increased.

The origin of any risk of the development of malignant disorders in SLE remains unclear, although it does not appear to be related to the use of immunosuppressive or cytotoxic agents; most cohorts are too small to allow detection of a statistically meaningful risk increase in rare events over short periods of observation. Race and ethnicity have not been identified as major factors in cancer risk in SLE. Antimalarial drug use does not appear to affect the relative risk of malignancy, as was postulated in early studies.

Risk factors for the development of hematologic malignancies may relate to inflammatory burden and disease activity, immunologic defects and overexpression of Bcl-2 oncoproteins, and viruses, especially Epstein-Barr virus (EBV). A nested case-control study that included 6438 patients with SLE linked to the national cancer registry in Sweden found that leukopenia, independent of immunosuppressive treatment, was a risk factor for developing these leukemias. Bone marrow investigation was suggested for SLE patients with long-standing leukopenia and anemia.

Disease characteristics predisposing to non-Hodgkin’s lymphoma include longer disease duration and increased disease activity with moderately severe end-organ damage. Women with SLE, likely out of concern for treatment side effects and effects of pregnancy on disease control, are less often exposed to oral contraceptives and are more likely to be nulliparous, which may affect their malignancy risk. On the other hand, the possibly increased breast cancer risk to be nulliparous, which may affect their malignancy risk. Women with SLE, likely out of concern for breast cancer risk may be related to inflammation and disease activity. Immunologic defects and overexpression of Bcl-2 oncoproteins, and viruses, especially Epstein-Barr virus (EBV).

A nested case-control study that included 6438 patients with SLE linked to the national cancer registry in Sweden found that leukopenia, independent of immunosuppressive treatment, was a risk factor for developing these leukemias. Bone marrow investigation was suggested for SLE patients with long-standing leukopenia and anemia.

Disease characteristics predisposing to non-Hodgkin’s lymphoma include longer disease duration and increased disease activity with moderately severe end-organ damage. Women with SLE, likely out of concern for treatment side effects and effects of pregnancy on disease control, are less often exposed to oral contraceptives and are more likely to be nulliparous, which may affect their malignancy risk. On the other hand, the possibly increased breast cancer risk may be related to inflammation and disease activity. Immunologic defects and overexpression of Bcl-2 oncoproteins, and viruses, especially Epstein-Barr virus (EBV).

A nested case-control study that included 6438 patients with SLE linked to the national cancer registry in Sweden found that leukopenia, independent of immunosuppressive treatment, was a risk factor for developing these leukemias. Bone marrow investigation was suggested for SLE patients with long-standing leukopenia and anemia.

Disease characteristics predisposing to non-Hodgkin’s lymphoma include longer disease duration and increased disease activity with moderately severe end-organ damage. Women with SLE, likely out of concern for treatment side effects and effects of pregnancy on disease control, are less often exposed to oral contraceptives and are more likely to be nulliparous, which may affect their malignancy risk. On the other hand, the possibly increased breast cancer risk may be related to inflammation and disease activity. Immunologic defects and overexpression of Bcl-2 oncoproteins, and viruses, especially Epstein-Barr virus (EBV).

A nested case-control study that included 6438 patients with SLE linked to the national cancer registry in Sweden found that leukopenia, independent of immunosuppressive treatment, was a risk factor for developing these leukemias. Bone marrow investigation was suggested for SLE patients with long-standing leukopenia and anemia.

Disease characteristics predisposing to non-Hodgkin’s lymphoma include longer disease duration and increased disease activity with moderately severe end-organ damage. Women with SLE, likely out of concern for treatment side effects and effects of pregnancy on disease control, are less often exposed to oral contraceptives and are more likely to be nulliparous, which may affect their malignancy risk. On the other hand, the possibly increased breast cancer risk may be related to inflammation and disease activity. Immunologic defects and overexpression of Bcl-2 oncoproteins, and viruses, especially Epstein-Barr virus (EBV).

A nested case-control study that included 6438 patients with SLE linked to the national cancer registry in Sweden found that leukopenia, independent of immunosuppressive treatment, was a risk factor for developing these leukemias. Bone marrow investigation was suggested for SLE patients with long-standing leukopenia and anemia.

Disease characteristics predisposing to non-Hodgkin’s lymphoma include longer disease duration and increased disease activity with moderately severe end-organ damage. Women with SLE, likely out of concern for treatment side effects and effects of pregnancy on disease control, are less often exposed to oral contraceptives and are more likely to be nulliparous, which may affect their malignancy risk. On the other hand, the possibly increased breast cancer risk may be related to inflammation and disease activity. Immunologic defects and overexpression of Bcl-2 oncoproteins, and viruses, especially Epstein-Barr virus (EBV).

A nested case-control study that included 6438 patients with SLE linked to the national cancer registry in Sweden found that leukopenia, independent of immunosuppressive treatment, was a risk factor for developing these leukemias. Bone marrow investigation was suggested for SLE patients with long-standing leukopenia and anemia.

Disease characteristics predisposing to non-Hodgkin’s lymphoma include longer disease duration and increased disease activity with moderately severe end-organ damage. Women with SLE, likely out of concern for treatment side effects and effects of pregnancy on disease control, are less often exposed to oral contraceptives and are more likely to be nulliparous, which may affect their malignancy risk. On the other hand, the possibly increased breast cancer risk may be related to inflammation and disease activity. Immunologic defects and overexpression of Bcl-2 oncoproteins, and viruses, especially Epstein-Barr virus (EBV).

A nested case-control study that included 6438 patients with SLE linked to the national cancer registry in Sweden found that leukopenia, independent of immunosuppressive treatment, was a risk factor for developing these leukemias. Bone marrow investigation was suggested for SLE patients with long-standing leukopenia and anemia.

Disease characteristics predisposing to non-Hodgkin’s lymphoma include longer disease duration and increased disease activity with moderately severe end-organ damage. Women with SLE, likely out of concern for treatment side effects and effects of pregnancy on disease control, are less often exposed to oral contraceptives and are more likely to be nulliparous, which may affect their malignancy risk. On the other hand, the possibly increased breast cancer risk may be related to inflammation and disease activity. Immunologic defects and overexpression of Bcl-2 oncoproteins, and viruses, especially Epstein-Barr virus (EBV).

A nested case-control study that included 6438 patients with SLE linked to the national cancer registry in Sweden found that leukopenia, independent of immunosuppressive treatment, was a risk factor for developing these leukemias. Bone marrow investigation was suggested for SLE patients with long-standing leukopenia and anemia.

Disease characteristics predisposing to non-Hodgkin’s lymphoma include longer disease duration and increased disease activity with moderately severe end-organ damage. Women with SLE, likely out of concern for treatment side effects and effects of pregnancy on disease control, are less often exposed to oral contraceptives and are more likely to be nulliparous, which may affect their malignancy risk. On the other hand, the possibly increased breast cancer risk may be related to inflammation and disease activity. Immunologic defects and overexpression of Bcl-2 oncoproteins, and viruses, especially Epstein-Barr virus (EBV).

A nested case-control study that included 6438 patients with SLE linked to the national cancer registry in Sweden found that leukopenia, independent of immunosuppressive treatment, was a risk factor for developing these leukemias. Bone marrow investigation was suggested for SLE patients with long-standing leukopenia and anemia.

Disease characteristics predisposing to non-Hodgkin’s lymphoma include longer disease duration and increased disease activity with moderately severe end-organ damage. Women with SLE, likely out of concern for treatment side effects and effects of pregnancy on disease control, are less often exposed to oral contraceptives and are more likely to be nulliparous, which may affect their malignancy risk. On the other hand, the possibly increased breast cancer risk may be related to inflammation and disease activity. Immunologic defects and overexpression of Bcl-2 oncoproteins, and viruses, especially Epstein-Barr virus (EBV).
scleroderma, has not been associated with increased risk of cancer.

**KEY POINTS**

<table>
<thead>
<tr>
<th>The risk of cancer in patients with idiopathic inflammatory myopathies is about five to seven times higher than in the general population.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malignancy is strongly associated with dermatomyositis and, if present, is often detectable at disease onset.</td>
</tr>
<tr>
<td>The most common malignancies in inflammatory myositis are adenocarcinomas.</td>
</tr>
<tr>
<td>Suspicion for cancer should be high, especially in patients with active muscle and skin inflammation but normal creatine kinase levels, age over 50 years, and periungual erythema.</td>
</tr>
</tbody>
</table>

Both dermatomyositis and polymyositis occurring in adults have been associated with malignancies. The link to malignancy in newly diagnosed patients with dermatomyositis is strongest, although in neither condition is the origin of the association well understood. The association between malignancy and polymyositis and inclusion body myositis is less strong.

Assessment of risk is complicated by the temporal relationship between the development of malignancy and cancer. In particular, some cancers pre-date the onset of inflammatory myopathy, so that the inflammatory myopathy can be better considered a paraneoplastic syndrome (see Chapter 122); it is also likely that the presence of inflammatory myopathies represents a risk factor for the subsequent development of malignancy.

The incidence of cancer occurring in patients with inflammatory myositis is approximately five to seven times higher than in the general population. The prevalence of malignancy is about 25%; cancers are reported more frequently in dermatomyositis, occurring in 6% to 60% of patients, and in 0 to 28% of patients with polymyositis.

In most studies, cancers manifest within 2 years before or after the initial diagnosis of inflammatory myopathy. Inflammatory myopathies may initially manifest with the recurrence of a previously diagnosed cancer. A previously diagnosed but inactive inflammatory myopathy may become reactivated with occurrence of a cancer, supporting the hypothesis of autoantigens as drivers of the inflammatory disease.

The strength of the association between malignancy and inflammatory myositis varies. One study done at the Mayo Clinic failed to reveal an increased risk of malignancy in patients with inflammatory myopathy, and a more recent registry study from Sweden of 788 patients diagnosed with dermatomyositis or polymyositis between 1963 and 1987 revealed that 15% of 392 patients with dermatomyositis had cancer diagnosed concurrently with or after the diagnosis of dermatomyositis, with a relative risk of cancer of 2.4 (95% CI, 1.6 to 3.6) for males and 3.4 (95% CI, 2.4 to 4.7) for females. Of 396 patients with polymyositis, 9% had cancer at or after the time of diagnosis of polymyositis, with the relative risk of 1.8 for development of cancer (95% CI, 1.1 to 2.7) in males and 1.7 (95% CI, 1.0 to 2.5) in females.

A population-based retrospective cohort study from Victoria, Australia, of 537 patients with biopsy-proven dermatomyositis and polymyositis reported a relative risk for malignancy in dermatomyositis compared with polymyositis of 2.4 (95% CI, 1.3 to 4.2) with a higher SIR for dermatomyositis than polymyositis (6.2 vs. 2.0). Finally, odds ratios for association of cancer with dermatomyositis from a large meta-analysis were reported to be 4.4, and for polymyositis, 2.1.

A wide range of malignancies are associated with dermatomyositis and polymyositis. The most common malignancies in populations of Northern European descent are adenocarcinomas of the cervix, lungs, ovaries, pancreas, bladder, and stomach, which account for more than two-thirds of these cancers. In patients from Southeast Asia, a higher proportion of nasopharyngeal cancers are found, followed by lung cancer.

The association of cancer is less well understood for more unusual forms of inflammatory myopathies. Amyopathic dermatomyositis, a rare form of dermatomyositis with typical cutaneous but no muscle involvement, can be associated with the development of cancer, but the frequency of this condition is low, so that no stable estimates of cancer risk are available. Inclusion body myositis has not been well studied for the same reasons, although the overall risk of cancer of 2.4 suggests a possible link.

It is likely that relevant antigens are expressed in the underlying tumor and affected muscle. Myositis-specific antigens develop during the process of regeneration in patients who have myositis; these are the same antigens expressed in some cancers known to be associated with the development of inflammatory myopathies. A link between malignancy and inflammatory myositis is further supported by observations that in many cases, myositis improves after removal of the malignancy.

Clinical disease characteristics that may portend higher malignancy risk include active inflammatory disease with normal serum levels of creatine kinase, distal extremity weakness, pharyngeal and diaphragmatic involvement, and leukocytoclastic vasculitis. Other independent risk factors for the development of cancer in patients who have dermatomyositis in one study of 92 patients included age at diagnosis greater than 52 years (hazard ratio [HR], 7.24; 95% CI, 2.35 to 22.41), rapid onset of skin and/or muscle symptoms (HR, 3.11; 95% CI, 1.07 to 9.02), periungual erythema (HR, 3.93; 95% CI, 1.16 to 13.24), low baseline level of complement factor C4 (HR, 2.74; 95% CI, 1.11 to 6.75), and possibly topoisomerase I. A low baseline lymphocyte count was a protective factor for malignancy (HR, 0.33; 95% CI, 0.14 to 0.80), although the number of assessable patients was small.

**Sjögren’s Syndrome**

**KEY POINTS**

The risk of lymphoproliferative cancers, especially various types of lymphoma, is at least sixfold increased in patients with primary Sjögren’s syndrome.
Immunologic perturbations, including p35 mutations and B cell activation, as well as Helicobacter pylori, are likely predisposing risk factors.

Patients with primary Sjögren’s syndrome are at increased risk of lymphoproliferative diseases, especially non-Hodgkin’s lymphoma. The relative risk for the development of lymphoproliferative disorders in these patients ranges from 6 to 44 in individual studies, and a meta-analysis of cohort studies reported a pooled SIR of 18.8. Lymphoproliferative disorders eventually occur in between 4% and 10% of patients with primary Sjögren’s syndrome, with a lifetime risk of non-Hodgkin’s lymphoma of about 5%, 67,71,72

In addition to non-Hodgkin’s lymphoma, forms of lymphoproliferative disease seen in patients with Sjögren’s syndrome include low-grade B cell lymphoma and diffuse large B cell lymphoma, including follicular center lymphoma. Less commonly seen lymphoproliferative diseases include lymphocytic leukemia, Waldenström’s macroglobulinemia, and multiple myeloma. The risk of other cancers does not appear to be particularly high in patients with Sjögren’s syndrome. 67,71 The development of lymphoma and malignancy in patients with Sjögren’s syndrome does not appear to affect or cause mortality.

The pathoetiology of lymphoproliferative diseases occurring in patients with Sjögren’s syndrome is unclear. It is likely that the B cell activation characteristic of Sjögren’s syndrome is a predisposing risk factor. Most lymphomas in Sjögren’s syndrome appear to arise from lymphoepithelial sialadenitis or benign lymphoepithelial lesions, perhaps associated with p35 mutations. 73 Infectious agents such as hepatitis C and Epstein-Barr virus have been implicated, although the nature of this relationship remains speculative. Helicobacter pylori is associated with MALT (mucosa-associated lymphoid tissue) lymphoma in Sjögren’s syndrome. 74

Appropriate evaluation in symptomatic patients should include diagnostic testing for H. pylori in this setting. 75 The link of cancer in Sjögren’s syndrome to the proto-oncogene BCL-2 translocation is, as yet, not clearly defined but may be helpful for early detection of malignancy. 75,76

Vasculitis

KEY POINT

It is unclear whether the risk of cancer development is increased in vasculitis independent of drug treatment effects.

Vasculitis as a paraneoplastic syndrome may be present in about 8% of patients with malignancy (see Chapter 122). 77 Not as well studied is the risk of primary malignancy in patients with vasculitis. Most cases appear to be related to treatment, although one study using the Danish Cancer Registry suggested an increased risk of nonmelanoma within 2 years of the vasculitis diagnosis in antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis (granulomatosis with polyangiitis formerly Wegener’s granulomatosis) (OR, 4.0; 95% CI, 1.4 to 12). 78

Current data do not support a link between ANCA-associated vasculitis (granulomatosis with polyangiitis) and malignancy as a trigger for vasculitis, or the vasculitis itself as a trigger for malignancy, independent of treatment effects.

The risk of malignancy among patients with giant cell arteritis was not increased in a population-based study of 204 patients with giant cell arteritis and 407 age- and sex-matched controls. 79

Seronegative Spondyloarthropathy

KEY POINT

The overall risk of cancer does not appear to be increased in the seronegative spondyloarthropathies.

The risk of cancer among patients with spondyloarthropathies is not as well studied as that of patients with rheumatoid arthritis and other connective tissue diseases. A cohort study of 665 patients from Canada with psoriatic arthritis revealed an SIR for all cancers of 0.98 (95% CI, 0.77 to 1.24) without evidence of increase in cancer type–specific SIR for hematologic, lung, or breast cancer. 80

No overall increased risk of cancer has been noted among patients with ankylosing spondylitis, as was reported in a Swedish national cohort of patients with ankylosing spondylitis admitted to Swedish hospitals from 1965 to 1995, linked to the Swedish Cancer Registry and Death Registry. These patients appear to not have increased risk of malignant lymphoma. 81,82 A population-based cohort study of patients with ankylosing spondylitis from Australia suggests a prevalence of malignancy of 6.8%, which parallels baseline population prevalence rates reported in Western populations. 83

Cancer Risks Associated with Antirheumatic Drug Therapies

KEY POINTS

The effect on cancer risk of therapies used in the treatment of autoimmune rheumatic diseases can be difficult to separate from the underlying risk intrinsic to the diseases themselves.

Nonsteroidal anti-inflammatory drugs and glucocorticosteroids have not been associated with increased risk of cancer.

The risk of cancer with chemotherapeutic drugs is greatest in patients who have had the highest cumulative exposure to these agents.

Assessment of risk associated with both nonbiologic and biologic DMARDs is challenging because of the overall generally high burden of cancer in the population, the variable rheumatic disease–related cancer risk, and the potential risk of cancer associated with agents used to treat them. Disease severity may be a risk factor for developing cancer, introducing confounding or channeling bias if patients with
severe disease are treated more intensively with immunomodulatory agents. The sequential and combined use of immunomodulatory agents further complicates the assessment of risk related to individual agents. A further concern, as with all immunosuppressive drugs, is the oncogenic potential of immunosuppressive therapies in patients who have a pre-existent or concurrent cancer, and whether such patients should be treated with DMARDs, and if so, which DMARDs should be given.

Nonsteroidal anti-inflammatory agents and glucocorticosteroids do not appear to be associated with increased risk of malignancy in patients with rheumatoid arthritis nor other rheumatic diseases.\textsuperscript{84,85} In a large population-based cohort study from Sweden, a total duration of oral steroid treatment of less than 2 years was not associated with lymphoma risk (OR, 0.87; 95% CI, 0.51 to 1.5), whereas treatment lasting longer than 2 years was associated with a lower lymphoma risk (OR, 0.43; 95% CI, 0.26 to 0.72).\textsuperscript{86} The duration of rheumatoid arthritis at initiation of oral corticosteroids did not affect lymphoma risk. Whether this observed reduced lymphoma risk may be due to decreased disease activity, is a generic effect of steroids, or is specific to rheumatoid arthritis is uncertain.\textsuperscript{86}

**Nonbiologic Disease-Modifying Antirheumatic Drug Therapy**

**KEY POINTS**

| The nonbiologic disease-modifying antirheumatic drugs gold, hydroxychloroquine, penicillamine, and sulfasalazine are not associated with increased risk of cancer. |

| The risk of lymphoma, especially B cell lymphoma, may be increased with methotrexate. Some patients may experience regression with discontinuation of methotrexate. |

| Use of azathioprine is associated with increased risk of lymphoma, and alkylating agents are particularly associated with increased risk of several types of malignancies, including bladder and lung cancer and leukemia in patients treated with cyclophosphamide. |

The nonbiologic (nb-)DMARDs sulfasalazine and hydroxychloroquine and gold and penicillamine do not appear to be associated with increased risk of cancer. Radiation is no longer used for the treatment of rheumatic diseases and will not be further addressed. Although no increase in cancer occurrence has been reported, a paucity of data is available regarding the long-term risks of malignancies occurring with leflunomide.\textsuperscript{87}

All of the antimetabolites used in prevention of transplant rejection and treatment of cancer have tumorigenic potential and, in general, may, or do, confer an increased risk of malignancy. The risk of cancer is increased with chemotherapeutic nb-DMARDs, particularly cyclophosphamide, and the risk of certain cancers, particularly lymphoproliferative disorders, may be increased with the use of other nb-DMARDs, such as azathioprine, methotrexate, and cyclosporine, although data are relatively sparse. Data regarding cancer risk in rheumatic diseases in patients treated with mycophenolate mofetil are few, although cancers clearly do occur in patients treated with this drug.\textsuperscript{88}

Overall risk for lymphoproliferative disorders in patients treated with nb-DMARDs appears to be greatest in patients who have had highest cumulative exposure to DMARDs, compared with patients with less than 1 year of exposure (SIR, 4.82).\textsuperscript{89} A case-control study of 378 Swedish patients with rheumatoid arthritis who developed lymphoma and controls demonstrated that high accumulated rheumatoid arthritis disease activity was associated with increased lymphoma risk (relative risk [RR] of 62 for highest vs. lowest deciles of accumulated disease activity; 95% CI, 21 to 181). When adjusted for disease activity, DMARD use did not emerge as a risk factor (RR, 0.9; 95% CI, 0.6 to 1.2), with the exception of azathioprine, which was associated with a fourfold increased risk of lymphoma.\textsuperscript{84}

**Methotrexate**

The overall malignancy risk attributable to methotrexate treatment in patients with rheumatic diseases does not appear to be increased, although numerous studies suggest that the risk of lymphoproliferative disease may be increased.

Most cases of methotrexate-associated lymphomas reported in the literature are B cell lymphomas, often with extranodal involvement.\textsuperscript{90} Assays for EBV in one study revealed that 7 of 17 patients (41%) were positive.\textsuperscript{88} Further evidence supporting a link between methotrexate use and the development of lymphoma comes from observations of spontaneous remission of B cell lymphoma in 8 of 50 cases, including 4 that were positive for EBV.\textsuperscript{90} This suggests that methotrexate may potentiate persistent immunologic stimulation, clonal selection, and malignant transformation of B cells by direct oncogenic action, decreased apoptosis of infected B cells, and decreased natural killer cell activity.\textsuperscript{90}

**Azathioprine**

The use of azathioprine may be associated with increased risk of lymphoproliferative disorders. Studies from a Canadian azathioprine registry revealed increased rates of lymphoproliferative disorders in patients with rheumatoid arthritis compared with the general population (SIR, 8.05).\textsuperscript{91} A study from Britain also revealed a markedly increased risk of development of lymphoma in rheumatoid arthritis patients treated with azathioprine, with an estimated 1 case of lymphoma per 1000 patient-years of azathioprine treatment.\textsuperscript{92} Risk was highest in patients on higher daily doses of azathioprine of up to 300 mg per day.

The risk of leukemia in patients with systemic lupus erythematosus treated with azathioprine has been generally reported as not increased, although at least one study reported an increase in risk.\textsuperscript{91,93} In one study with up to 24 years of longitudinal follow-up, 5.4% of patients treated with azathioprine developed malignancies, none of which were lymphomas, compared with 6.7% of patients who had never received azathioprine, 3 of whom developed lymphoma.\textsuperscript{93}

**Cyclosporine**

Relatively few patients with rheumatic diseases treated with cyclosporine have been followed for protracted periods,
making assessment of malignancy risk in these patients difficult. Similar to methotrexate, cyclosporine has been associated with the development of EBV-associated lymphomas in a few patients with rheumatoid arthritis.44 A retrospective study aggregating experience from clinical trials of more than 1000 patients with rheumatoid arthritis treated with cyclosporine failed to demonstrate an increased risk of malignancy, at least beyond that seen with other DMARDs.45

**Alkylating Agents**

The use of alkylating agents in patients with rheumatoid arthritis, systemic lupus erythematosus, and vasculitis has been associated with an increase in non-Hodgkin’s lymphoma, leukemia, skin cancer, bladder cancer, and solid malignancies.6,96-98 Considerably more experience with cyclophosphamide than chlorambucil has been documented in these diseases.

Cyclophosphamide use in rheumatic diseases is associated with an overall increased risk of developing malignancy of between 1.5 and 4.1, compared with controls. This risk is best studied in ANCA-associated vasculitis (especially necrotizing granulomatosis with polyangiitis), in which risk is highest for bladder cancers (SIR, 4.8), leukemia (SIR, 5.7), and lymphoma (SIR, 4.2). Bladder cancer is a particular concern, and patients who have been treated with higher doses over longer periods and those who smoke appear to be at especially high risk.97 Bladder cancer related to cyclophosphamide use may occur within 1 year of initiation of therapy and up to 15 years or longer after discontinuation of cyclophosphamide treatment.97

Increased risk of hemorrhagic cystitis of the urinary bladder or development of bladder cancer is due to cyclophosphamide metabolites, especially acrolein. For this reason, current recommendations are to attempt to restrict the use of cyclophosphamide to 6 months or less, and to use it only in life-threatening or organ-threatening disease. The risk of bladder cancer, although probably not the overall malignancy risk, may be less with the use of pulse intravenous cyclophosphamide than with daily oral administration. Some authors advocate concurrent administration of mesna, which inactivates acrolein in the urine. Mesna may be administered intravenously at the time of pulse cyclophosphamide dosing, or by mouth daily, although this is rarely done because of its disagreeable taste.

**Biologic Response Modifiers**

**KEY POINTS**

- The overall risk of cancer associated with biologics used in the treatment of rheumatoid arthritis does not appear to be markedly increased from baseline cancer risk in these patients.
- The risk of skin cancer, especially nonmelanotic skin cancer, does appear to be somewhat increased by about 1.5-fold in patients treated with anti-tumor necrosis factor therapy.

Biologic response modifiers target specific pathways involved in the pathogenesis of some rheumatic diseases such as rheumatoid arthritis and spondyloarthritis. The term targeted should not imply absolute selectivity between physiologic and pathologic processes with these drugs.

**Anti–Tumor Necrosis Factor Agents**

More than 40 randomized controlled trials and several large cohort studies have investigated the use of anti–tumor necrosis factor (TNF) agents over a wide range of indications. Table 37-2 contains a list of meta-analyses and cohort studies undertaken to explore anti-TNF treatment and solid cancer/lymphoma in rheumatoid arthritis. The concern regarding cancer arises from animal models of TNF action, in vitro studies, and studies in humans suggesting that TNF is important in cancer initiation and promotion; indeed, anti-TNF agents may even be beneficial for patients with cancers, although clinical evidence of such a benefit is lacking.99 An interesting observation of regression of non–small cell lung cancer in a patient receiving anti-TNF therapy has been reported, which adds to the biologic plausibility of a connection between anti-TNF therapies and malignancy in individual patients who may have particular susceptibility.100

A report from a large U.S. databank about malignancies in patients with rheumatoid arthritis treated with anti-TNF therapies and followed for more than 10,000 patient-years compared with all patients with rheumatoid arthritis regardless of therapy followed for more than 29,000 patient-years reported an SIR for lymphoma of 2.9 in those patients receiving anti-TNF compared with 1.9 for all rheumatoid arthritis patients treated with biologics.101

**Table 37-2 Meta-analyses and Cohort Studies Exploring Anti–Tumor Necrosis Factor Treatment and Solid Cancer/Lymphoma in Rheumatoid Arthritis**

<table>
<thead>
<tr>
<th>References</th>
<th>Study Design</th>
<th>Number of Patients</th>
<th>Risk Estimate (All Anti-TNF* vs. Control Unless Stated Otherwise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>103</td>
<td>RCT meta-analysis</td>
<td>5014</td>
<td>OR, 3.3 (95% CI, 1.2 to 9.1)</td>
</tr>
<tr>
<td>104</td>
<td>RCT meta-analysis</td>
<td>5788</td>
<td>OR, 2.4 (95% CI, 1.2 to 4.8)</td>
</tr>
<tr>
<td>101</td>
<td>Cohort (NDB)</td>
<td>13,869</td>
<td>OR, 1.0 (95% CI, 0.8 to 1.2)</td>
</tr>
<tr>
<td>107</td>
<td>Cohort (pooled data from three health care utilization databases)</td>
<td>7830 subjects ≥ age 65</td>
<td>HR, 0.98 (95% CI, 0.73 to 1.31), excluding NMSC</td>
</tr>
<tr>
<td>105</td>
<td>RCT meta-analysis</td>
<td>8808</td>
<td>OR, 1.31 (95% CI, 0.69 to 2.48)</td>
</tr>
<tr>
<td>106</td>
<td>Cohort SBR</td>
<td>6366</td>
<td>OR, 1.21 (95% CI, 0.63 to 2.32) <em>Exposure adjusted</em></td>
</tr>
<tr>
<td>108</td>
<td>Nested case-control (RABBIT)</td>
<td>cases: 74; cohort overall 5120</td>
<td>RR, 1.00 (95% CI, 0.86 to 1.15) NMSC excluded</td>
</tr>
</tbody>
</table>

*Etanercept, infliximab, and adalimumab account for the vast majority (>90%) of anti-TNF agents in these studies.

CI, confidence interval; HR, hazard ratio; NDB, National Databank for Rheumatic Diseases; NMSC, nonmelanotic skin cancer; OR, odds ratio; RABBIT, German Biologic Register; RCT, randomized controlled trial; RR, relative risk; SBR, Swedish Biologics Register; TNF, tumor necrosis factor.
arthritists, regardless of therapy. A biologics registry study from Sweden found an SIR of 2.9 for lymphoma compared with the general population, but when compared with TNF-naive patients, the risk was not elevated (RR, 1.1).

The possibly increased risk of cancer early after the start of anti-TNF therapy may be a factor in meta-analyses of randomized clinical trials, which have suggested a possibly increased risk of cancer in patients with rheumatoid arthritis treated with these agents. One meta-analysis of nine randomized clinical trials including infliximab and adalimumab found an odds ratio of 2.4 for developing any malignancy for patients receiving infliximab or adalimumab compared with patients receiving placebo.

Another analysis based on individual patient data from all nine available randomized etanercept trials in patients with rheumatoid arthritis included 3316 patients, 2244 of whom received etanercept and 1072 who were TNF naïve. Incident malignancies were diagnosed in 26 patients in the etanercept group and 7 patients in the control group, yielding a hazard ratio of 1.84 (95% CI, 0.79 to 4.28).

A pooled analysis of randomized controlled trials using etanercept, infliximab, or adalimumab attempted to distinguish between the effects of recommended versus higher doses of anti-TNF on the development of malignancy, excluding nonmelanoma skin cancers. Exposure-adjusted analysis revealed odds ratios of 1.21 (95% CI, 0.79 to 2.82) and 3.04 (95% CI, 0.05 to 9.68) in patients treated with recommended and high doses of anti-TNF agents, respectively.

Results of larger observational studies have not replicated the increased risk of malignancy observed with meta-analytical approaches. In the Swedish Biologics Registry, overall cancer risk was similar in anti-TNF-treated patients with rheumatoid arthritis compared with three different control cohorts. In this database, no trend toward increased cancer incidence was noted with longer duration of TNF exposures. Studies from other databases including the German and British Biologic Registries and a large North American cohort have detected no significant safety signals with respect to overall cancer risk.

Use of anti-TNF agents may be associated with increased risk of nonmelanotic skin cancer. An odds ratio for nonmelanotic skin cancer of 1.5 (95% CI, 1.2 to 1.8) was reported from the U.S. National Databank of Rheumatic Diseases. A cohort study from this population suggested that the combination of TNF plus methotrexate versus control was associated with a higher risk of nonmelanoma skin cancer (HR, 1.97; 95% CI, 1.51 to 2.58) compared with a hazard ratio of 1.24 (95% CI, 0.97 to 1.58) in patients receiving anti-TNF monotherapy versus controls, emphasizing the possible potentiating effects of combination therapies for the development of cancers. This point is perhaps underlined by observations from a randomized clinical trial of 180 patients with ANCA-associated (granulomatosis with polyangiitis) vasculitis. The occurrence of solid and skin malignancies in the group assigned to etanercept was increased above that expected from treatment with cyclophosphamide alone.

Discordant results regarding cancer risk are likely explained by different patient populations and differing drug exposures. Meta-analyses of clinical trials generally reflect relatively short-term effects but offer the advantage of randomization, which should largely neutralize the variability introduced by individual comorbidities and previous drug exposures; long-term observational studies place more emphasis on mid- and long-term results. In either case, if cancer events occur in a nonlinear fashion with early dropout of individuals at risk for cancer, a conclusive analysis, even after several years of follow-up, may be unable to capture such a signal.

A crucial clinical question is whether patients with pre-existent cancers should be exposed to anti-TNF or other immunomodulatory therapies. Patients with pre-existent malignancies are generally excluded from clinical trials, and in clinical practice, clinicians may be reluctant to treat such patients with anti-TNF therapy, resulting in channeling of treatment with these agents toward low-risk cohorts. An analysis from the British Biologics Registry detected no increased risk of recurrent cancer in patients with pre-existent malignancy (IRR, 0.53; 95% CI, 0.22 to 1.26). Also, data from the German Biologic Registry revealed no significantly increased risk of recurrence in patients with a previous malignancy treated with anti-TNF agents (IRR, 1.4; 95% CI, 0.5 to 5.5). However, very few events were included in these analyses, so definite conclusions about overall or cancer-specific risks in individual patients cannot be drawn.

**Rituximab**

B cells appear to be involved in generation of antitumor responses and are important in maintaining inflammatory states, which promote carcinogenesis and tumor growth. The absence of B cells, for example, in hypogammaglobulinemia, has not been associated with increased susceptibility to cancers, and B cell depletion with rituximab has been shown to slow the growth of solid nonhematopoietic murine tumors.

Pooled analysis of safety data from patients with rheumatoid arthritis treated with rituximab in randomized controlled trials with more than 5000 patient-years of exposure revealed an incidence of malignancy excluding nonmalignoma skin cancer of 0.84 per 1000 patient-years (SIR, 1.05; 95% CI, 0.76 to 1.42). The incidence appeared to be stable over multiple courses of rituximab, and no unusual pattern of malignancy type was observed.

**Abatacept**

Abatacept is a fusion protein consisting of the extracellular domain of human cytotoxic T lymphocyte–associated antigen-4 (CTLA-4) linked to the Fc portion of human immunoglobulin. CTLA-4–mediated T cell suppression has been suggested to be important in the pathogenesis of several types of malignancies, especially malignant melanoma. Because abatacept blocks the same signal that new anticancer treatments based on this interaction are intended to enhance, theoretically a potentially increased risk of malignancies may be seen with the use of this biologic. So far, however, no signal for increased malignancies has been noted in patients treated with this agent. However, as with other contemporary trials, the abatacept trials in rheumatoid arthritis have excluded patients with a history of
cancer, and some studies have required negative screening mammograms before administration.

**Tocilizumab**

Tocilizumab is a humanized monoclonal antibody to the interleukin (IL)-6 receptor. IL-6 has an important role in inflammation and in the promotion of various types of malignancies, including suppression of apoptosis, promotion of angiogenesis, and induction of genes that mediate cell proliferation. IL-6 antagonists have been used in advanced multiple myeloma, with some evidence of efficacy. Data from studies of rheumatoid arthritis, the disease in which experience with this agent is most extensive, so far have not revealed any signals of an increased incidence of cancer during the 6-month controlled period of randomized clinical trials or follow-up. Labeling information for tocilizumab contains a general statement that “treatment with immunosuppressants may result in an increased risk of cancer,” but no specific warnings are included.

**Anakinra**

The IL-1 receptor antagonist, anakinra, has been best studied in rheumatoid arthritis. As is the case for most studies of biologic response modifiers, methotrexate usually has been administered concomitantly. The case rate for the development of malignant lymphoma with anakinra is 0.12 cases per 100 patient-years, with 8 cases of lymphoma observed among 5300 patients with rheumatoid arthritis treated with this drug in clinical trials for a mean of 15 months. This represents a 3.6-fold higher than expected rate of lymphoma compared with the general population. The SIR for lymphoma in this study (3.71; 95% CI, 0.77 to 11.0) is consistent with odds ratios from other studies of patients with rheumatoid arthritis. A number of solid tumors have also been reported with this agent. Whether anakinra is carcinogenic is unclear but cannot be excluded. The overall incidence of malignancies with anakinra use in rheumatoid arthritis is consistent with expected rates reported in the U.S. National Cancer Database.

**Cancer Screening in Patients with Rheumatic Disease**

Evidence of cancer risk in patients with rheumatic diseases forms the basis for clinically useful recommendations regarding cancer screening. First, with respect to management of the inflammatory disease, it is imperative to achieve optimal disease control and the lowest level of clinical disease activity possible using the least intensive treatment regimen available. Second, patients for whom immunomodulatory therapy including nb-DMARDs and biologic DMARDs are being contemplated should undergo routine cancer screening that is appropriate to their age, sex, familial cancer burden, and risk factors such as smoking. Third, because cancers may develop at an accelerated rate in the first few months to the first year or so of treatment, patients should be seen at frequent intervals and closely questioned and examined for signs and symptoms of malignancy, especially during this initial treatment period, and throughout the course of their disease.

Routine blood counts and differential blood cell counts should be performed at the initiation of treatment and as appropriate for the specific drug therapy used. Age- and gender-appropriate cancer screening for colorectal cancer, prostate cancer, breast cancer, and cervical cancer is advisable. In patients with particularly high cancer risk, such as dermatomyositis, assessment for tumor markers such as CA-125 and radiographic imaging of chest, abdomen, and pelvis may be appropriate yearly in the first year or two of the disease, and then as otherwise clinically indicated.

Especially patients taking alkylating agents such as cyclophosphamide may be at particularly high risk of cancer. These patients certainly should undergo routine cancer screening Pap smears and urinalyses for at least 15 years from cyclophosphamide therapy. With these considerations, the morbidity and mortality experienced by patients with rheumatic disease can be favorably managed.

**CONCLUSION**

Assessment of risk for malignancy in patients with rheumatic diseases is complex. Some rheumatic diseases such as dermatomyositis and Sjögren’s syndrome appear to confer a particularly high risk of cancers, particularly lymphoproliferative disorders. Many of these diseases are relatively rare, so that large patient cohorts required for more precise assessment of risk are not available or must be studied over long periods of time to develop stable risk estimates.

Malignancy and perineoplastic syndrome should be considered when patients present with musculoskeletal symptoms caused by an underlying malignancy, or noted as symptoms and signs associated with the presence of an underlying malignancy in a patient with a pre-existent autoimmune disease.

Most of the agents used in the treatment of these diseases are purposefully employed to modulate the immune response, and some, including the alkylating agents, are known carcinogens. Others may modulate immune response to decrease tumor surveillance. Further complicating the assessment are the individual susceptibility host factors, including the presence of oncogenic genes such as Bcl-2, and family history and environmental factors such as viruses, which may enhance the carcinogenic potential of the treatments.

Recommendations for treatment must include a general understanding of the disease and treatment-related malignancy and individualized discussion with the patient regarding risks and benefits of treatment as they relate to disease activity and severity and, most important, patient preference.

**References**

BRIEF HISTORY OF REHABILITATION IN ARTHRITIS

Early arthritis rehabilitation focused on bed rest, splinting, and gentle range-of-motion exercises. Physicians and rehabilitation specialists believed that physical activity and strenuous exercise would produce pain, increase joint swelling and temperature, and accelerate joint damage. In the 1940s, with the introduction of steroids, highly efficient and potent anti-inflammatory drugs, the rehabilitation focus shifted toward splinting and mobilizing patients with assistive devices to promote function. Surgical approaches such as joint replacements were used in the 1960s and 1970s to "stop" the progression of disease. At this time, rehabilitation interventions focused on postoperative protocols to enable patients to regain function and maximize independence. Through the 1970s and 1980s, with the proliferation of disease-modifying antirheumatic drugs (DMARDs) in clinical practice, rehabilitation research focused on evaluation of the effects of various intensities, frequencies, and modes of strengthening exercises on patient outcomes. From the mid-1990s onward, studies of aerobic exercise illuminated the benefits of aerobic conditioning for cardiovascular function without deleterious effects on joints and soft tissue. In the 21st century, with the development of advanced radiologic techniques and the advent of biologic therapies, researchers are investigating the impact of weight-bearing activities on joint integrity in people with arthritis and are embracing a public health perspective through promotion of community physical activity programs designed to improve quality of life and function.

GOALS OF REHABILITATION, REHABILITATION TEAM MEMBERS, AND MODELS OF TEAM CARE

Rehabilitation interventions address all aspects of the patient’s condition with the aim of maximizing function and
independence. As such, interventions range from prescription of adaptive devices, such as splints, orthotics, and ambulatory devices, to use of physical modalities, such as ultrasound, heat, and cold, to instruction in exercise and self-management strategies (relaxation, proper rest). Rehabilitation specialists are educated to assess environmental factors so they can best address barriers and facilitators impacting their patients.29

The initial phase of the rehabilitation process involves a comprehensive assessment of all dimensions of the patient’s life and condition. Given the fluctuating course of many arthritides, a coordinated effort by a variety of skilled multiprofessional rehabilitation specialists is required. The structure, function, and resources of the medical system in which the patient resides influence the composition of the team.30,31 Psychiatrists, or rehabilitation medicine physicians, are educated in medicine and rehabilitation to treat patients and refer to, and/or supervise, other skilled rehabilitation professionals.29 These physicians typically are responsible for the care of arthritis patients. However, in some circumstances, the rheumatologist may be the only physician involved in the patient’s care. Other rehabilitation professionals involved in the care of these patients are primary care physicians, nurse practitioners, nurses, physical therapists, occupational therapists, social workers, nutritionists, psychologists, podiatrists, and vocational rehabilitators, who work with the patient’s family.

Comparative studies of team care delivery versus individual practitioner models indicate that coordinated team efforts yield better outcomes.32-34 In all cases, the patient is the focal point of the team. Traditional models of team care are categorized as interdisciplinary, multidisciplinary, or transdisciplinary. In the interdisciplinary model, each professional conducts an independent patient evaluation and shares this information during team meetings to facilitate the development of integrated team goals. Negotiation and collaboration are prevalent principles. Multidisciplinary team care does not foster intercommunication between professionals in a coordinated manner.35 Rather, each professional conducts an examination and develops his/her goals for care in concert with the patient, and separately documents findings. The transdisciplinary model allows for the transference of professional roles. This model crosses professional boundaries (e.g., a physical therapist can do some functions of a nurse).36-38 Any of these models of care may occur in an inpatient or outpatient setting.

Although team care has been widely used in Europe, this is not the case in North America, in part because of budgetary constraints and shortages of health care professionals, especially in rural areas.39 Consequently, innovative care models such as the clinical nurse-specialist model,17 the primary therapist model,17 and telemedicine have emerged.

INTERNATIONAL CLASSIFICATION OF FUNCTIONING, DISABILITY, AND HEALTH: A FRAMEWORK FOR REHABILITATION MANAGEMENT

The International Classification of Functioning, Disability, and Health (ICF) of the World Health Organization provides a framework for the development and implementation of rehabilitation interventions.1 The ICF systematically organizes various aspects of an individual’s health condition. The term functioning is used to describe body functions, activities, and participation; the term disability refers to impairments, activity limitations, and participation restrictions. The ICF also considers environmental and personal factors that interact with body functions, body structures, activities, and participation.1 This framework integrates aspects of the biopsychosocial medical model with an ecological perspective. The equal emphasis on ecological factors, such as personal and environmental contextual factors, helps rehabilitation specialists identify and address elements that may facilitate or present barriers to obtaining independence. The ICF definitions serve as a common language for providers to communicate about patients’ health conditions, interventions, and performance of activities and participation in society1-8 (Figure 38-1). Researchers use the ICF to solidify and examine core constructs relevant to clinical studies of arthritis rehabilitation.3,6,7 These core sets have been compared with well-validated clinical outcome measures and help to identify the most salient outcome measures for use in rehabilitation clinical trials.3,9

ASSESSMENT TOOLS AND THE REHABILITATION CYCLE

The first step in rehabilitation is problem identification based on a comprehensive history and physical examination using reliable and valid health measurements. Results from such measurements are used in turn to develop the intervention and to evaluate outcomes. Rehabilitation health professionals use a wide variety of assessments, including the following:

- Technical measures: electrophysiologic, biomechanical, and computerized devices
- Clinical tests: ligament laxity tests, range-of-motion and strength testing
- Performance measures: gait velocity, mobility tests
- Patient-centered measures: patient and proxy self-reports on health status, quality of life, and health preferences

Because the primary goal in rehabilitation is to restore function and enable a return to normal life, the emphasis of a rehabilitation examination is on assessment of functioning and societal participation.15

Body structures and function elements such as joint motion may be assessed subjectively through simple observation, goniometry, or high-speed cinematography. Goniometry is most commonly used in the clinical setting because it is inexpensive, is easy to perform, and is psychometrically sound. Muscle strength is frequently assessed using manual muscle testing procedures, although its reliability and validity varies by joint location and by disease. Quantitative methods of maximal isometric strength measurement with hand-held dynamometers are reliable in patients with inflammatory arthritis35,40,41 and in those with degenerative lumbar spinal stenosis.42 Hand-grip strength can be measured reliably with a hydraulic hand dynamometer.43
Nonpharmacologic interventions used to manage arthritis symptoms include rest (total body or local), massage and trigger point techniques, exercise, assistive devices, orthotics and splints, counseling, education and self-management, manual therapy/mobilization, gait training and instruction in ambulatory devices, mobility devices, ergonomic modifications, and vocational rehabilitation. Table 38-1 provides a description of each of these interventions and their role in the management of arthritis. Among these, exercise, patient education programs, and self-management interventions are the best studied and the most effective. Across
Electrotherapy

**Superficial heat/cold therapy**

Radiation (infrared light) and conduction (hot packs, paraffin, or water) are mechanisms for superficial heat/cold generation that are applied for 20 minutes. Superficial heat increases the pain threshold, decreases muscle spasms, and produces analgesia by acting on free nerve endings. **Note:** AVOID heat therapy in the presence of acute inflammation or in persons with altered sensation. Cold therapy may be applied using ice packs, ice massage, vapocoolant sprays, or cold water baths. Cold therapy induces superficial and intra-articular tissue vasoconstriction, reduces local metabolism, and slows nerve conduction, thereby reducing pain and inflammation. Complications of cryotherapy include frostbite, cold-induced urticaria, and nerve damage. Use cryotherapy with caution in patients with Raynaud’s phenomenon or with cryoglobulinemia. Although evidence for physiologic benefits of heat and cold therapy indicates small effects,5 patients report psychological benefits.

**Electrotherapy**

Uses electricity to stimulate nerves and muscles and to alleviate pain. Surface electrodes are the common transfer medium. Only electro-acupuncture and dorsal horn stimulation use needle electrodes percutaneously. Electrotherapy uses direct continuous galvanic currents and modulated direct currents. Galvanic currents decrease pain conduction in slow unmyelinated nerve fibers (C fibers) to reduce pain. Modulated middle-frequency electrotherapy results in inhibition of pain-related potentials at spinal and supraspinal levels. Electrical stimulation of fast-conducting myelinated nerve fibers can partially decrease pain through inhibition of pain impulses carried more slowly by unmyelinated fibers. Faster impulses arrive at the level of the dorsal horn first and “close the gate.” Transcutaneous electrical nerve stimulation (TENS) is used for musculoskeletal pain, posttraumatic or postsurgical pain, peripheral nerve injury, neuropathic pain, and sympathetically mediated pain. **Note:** Electrotherapy is contraindicated in patients with cardiac pacemakers or implanted cardiac defibrillators, and is used with caution in patients with atrophic skin. A systematic review of randomized controlled trials of TENS in arthritis reports inconsistent results in managing RA symptoms1 and low back pain, and some benefits for pain and knee stiffness in patients with knee osteoarthritis.76

---

### Table 38-1 Definitions, Descriptions, and Purposes of Common Rehabilitation: Interventions to Manage Symptoms of Arthritis

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy conservation/total body rest</td>
<td>Important with inflammatory rheumatic diseases to have intermittent rests during the day (half hour) and 8-10 hours of rest per night.67</td>
</tr>
<tr>
<td>Local rest/resting splints</td>
<td>Resting splints (worn at night or during periods of total body rest to prevent joint movement) provide local rest to joints and maintain joint alignment. In RA and other systemic inflammatory disorders, resting wrist splints are well tolerated and effective.68</td>
</tr>
<tr>
<td>Manual therapy/joint mobilization/ manipulation</td>
<td>Performed as low- or high-velocity, small- or large-amplitude passive movement techniques. Flexibility and strengthening exercises follow manual therapy/mobilization to gain full benefit. Evidence is strongest in treating hip and low back disorders, but effects are small.67</td>
</tr>
<tr>
<td>Trigger point therapy</td>
<td>Used for muscle pain. Consists of ischemic compression of trigger points, followed by isotonic contractions or muscle tissue stretching. Additional techniques include myofascial release, trigger point injections, trigger point dry needling, and intramuscular manual therapy. A local twitch response, provocation of referred pain, and subsequent relaxation of the taut band indicate successful application.26</td>
</tr>
<tr>
<td>Massage</td>
<td>Increases local blood and lymph flow, facilitates muscle relaxation, and reduces muscle stiffness, pain, and spasm. Techniques include gliding, kneading, deep friction, and percussion. Gliding and kneading reduce muscle tension, improve circulation, and decrease edema. Friction breaks up adhesions. Lymph drainage increases lymph flow and decreases edema.67 Massage is contraindicated over malignant tumors, open wounds, thrombophlebitis, and infected tissues. Lymph drainage should be avoided in patients with congestive heart failure.26,69</td>
</tr>
</tbody>
</table>

#### Exercise

| Range-of-motion (ROM) and flexibility exercises | Maintain joint movement and function and may be performed passively (by therapist) or may require active patient participation. In active-assisted exercises, the patient exerts some force with joint movement but is assisted by the therapist. Active ROM exercises require the patient to exert muscle effort to achieve the desired ROM.35,68 Flexibility or stretching exercises enhance extensibility of muscle tissue. Stretching is best performed using gentle, smooth movement, and then holding the stretch for 2-15 seconds. |
| Isometric (static) exercise | Muscle contraction performed without a change in joint range or muscle length. These exercises produce less strain on joints than is produced by dynamic exercise.67 |
| Isotonic/dynamic/isokinetic exercise | Isotonic exercises require changes in muscle fiber length by elongating (eccentric) or shortening (concentric). These exercises involve movement through a fixed ROM at a fixed rate (velocity) against variable resistance. A machine provides resistance and rate of movement that matches exactly the force generated by the patient at any point in the range. This equipment can calculate the torque developed during the exercise activity.25,26 **Note:** AVOID in the presence of inflammation, popliteal cysts, and joint derangement. |
| Aerobic conditioning or endurance exercise | Provides cardiovascular pulmonary benefits, improves muscle strength, and reduces inflammation and weight. Modest effect sizes for these outcomes are achieved when exercises are performed at moderate levels of intensity for extended periods. Modes of aerobic exercise include walking, running, hiking, cycling, swimming, and stair climbing.25,33 |
| Aquatic exercise, spa therapy, balneotherapy | Based on physical properties of water (i.e., buoyancy, molecule adhesion, temperature), provide physiologic benefits such as muscle relaxation and ease of movement. Exercises performed in water allow buoyancy to support body weight and unload joints. Water adhesion properties may provide resistance when exercising. Diuresis and hemodilution are physiologic effects experienced with aquatic exercise. The recommended water temperature is 33° C–34° C (92° F–94° F). A systematic review of randomized controlled trials investigating balneotherapy has reported positive findings. However, the evidence was insufficient to allow formal conclusions about the efficacy of balneotherapy for patients with arthritis.26 |

#### Physical Modalities

| Superficial heat/cold therapy | Uses electricity to stimulate nerves and muscles and to alleviate pain. Surface electrodes are the common transfer medium. Only electro-acupuncture and dorsal horn stimulation use needle electrodes percutaneously. Electrotherapy uses direct continuous galvanic currents and modulated direct currents. Galvanic currents decrease pain conduction in slow unmyelinated nerve fibers (C fibers) to reduce pain. Modulated middle-frequency electrotherapy results in inhibition of pain-related potentials at spinal and supraspinal levels. Electrical stimulation of fast-conducting myelinated nerve fibers can partially decrease pain through inhibition of pain impulses carried more slowly by unmyelinated fibers. Faster impulses arrive at the level of the dorsal horn first and “close the gate.” Transcutaneous electrical nerve stimulation (TENS) is used for musculoskeletal pain, posttraumatic or postsurgical pain, peripheral nerve injury, neuropathic pain, and sympathetically mediated pain. **Note:** Electrotherapy is contraindicated in patients with cardiac pacemakers or implanted cardiac defibrillators, and is used with caution in patients with atrophic skin. A systematic review of randomized controlled trials of TENS in arthritis reports inconsistent results in managing RA symptoms and low back pain, and some benefits for pain and knee stiffness in patients with knee osteoarthritis.76 |

Continued
A multitude of devices may be used to assist function and reduce barriers to independence. Examples include orthotic devices such as braces, splints, corsets, collars, and shoe modifications. Orthotic devices restore or maximize function by altering biomechanics through stabilizing, realigning, and/or maximizing joint position, thereby reducing pain. Orthotics provide some pain relief but inconclusive evidence of improved function long term. Successful prescription and use of orthotics require the identification of functional limitations (on all ICF levels) and patient collaboration to adjust the orthotic. Devices can be simple and inexpensive, but may be specially designed and consequently expensive. Bracing is prescribed to stabilize joints. The most common are knee braces typically prescribed for OA.

Orthotics/braces: Orthotic devices include braces, splints, corsets, collars, and shoe modifications. Orthotic devices restore or maximize function by altering biomechanics through stabilizing, realigning, and/or maximizing joint position, thereby reducing pain. Orthotics provide some pain relief but inconclusive evidence of improved function long term. Successful prescription and use of orthotics require the identification of functional limitations (on all ICF levels) and patient collaboration to adjust the orthotic. Devices can be simple and inexpensive, but may be specially designed and consequently expensive. Bracing is prescribed to stabilize joints. The most common are knee braces typically prescribed for OA.

Dynamic splints: Dynamic splints maintain joint alignment and reduce pain during functional activities. The most common dynamic splints used in RA are functional wrist splints, although many others are available for the small hand joints. In OA, thumb splints are common. A Cochrane review from 2003 concluded that dynamic wrist splints significantly increase grip strength but do not impact pain, morning stiffness, pinch grip, or quality of life. No evidence suggests that resting splints changed pain, grip strength, or the number of painful or swollen joints. Patients preferred using these splints to nonuse.

Collars/corsets: Cervical collars include soft collars, Philadelphia collar, and sterno-occipitomandibular plaster immobilization. These provide varying levels of stability. None prevent subluxation or displacement. Patients with night pain resulting from a cervical disk syndrome may profit from wearing a soft collar at night. Corsets and abdominal binders provide feedback but limited stability to patients in terms of body position.

Assistive devices: A multitude of devices may be used to assist function and reduce barriers to independence. Examples include button hooks, long-handled reachers, sock aids, modified eating utensils (padded handles), bottle openers, and modified container lids. A physical or occupational therapist may recommend modified door handles, a raised toilet seat, a commode, safety bars on the bathroom wall, or a lift in the bath to ensure safety and maximize independence with hygiene activities.

Mobility devices: Mobility devices, such as canes, crutches (axillary, forearm, and platform), wheeled walkers, and wheelchairs, are used when walking is limited by lower extremity joint instability, pain, weakness, and fatigue or balance problems. These devices are easily accessible and provide immediate assistance, but they require more physical effort than is required for normal ambulation. The choice of a mobility device is based on impairments and resulting disability (e.g., patients with RA with bilateral upper extremity involvement and leg weakness may be safer ambulating with platform than axillary crutches). Patients with balance disturbances and leg weakness may be better suited for walkers. The device type also impacts the weight-bearing status. For example, a single-point cane can bear about 25% of body weight, axillary or forearm crutches about 50%, and a wheeled walker more than 50%. Wheelchairs are necessary if patients are unable to walk with devices. In these cases, a wheelchair can improve the quality of life by maintaining mobility. Persons unable to navigate a manual wheelchair can profit from an electric chair. Accidents can occur while traversing ramps, sidewalks, and streets.

Vocational rehabilitation: Vocational rehabilitation is an interdisciplinary approach that enables patients to acquire and maintain gainful employment and varies among countries. A common feature consists of problem assessment at work and the development of individual solutions. Studies demonstrate that vocational support prevents or delays work disability and improves fatigue and mental health.

Work hardening/functional restoration programs: These highly structured, goal-oriented, individually tailored programs are provided by an interdisciplinary professional team to address functional, physical, behavioral, and vocational needs to facilitate return to work. Work hardening programs bridge the gap between initial injury and return to work. Most programs include a formal worksite ergonomic to address workplace design issues and to reduce injury risk while improving function, as well as social interventions. The Commission on Accreditation of Rehabilitation Facilities defined and developed standards for work hardening practice to standardize program content and implementation. Work hardening programs integrate real or simulated work activities to model appropriate behaviors and assess functional, biomechanical, neuromuscular, cardiovascular/metabolic, behavioral, attitudinal, and vocational performance. A systematic review of these programs provides evidence for significant reductions in sick days.

Self-management and patient education: Patient education is defined as a set of planned educational activities used to improve a patient's health behaviors and/or health status. Most patient education programs are provider driven and developed. Self-management programs are patient oriented and patient active, and combine education with cognitive-behavioral strategies to influence patients’ attitudes toward disease and disease management.

Cognitive-behavioral therapy (CBT): CBT acknowledges that pain and its resulting disability are influenced by somatic pathology, and by psychological and social factors. In general, three behavioral treatment approaches are distinguished: operant, cognitive, and respondent. The primary focus is to reduce disability. A review of 21 studies found no significant differences between the various types of CBTs; also, CBTs did not differ from exercise. A conclusion regarding whether clinicians should refer patients with chronic low back pain to behavioral treatment programs or to active conservative treatment was not possible.

Table 38-1 Definitions, Descriptions, and Purposes of Common Rehabilitation: Interventions to Manage Symptoms of Arthritis—cont’d

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep tissue heating/ultrasound and diathermy</td>
<td>Parameters for application vary by the device; the device must be applied by a skilled provider owing to the risk of deep tissue burns of the skin, fat, muscles, and bones. Heating occurs mostly at tissue interfaces (e.g., bone–soft tissue interfaces). Ultrasound may penetrate 7-8 cm of fat, but less than 1 mm of bone, depending on the energy level and frequency chosen. In practice, ultrasound with frequencies of 0.1-1 MHz can increase temperature by 4°C-5°C at depths of 7-8 cm. Two systematic reviews of ultrasound reported small benefits for pain, knee and hand stiffness, grip strength, and tender and swollen hand joints. Note: Avoid deep heating in patients with altered sensation, implants, or history of cancer. Ultrasound must be applied using continuous movement over joints and bony surfaces or in a water bath to avoid heating of bone.</td>
</tr>
<tr>
<td>Devices to Stabilize and Protect Joints</td>
<td></td>
</tr>
<tr>
<td>Orthotics/braces</td>
<td>Orthotic devices include braces, splints, corsets, collars, and shoe modifications. Orthotic devices restore or maximize function by altering biomechanics through stabilizing, realigning, and/or maximizing joint position, thereby reducing pain. Orthotics provide some pain relief but inconclusive evidence of improved function long term. Successful prescription and use of orthotics require the identification of functional limitations (on all ICF levels) and patient collaboration to adjust the orthotic. Devices can be simple and inexpensive, but may be specially designed and consequently expensive. Bracing is prescribed to stabilize joints. The most common are knee braces typically prescribed for OA.</td>
</tr>
<tr>
<td>Dynamic splints</td>
<td>Dynamic splints maintain joint alignment and reduce pain during functional activities. The most common dynamic splints used in RA are functional wrist splints, although many others are available for the small hand joints. In OA, thumb splints are common. A Cochrane review from 2003 concluded that dynamic wrist splints significantly increase grip strength but do not impact pain, morning stiffness, pinch grip, or quality of life. No evidence suggests that resting splints changed pain, grip strength, or the number of painful or swollen joints. Patients preferred using these splints to nonuse.</td>
</tr>
<tr>
<td>Collars/corsets</td>
<td>Cervical collars include soft collars, Philadelphia collar, and sterno-occipitomandibular plaster immobilization. These provide varying levels of stability. None prevent subluxation or displacement. Patients with night pain resulting from a cervical disk syndrome may profit from wearing a soft collar at night. Corsets and abdominal binders provide feedback but limited stability to patients in terms of body position.</td>
</tr>
<tr>
<td>Assistive devices</td>
<td>A multitude of devices may be used to assist function and reduce barriers to independence. Examples include button hooks, long-handled reachers, sock aids, modified eating utensils (padded handles), bottle openers, and modified container lids. A physical or occupational therapist may recommend modified door handles, a raised toilet seat, a commode, safety bars on the bathroom wall, or a lift in the bath to ensure safety and maximize independence with hygiene activities.</td>
</tr>
<tr>
<td>Mobility devices</td>
<td>Mobility devices, such as canes, crutches (axillary, forearm, and platform), wheeled walkers, and wheelchairs, are used when walking is limited by lower extremity joint instability, pain, weakness, and fatigue or balance problems. These devices are easily accessible and provide immediate assistance, but they require more physical effort than is required for normal ambulation. The choice of a mobility device is based on impairments and resulting disability (e.g., patients with RA with bilateral upper extremity involvement and leg weakness may be safer ambulating with platform than axillary crutches). Patients with balance disturbances and leg weakness may be better suited for walkers. The device type also impacts the weight-bearing status. For example, a single-point cane can bear about 25% of body weight, axillary or forearm crutches about 50%, and a wheeled walker more than 50%. Wheelchairs are necessary if patients are unable to walk with devices. In these cases, a wheelchair can improve the quality of life by maintaining mobility. Persons unable to navigate a manual wheelchair can profit from an electric chair. Accidents can occur while traversing ramps, sidewalks, and streets.</td>
</tr>
<tr>
<td>Vocational rehabilitation</td>
<td>Vocational rehabilitation is an interdisciplinary approach that enables patients to acquire and maintain gainful employment and varies among countries. A common feature consists of problem assessment at work and the development of individual solutions. Studies demonstrate that vocational support prevents or delays work disability and improves fatigue and mental health.</td>
</tr>
<tr>
<td>Work hardening/functional restoration programs</td>
<td>These highly structured, goal-oriented, individually tailored programs are provided by an interdisciplinary professional team to address functional, physical, behavioral, and vocational needs to facilitate return to work. Work hardening programs bridge the gap between initial injury and return to work. Most programs include a formal worksite ergonomic to address workplace design issues and to reduce injury risk while improving function, as well as social interventions. The Commission on Accreditation of Rehabilitation Facilities defined and developed standards for work hardening practice to standardize program content and implementation. Work hardening programs integrate real or simulated work activities to model appropriate behaviors and assess functional, biomechanical, neuromuscular, cardiovascular/metabolic, behavioral, attitudinal, and vocational performance. A systematic review of these programs provides evidence for significant reductions in sick days.</td>
</tr>
<tr>
<td>Self-management and patient education</td>
<td>Patient education is defined as a set of planned educational activities used to improve a patient’s health behaviors and/or health status. Most patient education programs are provider driven and developed. Self-management programs are patient oriented and patient active, and combine education with cognitive-behavioral strategies to influence patients’ attitudes toward disease and disease management.</td>
</tr>
<tr>
<td>Cognitive-behavioral therapy (CBT)</td>
<td>CBT acknowledges that pain and its resulting disability are influenced by somatic pathology, and by psychological and social factors. In general, three behavioral treatment approaches are distinguished: operant, cognitive, and respondent. The primary focus is to reduce disability. A review of 21 studies found no significant differences between the various types of CBTs; also, CBTs did not differ from exercise. A conclusion regarding whether clinicians should refer patients with chronic low back pain to behavioral treatment programs or to active conservative treatment was not possible.</td>
</tr>
</tbody>
</table>

ICF, International Classification of Functioning, Disability, and Health; OA, osteoarthritis; RA, rheumatoid arthritis.
diseases, exercise interventions, whether dynamic or static, yield moderate improvements (effect sizes of 4 to 6) in strength, pain, and function, and produce small to moderate improvements in mood, quality of sleep, sleep patterns, and psychological well-being.\textsuperscript{16,17,61-65,77-81} Exercise also reduces inflammation.\textsuperscript{82-84}

Patient education is an integral component of rehabilitation management. Patient education programs provide structured information about disease, medications and their use, when and how to perform exercise, disease management strategies (e.g., energy conservation, rest), and community resources. These programs are distinct from self-management interventions in that they are often provider driven and formulated, and focus on information dissemination versus behavioral interventions. For example, ankylosing spondylitis education programs educate patients about their disease, medication use and side effects, importance of good posture, extension exercises, activities in life that promote extension (playing on the floor with toddlers), and coping skills.

Self-management programs are patient focused, patient driven, and action oriented.\textsuperscript{85} These programs combine education, behavioral interventions, and cognitive strategies to enable patients to problem-solve, cope, and develop strategies to maximize function and independence. A focus of these programs is to assess and influence patients’ attitudes and beliefs about disease and their ability to manage their disease. Self-management programs demonstrate benefits with respect to disease outcomes\textsuperscript{85,86} and have proved cost-effective.\textsuperscript{87}

Evidence of the benefits of physical modalities, therapeutic relaxation techniques (massage, trigger point therapy) and splinting, orthotics, and gait deviation interventions varies across intervention type and disease. Overall, the evidence is weak.\textsuperscript{8,10,13,57} However, it is important to note that relaxation and physical modalities are adjunct therapies designed to prepare a patient for dynamic exercise, flexibility training, or gait training and are not stand-alone interventions.\textsuperscript{88} Additionally, many patients report satisfaction with and preference for these therapies.\textsuperscript{89}

**PRINCIPLES GUIDING REHABILITATION IN PEOPLE WITH ARTHRITIDES**

Research and clinical management outcomes suggest the need for the following guiding principles for rehabilitation in the care of persons with arthritis:

- Primary, secondary, and tertiary programs are an integral aspect of the care of persons with arthritis. Primary prevention focuses on preventing disease. Secondary prevention refers to detection of disease before it becomes symptomatic. Tertiary prevention, the most common in arthritis, focuses on interventions for persons already experiencing disease symptoms to ameliorate symptoms and to maximize function and independence.
- Exercise prescriptions for persons with systemic inflammatory diseases are developed on the basis of the disease state (active vs. inactive disease), disease severity, patient preferences and goals, medications used and their corresponding side effects, and latency periods.\textsuperscript{87}
- Persons with arthritis are less aerobically fit than their healthy counterparts.\textsuperscript{21,24,89}
- Asymptomatic heart disease may be disproportionately present in persons with systemic lupus erythematosus, rheumatoid arthritis, and other inflammatory conditions; therefore, heart rate and blood pressure monitoring is needed before exercise sessions and at periodic intervals during exercise and cool-down.\textsuperscript{87}
- Exercise response in persons with systemic inflammatory disease is altered, and careful monitoring of cardiovascular pulmonary response is needed during exercise testing and participation.
- Dynamic resistance strengthening exercises should be avoided in the presence of joint derangement and popliteal cysts.\textsuperscript{87}
- During periods of active myositis and elevated creatine phosphokinase (CPK) levels, activities should be limited to range-of-motion exercise and functional activities of daily living, and rest should be encouraged.

The Centers for Disease Control and Prevention (CDC) in conjunction with the American College of Sports Medicine (ACSM) has established guidelines for physical activity to promote the health of all Americans, including persons with arthritis (Table 38-2). The CDC deliberately framed these recommendations as minutes of exercise versus traditional physiologic parameters to enable consumers to better interpret activity expectations and attempt to meet these recommendations. Presently, physical inactivity among persons with doctor-diagnosed arthritis in the United States is a major public health issue; only 17.2% of people meet the CDC recommendations for moderate physical activity.\textsuperscript{71}

### Table 38-2  Centers for Disease Control and Prevention and American College of Sports Medicine Guidelines for Physical Activity

<table>
<thead>
<tr>
<th>Activity Type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 minutes of moderate physical activity</td>
<td>5 days a week</td>
</tr>
<tr>
<td>20 minutes of vigorous activity</td>
<td>3 times per week</td>
</tr>
</tbody>
</table>

*Moderate activity is defined as an increase in breathing or heart rate; perceived exertion in the range of 11-14 out of 20 on a perceived exertion scale\textsuperscript{59} or 3-6 metabolic equivalents of task; or any activity that burns 3.5-7 calories per minute (kcal/min).\textsuperscript{59}*

**REHABILITATION OF SELECT ARTHRITIDES**

In this section, select rheumatic conditions are discussed, along with evidence for rehabilitation interventions. Given the variety of rheumatic diseases, conditions that have similar attributes are grouped together, although they are distinct entities. In this manner, the rationale for interventions based on similar impairments, functional limitations, and activity participation can be delineated. Although not all diseases will be discussed, this section provides the intervention algorithms used in management of these conditions.
Common Impairments and Rehabilitation Interventions in Rheumatoid Arthritis and Inflammatory Arthritis

Both rheumatoid arthritis (RA) and psoriatic arthritis (PsA) are chronic systemic inflammatory rheumatologic disorders affecting connective tissues and joints. Of the two conditions, RA is more prevalent. Although the origin and prevalence of RA and PsA differ, similarities have been noted with respect to symptoms. For example, patients present with polyarticular symmetric joint involvement, malaise, morning stiffness, pain that is relieved with activity, muscle weakness, enthesitis and/or plantar fasciitis, and cardiovascular pulmonary involvement. Both diseases are characterized by exacerbations and remissions. Differences between these diseases are noted with respect to the joints involved, integumentary involvement, dactylitis, and predominant pathology. The primary pathology in PsA is enthesitis, and the axial skeleton may be predominantly involved. Chronic inflammation eventually may lead to spinal ankylosis, significantly impacting trunk range of motion and rib cage excursion. RA is a disease of the synovium that typically involves the hands, feet, wrists, and weight-bearing joints. Although RA does not affect most of the spine, subluxation of the atlantodental joint can occur.

Systemic features of these conditions lead to impairments such as fatigue, malaise, fever, sarcopenia, reduced aerobic capacity, restricted joint range of motion, postural instability, cardiovascular pulmonary dysfunction, and depression. Both conditions are associated with significant disability and early death. Medical management of these conditions includes the use of biologics, disease-modifying antirheumatic drugs (DMARDs), steroids, and nonsteroidal anti-inflammatory agents. Although responsiveness to DMARDs varies by disease, data suggest that patients with RA are more responsive to methotrexate, whereas those with PsA are more responsive to leflunomide. Rehabilitation professionals must be aware of indications of the need for these medications, potential side effects, and the latency period to effectiveness if they are to develop treatment plans that adjust as responsiveness to medications increases.

Given the fluctuating nature of both diseases, the selection of rehabilitation interventions varies on the basis of disease activity. Figure 38-2 provides an intervention algorithm that illustrates the different modes of exercise and rehabilitation interventions used to manage these conditions during acute, subacute, and chronic stages of disease. When RA or PsA is highly active, patients may report low-grade fever, fatigue, or malaise, and may present with warm, swollen, and tender joints. Adaptive shortening of soft tissues, joints, and tendons may occur secondary to inflammation and as the result of protective splinting of joints in response to pain. In RA, subluxation of the metacarpophalangeal and interphalangeal joints, as well as ulnar deviation of the fingers and radial deviation of the wrists, is prevalent, whereas with PsA, dactylitis may be present. With both conditions, soft tissue inflammation may lead to atrophy of the foot intrinsic muscles and loss or subluxation of plantar fat pads. Because of this, patients may present with plantar fasciitis and enthesitis. Myositis may also be present. Eventually, loss of cartilage may result, with severe disease and joint subluxation and derangement. These joint changes may lead to functional limitations, gait abnormalities, and deconditioning.

Interventions used to manage acute inflammatory symptoms include patient self-management strategies, gentle range-of-motion exercises, isometric (static) exercises, physical modalities, total body rest, and splints. The frequency and intensity of exercises are generally set at three to five repetitions, twice a day. To enhance adherence, patients are instructed to conduct daily joint checks, that is, a simple check of each joint to determine which are most limited and can serve as the focus of exercises for that day. Self-management interventions focus on energy conservation strategies. Patients are advised to sleep 8 to 10 hours per night and to take frequent rests during the day. Resting and dynamic splints are provided to maintain joint alignment, reduce pain, and control inflammation. Although resting splints are used when sleeping and serve to hold the joints in a neutral position, restricting all movement, dynamic splints are designed to allow joint movement and accommodate functional activities. During the acute inflammatory stage, cold therapy (cold packs, ice massage) is preferred over heat to reduce swelling and alleviate pain. However, thermal modalities (heat/cold) are contraindicated in persons with altered sensation and vascular conditions. It is important to evaluate the need for assistive/adaptive devices (e.g., long-handled reachers) and ambulatory devices to maximize function and independence. Ambulatory devices and orthotics are particularly important in the presence of foot involvement and deformities in that they redistribute weight and alter lower extremity biomechanics. Orthotics are also shown to improve function and reduce pain. When wrists and hands are involved, platform attachments to assistive devices are warranted. In all cases, patient needs to be elicited and considered when ambulatory devices are selected, to ensure adherence.

When symptoms subside (i.e., subacute stage), more aggressive interventions can be incorporated. The program may be modified to include increased repetition of range-of-motion exercises with progression from isometric to dynamic exercises (five to ten repetitions of each exercise, three times per day). However, if myositis or popliteal cysts are present, active resistive exercise should be avoided.
Flexibility exercises are initiated to reduce soft tissue tightness and contractures. Before stretching, heat therapy and massage may be used to reduce muscle spasm and enhance joint mobility. Clinical trials of thermotherapy report small effects for relief of muscle spasm, pain, and extensibility of soft tissue. Transcutaneous electrical nerve stimulation (TENS), laser, massage, and trigger point therapy provide inconsistent results but small benefits. Randomized controlled trials of exercise at moderate intensity have been conducted and have demonstrated moderate effects in improving strength, function, and mood state, and in reducing pain, stiffness, and inflammation, without deleterious effects on joints. Aquatic exercises are encouraged at this time to promote flexibility, improve gait patterns, maximize strength, and increase independence. The buoyancy of the water facilitates movement and reduces joint loading.

With chronic or stable disease activity, patients should transition from gentle range-of-motion exercises to dynamic strengthening exercises with resistance (eight to ten repetitions a day). Studies suggest that dynamic exercises can effectively increase muscle strength, physical capacity, and aerobic capacity without causing deleterious effects. It is imperative to initiate aerobic exercises (30 minutes, three to five times per week) to address cardiovascular risk, improve aerobic performance, and increase muscle strength. Patient counseling regarding general physical activity and active lifestyles should be emphasized to prevent secondary effects of inactivity. The most common modes of physical activity and exercise include low-impact exercise such as walking programs, aquatics, dance, and cycling, and dynamic exercises with resistance. Recent clinical trials of high-impact physical activities report significant improvements in aerobic capacity, grip strength, physical activity, anxiety, and depression. Benefits were maintained after 1 year. These exercises appear to be safe and effective, except in the presence of joint changes in large weight-bearing joints, joint derangement, and cartilage loss before initiation of weight-bearing activities.

Common Impairments and Rehabilitation Interventions in Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) is a systemic inflammatory disease that predominantly affects women in their childbearing years and is more common among African-Americans. Its cause is unknown. Systemic lupus erythematosus affects multiple organ systems, including the joints, skin, kidneys, heart, lungs, brain, and gastrointestinal tract. Fatigue is prevalent and may arise from myositis, anemia, depression, pulmonary disease, or deconditioning. Unlike in RA and PsA, joint involvement is uncommon, although if present, the small joints of the hands, especially the proximal interphalangeal joints, are most commonly involved, and recommendations similar to those for RA apply. Aseptic necrosis of the hip may occur and may require total joint arthroplasty. Instruction in the use of assistive devices (crutches, canes) may help to unload painful hip joints. With unilateral hip disease, use of the device on the contralateral side reduces acetabular loading and hip abductor muscle activity, thereby reducing joint hip load and pain.

Asymptomatic coronary heart disease is a major cause of mortality. The physical examination should include a comprehensive cardiovascular pulmonary system review, and interventions should focus on exercise modes to enhance cardiovascular performance, such as biking, walking, and dynamic exercises at moderate intensity (50% to 70% one repetition maximum). For safety, vital signs are assessed at baseline and at regular intervals during the exercise session. Patients who are severely deconditioned may benefit from interval training (short bouts of 10 to 15 minutes of exercise twice a day) to enhance aerobic conditioning. Self-management interventions are recommended to promote medication adherence and consistency in performing physical activities and exercise.

Studies of exercise in SLE are limited. In an earlier trial of endurance training using stationary cycles with individually tailored resistance settings based on submaximal exercise testing, patients with SLE demonstrated small to modest effects on function but no significant impact on VO2max (maximal oxygen consumption). A recent review of exercise for patients with SLE concluded that patients with mild to moderate SLE benefit from exercise of moderate to high intensity, and significant benefits can be demonstrated for aerobic capacity, fatigue, physical function, and depression.

Common Impairments and Rehabilitation Interventions in Osteoarthritis

Osteoarthritis (OA) is a progressive, multifaceted disease of the cartilage that can lead to cartilage and bone loss. Typical joints involved include hips, knees, spine, and hands (distal and proximal interphalangeal joints, first carpometacarpal, and first metatarsal phalangeal). Intrinsic and extrinsic factors, such as joint injury, malalignment, trauma, obesity, and quadriceps weakness, are associated with the development of OA. Patients typically note short-term morning stiffness (<30 minutes) or stiffness after prolonged sitting or inactivity. The combination of excess weight and poor joint biomechanics increases the risk of OA about threefold.

Patients typically report specific joint pain, muscle weakness, and restrictions in range of motion. Osteophytes and bone fragments may produce radicular pain if a nerve root is compressed and can limit movement. Herberden’s nodes at the medial and dorsolateral aspects of the distal interphalangeal joints and/or Bouchard’s nodes at the proximal interphalangeal joints may be present. Hip pain is frequently located at the joint line but may radiate to the groin, buttocks, anterior thigh, or knee. Reduced hip range of motion impacts function and hygiene. Knee osteoarthritis may present as knee instability, restrictions in joint range of motion, crepitus, pain with loading activities, or buckling of the leg with transfers and ambulation.

Common interventions to manage OA include dynamic strengthening exercises (both open and closed chain exercises), core stability exercises, education, balance activities, aerobic exercises (biking, swimming), joint mobilization and manipulation, bracing (severe cases), gait training, use of orthotics, and appropriate footwear. Thermotherapy including deep heat applications (ultrasound) may help to

Chapter 38: Introduction to Physical Medicine, Physical Therapy, and Rehabilitation
alleviate pain, enhance tissue extensibility, and reduce stiffness. Paraffin is useful for hand treatment because it provides a medium for dispersing heat over small joints. Exercise therapy can effectively improve function, reduce stiffness, and maximize independence. Recent reviews and synthesis of the literature on exercise for knee OA and hip OA, provided at varying intensities, frequencies, and durations, report significant moderate gains in muscle strength function (effect sizes ranging from 0.3 to 0.6). The emphasis of patient education is on joint protection techniques, activity modification, weight reduction, and strategies to unload the joint.

Manual therapy techniques for patients with knee OA include tibiofemoral and patellofemoral joint mobilization. Manual therapy interventions are followed by stretching of the quadriceps, hamstrings, tensor fascia, iliotibial band (ITB), hip flexors, and gastrocnemius soleus, and soft tissue mobilization techniques. These combined interventions yield benefits for joint range and function. Open-chain isometric exercises such as quadriceps exercises are used with progression to closed-chain weight-bearing exercises such as mini squats and step-ups. Balance activities such as exercises performed on balance boards, single-leg stance, and hop are designed to facilitate co-contraction, and coordination of muscle firing is a useful exercise technique for promoting knee stability. Sensorimotor interventions may provide necessary proprioceptive input, but their efficacy is still unknown. Lateral wedge orthoses in patients with medial compartment knee OA assist with altering biomechanical loading at the knee and are well tolerated. With progressive disease and instability, knee braces are used to decrease medial compartment pressures resulting from varus positioning. Systematic reviews of the effects of knee bracing report limited benefits for pain relief and function. However, it is important to note that adherence to these devices is problematic.

Hip OA often results in restrictions in internal hip rotation and impacts gait. Manual techniques for the hip joint aim to improve the elasticity of the joint capsule and surrounding muscles. The combination of mobilization and exercise appears to be more effective than exercise alone. Ambulatory aids (canes, crutches, walkers) are used to unload the joint and allow patients to maximize independence.

Common Impairments and Rehabilitation Interventions in Ankylosing Spondylitis

Ankylosing spondylitis (AS) is a systemic inflammatory disease that targets the sacroiliac and axial joints. Systemic features include fatigue, malaise, and sarcopenia, which results in reduced engagement in physical activity. Enthesiopathy is common and impacts gait and transfers. With persistent inflammation, ankyloses of the spine can occur, limiting trunk range of motion and rib cage excursion. AS may also involve the hip and shoulder joints. Cardiovascular pulmonary manifestations may be seen; when they are coupled with trunk and rib cage restrictions, as well as the potential for costochondritis, aerobic capacity may be dramatically impacted.

Rehabilitation interventions for patients with AS include physical modalities such as thermal therapy to relax soft tissues and prepare the patient for flexibility exercises, manual therapy, assistive and ambulatory devices, orthoses, and functional activities. The principles of physical modalities for persons with AS are similar to those for those with PsA. Flexibility exercises are used to promote extension and help maintain a neutral spine. Incorporating posture and flexibility exercises into daily routines can enhance adherence and maintain joint range. For example, patients can be encouraged to read the newspaper while lying prone on the floor or to take stretch breaks during the day to avoid long periods of time sitting in front of a computer. Given the potential for reduced rib cage excursion, deep breathing exercises are used, although their efficacy is unproven. Dynamic progressive strengthening exercises of the postural muscles (at moderate intensity) and motor control activities are useful for maximizing strength and function. A Cochrane review of three randomized controlled trials reported greater improvements in spinal mobility and patient global assessment among patients participating in supervised group exercise.

Aquatic exercises provide a nice alternative, and the buoyancy of the water can facilitate movement. However, studies of aquatic therapy for persons with AS are limited and, at best, show small effects. Ambulatory and assistive devices may be warranted as the disease progresses or to unload a painful hip. Patient education typically focuses on smoking counseling (cessation or prevention), joint protection, energy conservation, instruction in exercise, and coping strategies.

Strategies to Increase Adherence to Exercise and Rehabilitation

Rehabilitation interventions require active participation of the patient and a commitment to adhere to the regimen for benefits to be acquired. The team must always maintain a patient-centered focus and must engage the patient in mutual goal setting and problem solving. Clear, concise communication about treatment expectations and the establishment of small, progressive, and achievable milestones are important. Potential barriers to exercise such as fear of joint damage or exacerbation of pain, lack of social support for exercise, fatigue, lack of transportation to a pool, or cost make it imperative to develop an appropriate and acceptable intervention plan. With systemic inflammatory diseases, disease activity primarily influences the mode, frequency, and intensity of exercises prescribed, although other factors influencing intervention selection include disease severity, systemic manifestations and other comorbidities, and patient preferences. Self-management programs are a vehicle to promote adherence to exercise and active lifestyles and have been found to be highly effective. Linking patients to community resources such as Arthritis Foundation exercise programs (Walk with Ease, Fit and Strong!) allows patients to develop social networks to share their experiences in living with arthritis and provides a venue for exercise.

SUMMARY AND CONCLUSIONS

Rehabilitation interventions comprise a vast array of treatments. Physical modalities are adjuncts to exercise and self-management and provide small benefits with respect to pain relief. Evidence of the benefits of orthotics and bracing is weak, although adherence is a factor that influences results. Exercise is the most intensively studied intervention; it provides modest benefits for pain relief, strength, function, mood state, and joint stiffness across most disease entities. Selection of rehabilitation interventions for systemic inflammatory diseases is influenced primarily by disease state, but is also dependent on disease severity, medication latency periods, comorbidities, disease severity, and patient preferences.

References


The references for this chapter can also be found on www.expertconsult.com.
Pregnancy in the Rheumatic Diseases

ELIZA E. CHAKRAVARTY • MEGAN E. CLOWSE

KEY POINTS

The majority of women with autoimmune diseases will have healthy and uneventful pregnancies, and the risk-benefit ratios for medication use during pregnancy will improve pregnancy outcomes and reduce maternal and fetal risks.

The best outcomes result from conception occurring when the disease is in remission or clinically quiescent.

With few exceptions, clinical or laboratory parameters are unable to predict whether underlying disease activity will improve, worsen, or remain the same during pregnancy.

Potentially teratogenic medications should ideally be discontinued several months before conception.

At times women with underlying autoimmune disease require ongoing disease-modifying or immunosuppressive therapy during pregnancy to treat persistently active disease or flare-ups.

U.S. Food and Drug Administration pregnancy categories for medication safety for use during pregnancy are not routinely updated and may not accurately reflect increased experience during pregnancy.

Increased rates of premature delivery or small-for-gestational age infants may be seen in many autoimmune diseases.

The majority of autoimmune diseases display a female predominance, and, with the exception of giant cell arteritis, may present during the childbearing years. With improved treatment options for many autoimmune diseases that may prevent or retard functional disability or progressive organ damage from active disease, an increasing number of affected women will be able to face the physical challenges of childbearing and childrearing and may consider pregnancy as an important component of a fulfilling life. When considering pregnancy in women with underlying autoimmune diseases, one must take several things into account: (1) the effects of underlying disease on pregnancy outcomes; (2) effects of a pregnancy, with all of its immunologic and physiologic changes, upon an autoimmune disease; and (3) safe use of medications for effective control of autoimmune disease while avoiding adverse effects to the pregnancy or fetus.

In order for the “fetal semiallograft” to survive and grow during 9 months of exposure to the maternal immune system, the maternal immune system must undergo a complex modulation of the innate and humoral immune system, much of which is not well understood. Pregnancy had long been understood as a Th2-predominant condition, whereby a shift of helper T cells toward a Th2 dominant state, possibly induced by increasing levels of progesterone, is necessary to establish and maintain a normal pregnancy. This theory was consistent with earlier observations that systemic lupus erythematosus (SLE, a Th2-predominant disease) may be exacerbated by pregnancy, whereas Th1-mediated autoimmune diseases (rheumatoid arthritis [RA], multiple sclerosis, and psoriasis) appear to be characterized by clinical improvement during pregnancy. More recently, however, it is becoming increasingly clear that many more components of both the innate and adaptive immune systems are involved in normal pregnancy. Furthermore, many of the immunologic changes during pregnancy may be preferentially located at the maternal-fetal interface and may not be accurately sampled using peripheral blood. The trophoblast and placenta, once considered passive mediators of maternal-fetal immune trafficking, have been increasingly recognized as playing active roles in mediating inflammation, as well as in host defense. Taken together, it is evident that understanding the immune regulation of a healthy pregnancy remains elusive, much less understanding how pregnancy-related immunologic changes interplay with an abnormal immune system stemming from preexisting autoimmunity.

Nonetheless, it is imperative to go beyond inferring that individual diseases “tend” to behave in well-defined ways during pregnancy and look at an individual woman’s risk for adverse pregnancy outcomes. Fortunately, the majority of women with autoimmune diseases who desire children will have safe and relatively uncomplicated pregnancies. However, despite significant advances in the management of pregnancy in women with autoimmune diseases, one can argue that the rates of adverse pregnancy outcomes remain unacceptably high for some conditions.

Further complicating the issue of pregnancy in women with autoimmune diseases is the continuation or discontinuation of disease-modifying agents in anticipation of or discovery of a pregnancy. Some medications are known or suspected teratogens and should be avoided during pregnancy, whereas others may be considered safer to use with less evidence of embryotoxicity. As much as it may be desirable to manage disease during pregnancy with the minimum amount of medications necessary, the risks of medication exposure during pregnancy must be balanced with the risks of untreated active inflammatory disease and its potential for adverse pregnancy outcomes.

A few general principles apply to all women with autoimmune diseases who are contemplating pregnancy or who discover an inadvertent pregnancy. First and foremost, open communication between provider and patient is critical to all aspects of contraception and pregnancy planning. In the
ideal situation, family planning discussions are conducted before a woman becomes pregnant. Open and frank discussion of the risks of pregnancy outcomes regarding underlying disease activity and medication use should be continued throughout pregnancy. It has become increasingly evident that pregnancy outcomes are optimized when conception occurs during periods of disease quiescence and when disease control can be maintained throughout pregnancy. This requires preconceptional management of autoimmune disease using medications that are safe should they be required to be continued into pregnancy to maintain disease control. Additionally, comorbid conditions including hypertension and hyperglycemia should be well controlled before conception. Consultation with specialists in maternal-fetal medicine may be helpful for management of the pregnant patient with autoimmune diseases to ensure the proper monitoring for and management of disorders of pregnancy are undertaken in a timely manner. For the patient who discovers an inadvertent pregnancy during periods of active disease or with exposure to potentially teratogenic medications, counseling by both a rheumatologist and maternal-fetal medicine specialist should be offered so that the woman can make well-informed choices regarding her pregnancy. Healthy infants have been born with documented exposure to the most teratogenic medications, and patients may want to continue pregnancies despite early antenatal exposure.

SYSTEMIC LUPUS ERYTHEMATOSUS

Because SLE primarily affects women in their reproductive years and these patients have normal fertility, pregnancy is a frequently encountered dilemma. Fortunately, the majority of these pregnancies will progress without complication and result in the delivery of a healthy baby by a healthy mother. Almost half of these pregnancies, however, will encounter complications including increased lupus activity, pregnancy loss, preterm birth, and preeclampsia. The key to pregnancy success is careful planning for pregnancy, which includes timing to avoid conception during periods of disease activity and alteration of medications to avoid teratogenicity but maintain disease quiescence.

Pregnancy Outcomes

Retrospective studies of women with SLE have noted an increased rate of pregnancy loss when compared with healthy women. Among 203 women with SLE, 166 of their friends, and 177 of their relatives, pregnancy loss was twice as common (21% in SLE vs. 14% in friends and 8% in relatives, \(P < .0001\)) and preterm birth was three times more frequent (12% in SLE vs. 4% in friends and relatives, \(P < .0001\)).\(^9\) Prospective cohorts of SLE pregnancies report a more modest pregnancy loss rate, but these cohorts probably miss some early pregnancy losses that occur before presentation in the rheumatology clinic. What they notably show is a rate of stillbirth, typically defined as a pregnancy loss after 20 weeks gestation that is 5- to 10-fold higher than that of the general population (Figure 39-1). In a meta-analysis of 37 studies of SLE pregnancies (12 prospective and 25 retrospective), there is a higher than expected rate for miscarriage, stillbirth, and neonatal death (infant death at <28 days of life).\(^7\) Several studies have documented that increased SLE activity around the time of conception, hypertension, prior or current lupus nephritis, and antiphospholipid syndrome increase the risk for pregnancy loss by twofold to fourfold.

Preterm birth, defined as a birth before 37 weeks' gestation, occurs in about one-third of SLE pregnancies (Figure 39-2). Among live births, the meta-analysis found that 39.4% of deliveries were preterm.\(^7\) The cause of preterm birth in the majority of these pregnancies has not been well quantified, but it is frequently induced because of maternal SLE disease activity, preeclampsia, slow fetal growth, or fetal distress. Spontaneous preterm birth, either prompted by contractions and preterm labor or premature rupture of membranes, is also increased in SLE. In one carefully collected prospective cohort of SLE pregnancies, half of the pregnancies that survived past 23 weeks' gestation delivered preterm. Among these 33 preterm deliveries, 13 (39%) were due to premature rupture of membranes compared with 18% in a control population of preterm births (\(P = .001\)).\(^8\) The main predictor of preterm birth in women with SLE is increased disease activity during pregnancy.

Preeclampsia is the syndrome of hypertension and proteinuria that occurs in the second half of pregnancy and resolves postpartum. The etiology is not entirely understood, but the current main hypothesis is that impaired trophoblast invasion of the uterus early in pregnancy results in poor development of the spiral arteries that deliver blood...
from the mother, through the placenta, and to the fetus. Later in pregnancies, these arteries are not able to expand to provide sufficient blood flow to the fetus, resulting in a cascade of endocrine, inflammatory, and endothelial changes that promotes maternal hypertension and proteinuria. In the general population, preeclampsia typically occurs in 5% to 8% of pregnancies and only 1/10 of these present before 34 weeks’ gestation. The risk for preeclampsia is increased if a woman has previously had preeclampsia, particularly if it occurred early in pregnancy and was severe (blood pressure over 160/110, proteinuria over 5 g/24 hr, or other severe organ manifestations). Also at risk are first pregnancies and pregnancies in women with obesity, diabetes, hypertension, age older than 40, and a family history of preeclampsia. Up to one quarter of pregnancies in women with SLE are complicated by either hypertension or preeclampsia. In a meta-analysis of SLE pregnancies, 7.4% reported preeclampsia, but the rate varied greatly between studies. Preeclampsia can be difficult to distinguish from active lupus nephritis. Other signs of SLE activity including arthritis, rash, low complement, or rising double-stranded deoxyribonucleic acid (dsDNA)-titer can help determine the etiology of hypertension and proteinuria. Treatment of these two conditions is different: immunosuppression for SLE and delivery for preeclampsia. In some situations, treatment for both is prudent and the ultimate determination of cause will be possible postpartum either through resolution of preeclampsia over the first month after delivery or renal biopsy for persistent disease.

Placental insufficiency, which leads to less blood and nutrient flow to the developing fetus, can lead to small babies. Definitions of this vary, but it can be measured both in utero via ultrasound and at birth by birth weight and is typically defined as a baby weighing less than the 10th percentile of expected weight for gestational age. In the meta-analysis of SLE pregnancies, 12.7% of babies were noted to be small. In an analysis of the Nationwide Inpatient Sample, which includes data for 20% of all deliveries in the United States, 5.6% of pregnancies in women with SLE were identified as intrauterine growth restricted, a rate that was 2.6-fold higher than for other pregnancies in the United States (P < .01). In the Hopkins Lupus Cohort, 23% of all live births weighed less than the 10th percentile for gestational age of delivery.

Although up to 30% of women with SLE will have a flare during pregnancy, the majority of these are fairly mild and readily controlled. Several studies in the 1980s and 1990s tried to determine whether SLE worsens with pregnancy, but the conclusions were often conflicting. It seems that many patients will proceed through pregnancy without difficulty, but an important minority will suffer significant disease activity that can harm both the pregnancy and the woman’s own health and survival.

Significant SLE activity is associated with increased pregnancy loss and preterm birth. The degree of SLE activity is important—it appears that mild activity of the skin or joints will have minimal impact on pregnancy outcomes. Internal organ involvement including the kidneys, hematologic cell lines, serositis, and the central nervous system (CNS) is more highly associated with pregnancy morbidity. Women with highly active SLE around the time of conception have an estimated 40% risk of pregnancy loss. This applies to women with high overall disease activity, with isolated proteinuria greater than 500 mg/24 hr, or isolated thrombocytopenia. Disease activity in the second and third trimesters is less often associated with pregnancy loss but can double the rate of preterm delivery.

The most important predictors of SLE activity in pregnancy are increased SLE activity in the 6 months before conception, the frequency of SLE flare in the years before pregnancy, and the cessation of immunosuppressive medications. Among women with high-activity SLE in the 6 months before conception, 58% had high-activity SLE during pregnancy. On the other hand, of women with relatively quiet SLE before pregnancy, only 8% had high activity in pregnancy. The cessation of hydroxychloroquine (HCQ) before or concurrent with conception is associated with a high rate of SLE flare, particularly in the second and third trimesters. Because these flares are predominantly in the skin and joints or manifest as fatigue, they do not have a statistically significant impact on pregnancy outcomes. They can, however, lead to significant discomfort and the use of higher doses of corticosteroids during pregnancy. Stopping immunosuppressants including mycophenolate mofetil or azathioprine because of pregnancy may lead to a flaring of SLE activity, particularly renal or hematologic, that could lead to significant pregnancy morbidity. For this reason, women who require these drugs to maintain quiet SLE activity before pregnancy will likely benefit from continuing azathioprine during pregnancy (see further discussion in Medications in Pregnancy and Lactation later).

Lupus nephritis, whether active or quiescent during pregnancy, affects pregnancy outcomes. Several prospective studies of lupus nephritis in pregnancy document that women with prior lupus nephritis have a higher rate of pregnancy loss, preterm birth, and preeclampsia than other women with and without SLE. Active lupus nephritis during pregnancy increases the pregnancy loss and preterm birth rates by twofold to threefold. Renal failure is a rare complication of active lupus nephritis in pregnancy.

Maternal death is 20-fold more common among SLE pregnancies than other pregnancies, but it is fortunately rare. When compared with the annual risk of death in all women with SLE (ranging between 0.5% and 1% of all women, depending on the study sample), the risk of death in pregnancy for women with SLE is lower (0.3%). This is likely because women with the most severe SLE, who are at the highest risk for death, avoid pregnancy. Reported causes of maternal death in women with SLE include thrombosis, in particular pulmonary embolism, and infection. Although risk factors for maternal death have not been studied due to the rarity of this event, patients at higher risk for death include those with prior major thrombosis; severe pulmonary hypertension, lung, or renal disease; and women requiring high-dose immunosuppression.

Several simple interventions can be employed to decrease the likelihood of fetal and maternal morbidity:

1. Time pregnancy to coincide with periods of disease quiescence. At least 6 months of quiet SLE is required to decrease the risk of disease flare and adverse pregnancy outcomes. Accomplishing this requires discussing and prescribing effective contraception to women with significant SLE activity to avoid conception. Active SLE does not impair fertility.
2. Avoid teratogenic medications, but continue immunosuppression when needed. The adverse effects of highly active SLE generally outweigh the teratogenic effects of many SLE medications. Women taking HCQ before conception should continue this drug to avoid an SLE flare. Women who are taking mycophenolate mofetil or cyclophosphamide to control SLE activity should either avoid pregnancy or switch to azathioprine before pregnancy (or when pregnancy is discovered). Discontinue angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers (ARBs), but continue blood pressure control with calcium channel blockers and labetalol.

3. Use prednisone to treat SLE flare, but do not use prophylactically in a woman with quiet disease.

4. Start women with SLE on low-dose aspirin (81 or 100 mg a day) to decrease the risk of maternal thrombosis, preeclampsia, and preterm delivery.

5. Have all women followed by, or at least seen in consultation by, a high-risk obstetrician (also called maternal-fetal medicine).

6. Continue close rheumatology follow-up for women with SLE during pregnancy. Do not assume that the obstetrician will continue to monitor for or will be able to identify SLE activity.

**RHEUMATOID ARTHRITIS**

RA is a systemic, autoimmune disease that leads to chronic inflammation of the joints and other structures. Without use of disease-modifying agents, RA often leads to progressive deformities of the hands, feet, and larger joints. Over the past few decades, improved therapies have led to better control of disease and significantly slower progression of joint damage. With better control of short-term and long-term effects of the disease, people with RA may be more apt to consider childbearing and childrearing than in previous decades.

RA is a female-predominant disease, affecting women more than men in a 3:1 ratio. The prevalence of RA in the United States is approximately 1.4% with a median age of onset of 59 years. The prevalence of women of childbearing age in the United Kingdom (data on prevalence in the United States are not available) is estimated to be 1 to 2/1000 of women between ages 16 and 44.19 A recent study estimated that approximately 1400 women with RA gave birth in the United States in 2002.20,21

**Fertility**

Although it has been apparent for many decades that women with RA tend to have fewer children or are more likely to be nulliparous than control populations, there is a paucity of data looking at RA associated with fertility. Katz22 studied pregnancy patterns and decisions regarding childbearing in a group of 411 women with RA in 1999. Results showed that family sizes were smaller among women who developed RA before having any children or to completing childbearing. Approximately 30% of respondents who developed RA during the childbearing years reported that the diagnosis affected their decision to have additional children, and 12% reported being advised by a provider to limit family size. From these data, it appears that reduced numbers of children among women with RA may be related to choice rather than to reduced fertility or increased pregnancy loss.22 Reduced sexual activity due to functional limitations from the disease are a further, nonbiologic cause of lower birth rates due to reduced opportunity for conception.23

Although several disease-modifying antirheumatic drugs (DMARDs) may assert teratogenic effects on the fetus,24 and nonsteroidal anti-inflammatory drugs (NSAIDs) have been associated with temporal and reversible infertility,25 most chronic medications taken to manage inflammatory arthritides have no known adverse effect on hormonal, ovarian, or endometrial function. Several studies have found that estrogen-containing oral contraceptives yield a protective effect or mildly ameliorating effect on RA, and increased use of oral contraceptive pills among women with RA may further contribute to lower birth rates. These medication issues are easily overcome in women with RA who desire a pregnancy. However, it is much more difficult to formally study biologic infertility in this population. Few studies have been able to disentangle true inability to conceive a child from the myriad of other reasons for which women with RA may have fewer pregnancies. There has been a suggestion of reduced ovulatory function among women with RA in a small study of menstruating women,26 but this has yet to be followed up with larger studies. A recent population-based study from Norway has found that women with RA used assisted reproduction technologies more than age-matched healthy women.27 Similar findings were seen in a case-control study of pregnant women in the Netherlands (17.8% RA patients used assisted reproductive techniques compared with 3% in healthy women).28

**Pregnancy Outcomes**

In the absence of exposure to potentially teratogenic medications (including methotrexate, leflunomide, and mycophenolate mofetil [MMF]), earlier observations showed that maternal-fetal outcomes did not appear to be changed by maternal RA.20 However, more recent studies have shown that there are, in fact, some differences in these outcomes comparing women with RA with the general population. Several recently published population-based studies have documented an increased risk of preterm birth (<37 weeks gestational age) and small for gestational age (SGA, <10th percentile weight for gestational age) among women with RA compared with healthy women (Table 39-1). Although the odds ratios are generally less than 2.0, the consistency of reports over several different populations suggests a true association between a diagnosis of RA and adverse fetal outcomes. Because most of these studies are population based, relying on administrative data, they do not contain data regarding medication use or disease activity before conception or during gestation. Two prospective cohort studies of pregnant women with RA compared with healthy pregnant women have demonstrated an inverse association between disease activity during the third trimester of pregnancy and birth weight.29,30 One study from the Netherlands found a linear relationship between third-trimester disease activity (assessed by a modified DAS28-CRP) and birth weight in a multivariate model that adjusted for gestational
age, maternal smoking, maternal age, education, parity, and prednisone use. A second study, from the United Kingdom, found that women with active disease during the third trimester had lower birth weight than women with RA in remission; women with inactive RA had birth outcomes indistinguishable from healthy control women. Together, these data suggest that RA is associated with higher risks of preterm delivery and lower birth weight infants; however, this may be mitigated by achieving good control of disease activity during at least the third trimester of pregnancy.

### Disease Outcomes

Since the initial report by Hench in 1938, it has been thought that RA disease activity improves during pregnancy. In reviewing the majority of the data looking at disease activity during pregnancy, it appears that approximately 75% of patients will have some degree of improvement. This may occur as early as the first trimester and continue through the end of pregnancy. The exact underlying immunologic mechanism as to why this occurs is unclear. However, more recent reports have suggested that this improvement is actually not as dramatic as was once perceived. De Man and colleagues looked prospectively at 84 patients with RA and determined improvement and deterioration by the DAS28 changes. For women who started the pregnancy with moderate disease activity, only 48% had a moderate improvement response. Those with low disease activity during the first trimester tended to remain stable. Only 27% of the patients as a whole were in remission by the third trimester. In comparison, Barrett and Nelson and colleagues in previous studies showed that 16% and 39%, respectively, were in remission during the third trimester (though different definitions of remission were used). Unfortunately, no clinical or laboratory predictors have been found to reliably predict the disease course during pregnancy. Conflicting reports have been published looking at the degree of maternal-fetal human leukocyte antigen (HLA) class II mismatch and maternal RA disease activity. The majority of studies have confirmed that a higher degree of maternal-fetal HLA mismatch is associated with improved disease activity during pregnancy; however, a study of 110 maternal-fetal pairs failed to find such an association. More recently, a prospective study of 118 pregnant women with RA from the Netherlands found that women who were seronegative for rheumatoid factor or anticyclic citrullinated peptide had a greater likelihood of improvement in disease activity during pregnancy. This has yet to be studied in other populations.

Despite a wide range of disease activity during pregnancy, the majority of studies have shown that RA may worsen or flare in the postpartum period. In addition, a recent study has demonstrated an increase in the incidence of RA during the first 24 months following delivery compared with 25 to 48 weeks postpartum among premenopausal women in Norway, suggesting that either the incidence of RA peaks in the postpartum months or pregnancy may delay the onset of RA in susceptible individuals.

### Ways to Improve Outcomes

Monitoring disease activity during pregnancy can be tricky because distinguishing underlying activity of RA from normal changes during pregnancy is often not clear. Both erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) are markers of systemic inflammation that are often used to follow disease activity in rheumatic disease. Unfortunately, both these levels can fluctuate in pregnancy with more dramatic increases seen in ESR rather than CRP. Therefore these levels must be interpreted with caution in women with RA. Furthermore, clinical changes of pregnancy such as fatigue, swelling (especially of the distal extremities), and carpal tunnel syndrome are commonly seen in pregnancy. Careful examination of the wrist and assessment of neurologic symptoms are important to distinguish carpal tunnel syndrome from RA flare.

Given the growing evidence that active disease during pregnancy is associated with increased risks of adverse fetal outcomes, it is important to achieve control of disease activity before conception and during pregnancy whenever possible. In the ideal situation, women with RA considering pregnancy should aim to achieve good control of disease after discontinuation of medications with teratogenic potential, using medications considered safer during pregnancy. Determining the right medications for a pregnant patient can be challenging and requires an individualized approach. Certain medications are strictly contraindicated during pregnancy and should always be avoided due to teratogenic potential: methotrexate and leflunomide. Additionally, these agents should be discontinued immediately on the discovery of pregnancy in a woman receiving these medications (leflunomide will require an additional washout with cholestyramine). If a disease-modifying agent is required to control chronically active disease during an anticipated or confirmed pregnancy, plaquenil, sulfasalazine (SSZ), and/or azathioprine carry the fewest risks to the developing fetus.

### Table 39-1  Odds Ratios of Preterm Delivery and Small for Gestational Age Infants among Women with Rheumatoid Arthritis Compared with Healthy Controls in Population-Based Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Location</th>
<th>Preterm Delivery (OR)</th>
<th>95% CI</th>
<th>Small for Gestational Age (OR)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wallenius</td>
<td>2010</td>
<td>Norway</td>
<td>1.85</td>
<td>1.09-3.13</td>
<td>1.6</td>
<td>1.00-2.56</td>
</tr>
<tr>
<td>Chakravarty</td>
<td>2006</td>
<td>United States</td>
<td>1.00</td>
<td>0.98-1.00</td>
<td>2.2</td>
<td>1.2-4.1</td>
</tr>
<tr>
<td>Lin</td>
<td>2010</td>
<td>Taiwan</td>
<td>1.17</td>
<td>0.98-1.40</td>
<td>2.2</td>
<td>1.05-1.38</td>
</tr>
<tr>
<td>Reed</td>
<td>2006</td>
<td>United States</td>
<td>1.78</td>
<td>1.21-2.60</td>
<td>1.51</td>
<td>0.94-2.43</td>
</tr>
<tr>
<td>Norgaard</td>
<td>2010</td>
<td>Sweden/Denmark</td>
<td>1.44</td>
<td>1.14-1.82</td>
<td>1.56</td>
<td>1.20-2.01</td>
</tr>
</tbody>
</table>

CI, confidence interval; OR, odds ratio.
and are generally considered relatively safe. However, these agents may have limited efficacy for moderate to severely active disease and may be insufficient to control disease during pregnancy. The use or continuation of anti-tumor necrosis factor (TNF) agents during pregnancy has been more controversial. Some studies have suggested an increased risk for congenital abnormalities, whereas others have not. Cumulatively, the data seem to show that exposure to anti-TNF agents in pregnancy does not seem to increase risk of miscarriage, preterm delivery, or congenital malformations. In cases of severe inflammatory disease before or during pregnancy, TNF inhibitors may be used with caution, provided that the mother is aware of and comfortable with the risk-to-benefit ratio.

For acute flares of disease, use of glucocorticosteroids is the best option because they have a rapid onset of action and can be titrated to the level of disease activity. There is vast experience with glucocorticoids during pregnancy for many indications. Prednisone is perhaps the most ideal choice for the treatment of maternal disease because a very small amount of active drug will enter the fetal circulation. In fact, some women may even require regular glucocorticoid use if background DMARD therapy is ineffective. The goal should always be to minimize the dose as best as possible and to prescribe adequate calcium and vitamin D supplementation during use. Isolated joints may be best managed through direct local injection of steroids, thus avoiding significant systemic exposure to the drug. Although intermittent NSAID use appears to be safe, there are some risks to regular use, particularly during the latter half of pregnancy because it can be associated with premature closure of the ductus arteriosus.

### SCLERODERMA

The hallmarks of scleroderma, vasculopathy and fibrosis, complicate surprisingly few numbers of pregnancies in women with this disease. Pregnancies are relatively rare in women with scleroderma, primarily because the onset of the disease is often after a woman has completed her family. About 200 pregnancies in women with scleroderma have been reported in the literature, however, with largely good results. The rate of miscarriage is about 15%, which is 50% higher than in the general population. Stillbirths, however, are rare. Preterm birth rates range from 25% to 40%, dependent on the degree of internal organ damage. Reports of preeclampsia are rare, but in a nationwide database of pregnancies, 23% of the 504 with scleroderma had hypertensive complications, including preeclampsia.

To maintain pregnancy, a woman’s body must make major hemodynamic changes including an increase in blood volume of 50%, increase in cardiac output, and decreased vascular resistance. These adaptations may be hindered by underlying vasculopathy in women with scleroderma. Pulmonary hypertension, in particular, can both limit pregnancy success and put the woman at significant risk for death from hemodynamic collapse. Pregnancy in a woman with pulmonary hypertension carries a maternal mortality rate of 17% to 33% with deaths caused primarily by right ventricular failure or pulmonary embolism. The risk for death is greatest around the time of delivery and in the weeks postpartum. Neonatal or fetal death after 22 weeks’ gestation occurred in 7% to 13% of pregnancies complicated by pulmonary hypertension, about one-third had intrauterine growth restriction, and more than 85% delivered preterm. Medical treatment of pulmonary hypertension may include prostacyclin, nitric oxide, and calcium channel blockers. Due to the high maternal and fetal morbidity and mortality, pulmonary hypertension is considered a contraindication to pregnancy.

Scleroderma renal crisis, though rare in pregnancy (occurring in 2% to 3% of scleroderma pregnancies), can be catastrophic. It can be difficult to distinguish from preeclampsia, both presenting with hypertension, proteinuria, and in severe cases hemolysis and thrombocytopenia. Women at particular risk for this complication include those with a prior history of it and those with recent onset of rapidly progressive diffuse disease. If the pregnancy is near term, delivery followed by aggressive ACE inhibitor therapy is indicated. If not near term, treatment with ACE inhibitors should be weighed. These drugs carry significant risk for permanent and fatal renal disease in the fetus but are also the essential treatment for scleroderma renal crisis.

Gastroesophageal reflux disease often worsens in pregnancy. Raynaud’s phenomenon, however, often improves due to decreased vascular resistance. Skin disease does not seem to change significantly in pregnancy. The peripheral edema of pregnancy may be particularly uncomfortable, however, for women with diffuse scleroderma.

An approach to improving pregnancy outcomes in women with scleroderma begins with careful preconception counseling and continues with close monitoring during pregnancy.

1. Women with pulmonary hypertension, a history of scleroderma renal crisis, or severe interstitial lung disease should be made aware of the significant risk to their life, health, and the survival of a pregnancy due to their disease.
2. Women without known pulmonary hypertension should undergo screening with an echocardiogram and pulmonary function testing before or early in pregnancy to identify previously subclinical disease.
3. Women taking ACE inhibitors or angiotensin-receptor blockers should discontinue these prior to pregnancy, if possible.
4. Due to the high risk of preterm birth and preeclampsia, a daily dose of low-dose aspirin should be considered.
5. Women with scleroderma should be followed closely in pregnancy by rheumatology and a high-risk obstetrical team. In addition, cardiology or pulmonologists may need to be involved on the basis of the degree of cardiopulmonary disease.
6. Repeated screening for intrauterine growth restriction should be performed in the third trimester.

### PSORIATIC ARTHRITIS

Psoriatic arthritis (PsA) is a chronic destructive autoimmune arthropathy often associated with the presence of psoriasis. Psoriasis is a chronic inflammatory skin condition
manifested by hypertrophic erythematous plaques with overlying scale that can be quite disfiguring and functionally limiting for affected patients. Both psoriasis and PsA affect men and women equally; the average age of onset of psoriasis is in the second and third decades with PsA often occurring approximately 10 years later among the 20% who will go on to develop PsA.51

Pregnancy and Disease Outcomes

Outcomes data for women with PsA are scarce32; however, preconception counseling for patients with these conditions is similar to those with RA. They should have a good understanding of drug toxicities, side effects, and benefits of use as detailed earlier. In addition, it appears as though women can expect some degree of improvement in their arthritis symptoms during pregnancy. At least half of women show some resolution of their arthritis in case series reported to date.53,54 However, the degree of improvement has not yet been elucidated because more recent data reviewing pregnant PsA cohorts have not occurred. Furthermore, most case series and reports of pregnancy in PsA have occurred before the use of biologics and the era of better disease control, so it is difficult to say with certainty if these women actually clinically improve during as opposed to before pregnancy. As in RA, a postpartum flare often occurs within 3 months.53 Studies looking at pregnancy and fetal outcomes appear to show no increase in premature labor, preeclampsia, or adverse fetal outcomes.53,54

Ways to Improve Outcomes

Management of the arthritis in PsA during pregnancy is similar to that of RA. NSAID use sparingly, corticosteroids at low dose, and DMARDs that are not fetal toxic such as SSZ may be good options. Plaquenil is generally not commonly used in patients with PsA due to the thought that it may increase flares in skin disease. Similarly, corticosteroids may be relatively contraindicated in some patients with profound cutaneous disease. As with other conditions, it is important to discontinue potentially teratogenic medications including methotrexate and leflunomide before conception. The use of TNF inhibitors may be considered in cases of severe disease where the benefits of improved control of disease outweigh potential risks associated with antenatal exposure.

ANKYLOSING SPONDYLITIS

Ankylosing spondylitis (AS) is a chronic autoimmune disease that most often affects the axial skeleton, as well as the hips and shoulders, that is often, but not always, associated with the HLA-B27 allele. In contrast to most autoimmune diseases, AS displays a male predominance in a ratio of 4:1. The overall prevalence of disease in Caucasian populations is approximately 0.03% to 0.1%.55 Prevalence estimates in women of childbearing age are not available, but the incidence in Rochester, Minnesota, is 3.6 to 6.4 cases per 100,000 person-years.56 Treatment options for AS are much more limited than for RA, and most therapies do not appear to abrogate the progressive spinal fusion that is characteristic of the disease. Because of the relatively low incidence of AS in women of childbearing potential, data regarding pregnancy experiences in this population are sparse.

Pregnancy Outcomes

Unlike SLE and RA, underlying maternal AS does not appear to be associated with increased risks of maternal or fetal morbidity. A questionnaire-based study of 649 women with AS in North America and Europe57 was performed looking at pregnancy outcomes. It was found that 15.1% ended with miscarriage, a rate not dissimilar to the general population. The overwhelming majority (93.2%) of cases delivered at term, but 58% delivered by cesarean section with AS, rather than other pregnancy-related conditions, as the stated indication. Neonates were healthy with a mean birth weight of 3339 g. Studies looking at pregnancy and fetal outcomes appear to show no increase in premature labor, preeclampsia, or adverse fetal outcomes.

Disease Outcomes

Most women with AS who become pregnant either remain stable or may have some clinical worsening of symptoms during the course of pregnancy. In addition, at least half will have some aggravation of symptoms in the postpartum period as well.53,54 In a smaller study of nine patients with AS followed prospectively during pregnancy with validated disease activity measures (BASDAI scores, morning stiffness, patient global assessment), the majority had active disease during pregnancy, four had a decrease of 20% during pregnancy, and no remissions were found. Furthermore, women with AS had high pain scores throughout pregnancy when compared with women with RA, with resultant increased utilization of NSAIDs and paracetamol.58

Special considerations specific to AS should be considered. Bone loss is a well-recognized feature in AS. Osteoporosis is common and is largely related to disease activity, whereas vertebral fractures appear to be related to duration of disease and structural severity of the disease.60 The prevalence of vertebral fractures has been reported but variable.61 Cooper and colleagues reported an increased odds ratio of 7.7 (4.3 to 12.6).62 These fractures are often spontaneous, low-trauma fractures. Pregnancy can also lead to osteoporosis and/or bone density loss, which most often occurs in the lumbar spine in the peripartum period or within a few months after the birth of a child. Bone density loss related to pregnancy in combination with underlying bone density loss or loss of integrity due to AS may increase the risk of fracture even further and has been reported.64,65 This bone fragility should especially be considered during delivery when tremendous intrauterine pressure, most of which is transmitted to the perineum and lumbar spine, is created during active labor and pushing. If a patient has severe osteoporosis and/or poor spinal mobility, an elective
cesarean section may be considered to reduce risks associated with active labor.

**Ways to Improve Outcomes**

Management of AS during pregnancy is similar to that of RA. However, data regarding the relationship between active disease and adverse pregnancy outcomes are scarce due to fewer numbers of pregnant patients who have been studied. NSAIDs should be used as needed to improve function, but use should be limited or discontinued in the third trimester. Corticosteroids at low dose and DMARDs such as SSZ may be good options to treat peripheral disease but do not have much utility in treating axial symptoms. Use of anti-TNF therapy during pregnancy remains controversial at this point and should be used only with caution in the absence of more definitive data regarding pregnancy outcomes.

The mechanical variants that result in patients with AS are important to consider during delivery. AS often results in inflammation of and/or fusion of the pubic symphysis or sacroiliac (SI) joints. Hormones released during pregnancy such as relaxin cause the symphysis pubis and SI joints to soften and separate to some degree to prepare for the birthing process. For example, the nonpregnant gap of the pubic symphysis is 4 to 5 mm but in every pregnancy there will be an increase in 2 to 3 mm. If fusion interferes in adequate separation, it can create a mechanical obstacle during labor and delivery. In some cases, this may cause increased pain and/or lead to need for cesarean section delivery as opposed to vaginal birth. As noted earlier, Ostensen and colleagues reported that, compared with healthy women, caesarean section was more frequently performed in patients with AS and in 58% of the cases AS was the indication for cesarean delivery.51 Aside from potential difficulties with labor mechanics, these ligament and/or joint changes in the setting of underlying inflammation may worsen pain symptoms and increase SI joint or pubic symphysis dysfunction (a disorder common in healthy pregnant women).

**VASCUITIS**

The data on pregnancy in women with vasculitis are limited to case reports and case series. Because of the dearth of reliable scientific data, many of our assumptions about the treatment of vasculitis in pregnancy are based on the principles we have learned about pregnancy in other systemic autoimmune diseases. Based on this, we assume that having significant vasculitis activity, particularly in the kidneys or lungs, is associated with poor pregnancy outcomes.

**Takayasu’s Arteritis**

Takayasu’s arteritis, a large vessel vasculitis that leads to stenoses and aneurysms in the aorta and its branches, primarily affects young women. More than 150 pregnancies in women with Takayasu’s arteritis have been published, most coming from five case series.66-70 The primary complication seen in these pregnancies was hypertension (≤30% of pregnancies) with almost 20% developing preeclampsia.71 Serious maternal complications including myocardial infarction, aortic aneurysm and rupture, renal insufficiency, and pulmonary embolism have been reported, but the majority of women appear to do well without progression of Takayasu’s arteritis during pregnancy. More than 80% of pregnancies resulted in the delivery of a healthy baby, though preterm delivery, either spontaneous or induced for preeclampsia, was common. Intrauterine fetal demise was reported in 8% of pregnancies, and intrauterine growth restriction (a marker of poor placental perfusion) occurred in almost 20% of pregnancies.71

Treatment of Takayasu’s arteritis during pregnancy depends on the current symptoms of the disease. If active inflammation is present, leading to arterial damage, treatment with corticosteroids or immunosuppressants with a low teratogenic risk may be reasonable. Among asymptomatic women, these medications may not be necessary during pregnancy and should not be started prophylactically. Close monitoring of blood pressure is essential during pregnancy, and treatment with acceptable antihypertensives will likely promote good placental health and minimize maternal organ damage. Blood pressure can increase significantly during labor, exacerbated by maternal exertion. It is important to keep in mind that maternal blood pressure measurements may be inaccurate with an arm cuff if subclavian stenosis is present. In these cases, internal blood pressure monitoring and/or avoidance of labor via a scheduled cesarean section may be prudent.

**Small Vessel Vasculitis**

Granulomatosis with polyangiitis (formerly Wegener’s granulomatosis), microscopic polyangiitis, and Churg-Strauss vasculitis are all rare small vessel vasculitides with a peak onset after the reproductive years. In addition, the primary treatment for severe disease includes cyclophosphamide, which leaves many women infertile. For these reasons, there are few systematically collected data on pregnancies in these diseases. A recent review of 20 pregnancies in 12 women with systemic necrotizing vasculitis (including WG, CSS, PAN, and MPA) from several centers in France included 2 elective abortions, 4 miscarriages, and 14 live births, half of which were delivered preterm. Complications included isolated episodes of pancreatic artery microaneurysms, transient cardiac failure, renal insufficiency from thrombotic microangiopathy, and severe pneumonia. Vasculitis remained quiescent in most pregnancies.72 Case reports of small vessel vasculitis in pregnancy are similar with relatively good pregnancy outcomes. Several report the onset of vasculitis during pregnancy.

The principles of treatment of these vasculitides are similar to other systemic rheumatologic disease: Increased disease activity is likely related to poor pregnancy outcomes, so maintenance of quiet disease is important. Continuation of low-dose prednisone and azathioprine seems prudent in women with recently active disease, but these medications do not need to be initiated prophylactically in women who have been in remission.

**Behçet’s Disease**

Behçet’s disease, which primarily manifests with oral and genital ulcers but can also involve ocular or CNS disease and arterial or venous thrombosis, is a rare disease with
limited pregnancy data. Case series are contradictory about the course of disease activity, with some reporting improvement and others worsening during pregnancy. Pregnancy outcomes are largely good, but some women have increased numbers of miscarriages and episodes of hypertension.\textsuperscript{13} Pregnancy is a period of hypercoagulability, so prophylactic treatment with low-molecular-weight (LMW) heparin for women with a history of thrombosis is important.

**ANTIPHOSPHOLIPID ANTIBODIES AND THE ANTI PHOSPHOLIPID ANTIBODY SYNDROME**

The antiphospholipid antibody syndrome (APS) is discussed in detail in Chapter 82. However, the APS, albeit primary or secondary in women with other autoimmune diseases, has important implications for women who are considering pregnancy. In some cases, the clinical manifestations of the syndrome are limited to the obstetric realm. The revised 2006 criteria for the APS include (1) laboratory criteria: lupus anticoagulant, anticardiolipin antibodies, or anti-\( \beta_2 \)glycoprotein I positive on two or more occasions at least 12 weeks apart; (2) vascular thromboses: at least one unequivocal episode of arterial, venous, or small vessel thrombosis; and (3) obstetric criteria: one or more unexplained deaths of a morphologically healthy fetus after 10 weeks’ gestational age; one or more births before 34 weeks’ gestation because of preeclampsia or placental insufficiency; or three or more consecutive spontaneous abortions (before 10 weeks’ gestation) without chromosomal, anatomic, hormonal, or other causes to explain pregnancy loss.\textsuperscript{19} The prevalence of lupus anticoagulant or antiphospholipid antibodies in the general population is unknown, but it is estimated that up to 8.5% of the general obstetric population has positive antibodies.\textsuperscript{25} However, it is important to recognize that the mere presence of antibodies/lupus anticoagulant does not in itself confer an increased risk of thrombosis/obstetric complications, and actual clinical events are required to make a diagnosis of the actual syndrome. For example, only about 40% of SLE patients with positive titers for antiphospholipid antibodies will develop thrombotic events, even though SLE patients have a high prevalence of antibodies.\textsuperscript{26} When thinking about pregnancy outcomes among women with antiphospholipid antibodies or the antiphospholipid antibody syndrome, it is helpful to divide women into three general categories: (1) women with APS and a history of arterial/venous thrombotic events; (2) women with APS with clinical manifestations limited to obstetric complications (no maternal thrombotic events); and (3) asymptomatic women found to have positive tests for lupus anticoagulant and/or antiphospholipid antibodies (no history of adverse pregnancy outcomes or maternal thromboses).\textsuperscript{27}

**Antiphospholipid Antibody Syndrome with Thrombotic Manifestations**

As is evident in the nonpregnant population of APS patients with a history of arterial or venous thromboses, there is a high rate of recurrent thromboses and lifelong anticoagulant therapy is the mainstay of treatment.\textsuperscript{16} In nonpregnant patients, this is usually achieved with the use of warfarin to a target international normalized ratio (INR) of 2.0 to 3.0. In some cases, recurrent thromboses recur despite therapeutic levels of warfarin and these patients may need to be managed with a higher target INR of 3.0 to 4.0, long-term therapy with heparin, or combination antithrombotic agents.\textsuperscript{16}

When women with APS and thrombotic manifestations consider pregnancy, aggressive antithrombotic therapy with LMW heparin must be continued to reduce risks of recurrent maternal thromboses during the increased hypercoagulable state of pregnancy and the postpartum period. In addition, studies have suggested that women with thrombotic manifestations have higher risks of adverse pregnancy outcomes than do women with only obstetric manifestations of APS despite using the more aggressive antithrombotic management protocol with low-dose aspirin and LMW heparin (daily heparin dosing for obstetric APS compared with twice-daily dosing for thrombotic APS patients).\textsuperscript{77} Overall, with appropriate antithrombotic prophylaxis and management, live-birth rates are greater than 70%.\textsuperscript{76} At this time, there is no role for prednisone or other immunosuppressive treatment in the prevention of thrombotic or obstetric complications due to APS.

**Antiphospholipid Syndrome with Obstetric Manifestations**

Optimal management of pregnancy in women who meet obstetric criteria for APS has been the subject of several randomized clinical trials and meta-analyses; unfortunately, some controversy remains regarding the optimal treatment regimen.\textsuperscript{76} Two randomized controlled studies found a significant reduction in pregnancy loss in women treated with unfractionated heparin plus aspirin when compared with aspirin alone.\textsuperscript{78,79} Other trials, comparing LMW heparin plus aspirin to aspirin alone, did not find a difference in live-birth rates.\textsuperscript{80,81} Studies that have evaluated the difference between LMW or unfractionated heparin (plus aspirin) have failed to find a difference in outcomes.\textsuperscript{82,83} A recent meta-analysis of heparin plus aspirin compared with aspirin alone has concluded that heparin (LMW or unfractionated) plus aspirin has a 30% increase in live-birth rate compared with aspirin alone.\textsuperscript{84} Experts generally agree that low-dose aspirin should be instituted in the preconceptional period, with the addition of once-daily subcutaneous heparin (LMW or unfractionated) on confirmation of pregnancy.\textsuperscript{76,77} Newer experimental evidence strongly suggests a critical role of tissue factor and activated complement in APS-mediated pregnancy loss.\textsuperscript{85} Inhibitors of these factors have yet to be studied in the clinical setting.

**Antiphospholipid Antibodies or Lupus Anticoagulant without a History of Clinical or Obstetric Events**

Although the presence of lupus anticoagulant or antiphospholipid antibodies is associated with increased risks of adverse pregnancy outcomes, there is currently no way to predict which asymptomatic seropositive women will go on to develop related obstetric complications in the absence of a prior history of a clinical event. Thus randomized
controlled trials or risk-benefit analyses of primary prophylaxis during pregnancy are not available. This can be particularly troubling in primigravidas with known positive antibodies because outcomes of previous pregnancies cannot be used to predict the course of a current or future pregnancy. In these cases, the risk of bleeding associated with use of heparin and aspirin is not offset by a theoretic improvement in pregnancy outcomes. Many practitioners elect to use low-dose aspirin alone in these situations, although there are no data to support this practice. In subjects with SLE and asymptomatic antiphospholipid antibodies, the use of HCQ before and during pregnancy may be considered because HCQ may have some antithrombotic properties by several different mechanisms.\(^5\) Again, no clinical data are available to support the routine use of HCQ in pregnant women for the purpose of reducing the risk of APS-related obstetric complications.

**MEDICATIONS IN PREGNANCY AND LACTATION**

The decision to use medications in pregnancy must always balance the risks and benefits to both mother and fetus. For some more systemic rheumatologic diseases, in particular SLE, the risk to the pregnancy of active SLE outweighs the risks of many medications. For other more limited diseases such as RA, the disease process has limited impact on the fetus, making medications more optional.

The U.S. Food and Drug Administration (FDA) rates each medication on the basis of its potential risk to the fetus (Table 39-2). Although this system can provide some general guidance, it is not a perfect system by which to treat a patient. For example, azathioprine and MMF are both FDA class D immunosuppressants, but current data indicate that azathioprine is probably far safer for a fetus than MMF. The FDA has plans to replace the current grading system with more objective data, but when this change will occur is unclear.

Breast-feeding is generally believed to be beneficial to both mother and infant in the best of circumstances. Some rheumatologic medications may transfer into breast milk to a degree that is potentially dangerous, whereas others have more limited transfer. Although current information on lactation and medications is included in this chapter (see Table 39-2), continually updated information is available from the Texas Tech University Infant Risk website at www.infantrisk.org/.

**Nonsteroidal Anti-inflammatory Drugs**

NSAIDs are classified by the FDA as C, with possible risks in the first and second trimesters, but D in the third trimester with well-defined risks.

Some evidence indicates that frequent use of NSAIDs in the week of conception can increase the risk for miscarriage. In a case-control study based on pregnant women in the Kaiser hospital-system database, those taking NSAIDs around conception and in the first trimester had a 1.8-fold increase in miscarriage. That risk increased to a 5.6-fold risk for women taking the NSAID in the week of conception and to an 8.1-fold risk for those taking daily NSAIDs.\(^5\) A similar study from a Danish birth registry found an increased risk for miscarriage, particularly for women who picked up a prescription for NSAIDs in the 6 weeks before the miscarriage (odds ratio [OR], 3.0 to 6.99).\(^6\) Suppression of prostaglandins by NSAIDs may lead to poor implantation or decreased placental perfusion, both of which can precipitate miscarriage. Avoidance of frequent NSAID use in the first trimester and around the time of conception may be prudent.\(^8\)

Many studies have not found an association between NSAIDs and a higher rate of congenital malformation, preterm birth, or low birth weight.\(^7\) Several studies, however, have identified a possible link between NSAID use and congenital heart defects.\(^8\)

In the second and third trimesters, NSAID use may decrease fetal renal perfusion, leading to decreased amniotic fluid production and oligohydramnios. This condition is typically reversible with cessation of the drug.

NSAIDs are contraindicated in the third trimester because they can prompt closure of the fetal ductus arteriosus (DA), the channel through which blood bypasses the fetal lungs. In previous years, NSAIDs were used as tocolytic agents to halt premature labor, allowing several elegant studies demonstrating the impact of NSAID use at different points in the third trimester. With exposure before 27 weeks’ gestation, the DA is largely unaffected. Between 27 and 34 weeks, however, about half of the fetuses will have constriction of the DA, a finding that is evident within hours of NSAID dosing but reverses within several days.\(^9\)

Preterm infants with third-trimester exposure to indomethacin have a higher rate of necrotizing enterocolitis, intraventricular hemorrhage, and a mild decrease in renal function for several days of life.\(^9\)

During lactation, the preferred NSAID is ibuprofen due to its relatively short half-life and low penetration into breast milk. For a woman taking 400 mg of ibuprofen every 6 hours, less than 1 mg of ibuprofen is excreted into the breast milk each day.\(^2\) This is far lower than the recommended maximum pediatric dose of 40 mg/kg per day.

**Corticosteroids**

Corticosteroids vary structurally and will affect the mother and fetus differently. Betamethasone and dexamethasone have fluorine at the 9-alpha position and are not well metabolized by the placenta. They will cross the placenta with direct effects on the fetus. Most other corticosteroids, however, are metabolized in the placenta by 11-beta-hydroxysteroid dehydrogenase to inactivated forms, leaving less than 10% of the active drug to reach the fetus.\(^3\) If the goal is to treat the mother, prednisone is the most ideal glucocorticoid because a small amount of active drug will enter the fetal circulation. If the goal is to treat the fetus, betamethasone and dexamethasone are better options and are commonly used in women at risk for preterm delivery to reduce the risk of respiratory distress and cerebral hemorrhage in the preterm infant. Some recent reports suggest betamethasone may be preferred to dexamethasone because it may offer better long-term neurodevelopmental outcomes for the fetus.\(^4\)

Corticosteroids carry the risk of elevated blood pressure, steroid-induced hyperglycemia, and osteopenia. These risks
should the findings be true, it should be noted that cleft palates occur at a rate of about 1/1000 in the general population and thus the risk with steroid use would increase this to 3/1000. This small increase in risk should be weighed against the need to treat active disease in the mother. There have also been rare isolated reports of neonatal cataracts and adrenal suppression in infants after corticosteroid use, which may be dose dependent.

The pharmacokinetics of prednisolone, the active form of prednisone, has been studied in breast milk. Levels measured in the milk are typically less than 0.1% of the total prednisolone dose ingested by the mother, which when small to moderate doses are being used, is less than 10% of the infant’s endogenous cortisol production. Ost and colleagues evaluated a patient taking up to 80 mg of daily

are pertinent in a pregnant woman because pregnancy requires tight blood pressure control and may induce insulin resistance (and often frank gestational diabetes), as well as osteopenia or osteoporosis of pregnancy. These issues can be exacerbated by steroids, so proper monitoring and calcium and vitamin D supplementation should be employed.

Nonfluorinated corticosteroids do not appear to put a fetus at risk for congenital abnormalities. The exception is a few reports of an increased risk of cleft lip and/or palate, a first-trimester risk. A meta-analysis looking at birth defects after maternal exposure to corticosteroids reported a 3.3 increase in the odds ratio of clefts after first-trimester exposure. Subsequent studies have been mixed, some showing similar results and some showing no difference. However, the findings are true, it should be noted that cleft palates occur at a rate of about 1/1000 in the general population and thus the risk with steroid use would increase this to 3/1000. This small increase in risk should be weighed against the need to treat active disease in the mother. There have also been rare isolated reports of neonatal cataracts and adrenal suppression in infants after corticosteroid use, which may be dose dependent.

The pharmacokinetics of prednisolone, the active form of prednisone, has been studied in breast milk. Levels measured in the milk are typically less than 0.1% of the total prednisolone dose ingested by the mother, which when small to moderate doses are being used, is less than 10% of the infant’s endogenous cortisol production. Ost and colleagues evaluated a patient taking up to 80 mg of daily
prednisolone and calculated that the prednisolone ingested from the milk would add, at most, 10% to the endogenous corticosteroid production of the infant, which may have little clinical significance. Peak steroid levels in breast milk occur about 2 hours after a dose is taken and rapidly decline. These small to moderate amounts of corticosteroids do not appear to have adverse effects on the developing infant; however, exposure may be minimized if nursing is done 3 to 4 hours after the dose is taken.

**Hydroxychloroquine**

HCQ crosses the placenta, and infants may be born with similar serum concentrations of the drug as the mother. The literature includes several reports of isolated women taking daily chloroquine during multiple pregnancies and having infants born with anomalies, most commonly ocular. This finding, however, has not been replicated in studies of HCQ. Researchers systematically evaluated almost 100 infants in several studies for ocular toxicity and did not find abnormalities. A recent study of 21 infants with in utero HCQ exposure found electoretinogram abnormalities in 6, with resolution in one after 7 months. Prior reports of six children and four children with in utero HCQ exposure did not find abnormalities with this testing. Further information about the lasting impact of HCQ on the retina of infants will be helpful. A study of cardiac conduction in exposed infants found no anomalies. Several larger studies have found no increase in fetal anomalies in infants exposed to daily dosing of HCQ for rheumatologic disease.

The discontinuation of HCQ early in pregnancy is associated with SLE flare and decreased gestational age. An increase or decrease in preterm birth and pregnancy loss has not been statistically associated with HCQ use. A case-control study of cardiac neonatal lupus has found a protective effect from HCQ. Although univariate analysis in this study found a 70% decrease in cardiac neonatal lupus with HCQ use, multivariate analysis did not maintain statistical significance for this finding. Prospective studies of HCQ for prevention of neonatal lupus would provide valuable clinical information for women with Ro/SSA antibodies.

HCQ does cross into breast milk but in limited quantities. Analysis of two lactating women taking HCQ estimated that a fully breast-fed infant would ingest an estimated 0.2 mg/kg/day of HCQ, compared with maternal dosing of 6 mg/kg/day. The American Academy of Pediatrics has designated HCQ as compatible with breastfeeding with only minimal risk to the infant. For this reason, breast-feeding is generally allowed in women taking HCQ.

**Sulfasalazine**

Fertility in women does not appear to be affected by SSZ use; however, it can lead to reduced fertility in men due to oligospermia and reduced sperm motility. Fortunately, this is not an irreversible event; normal spermatogenesis should return approximately 2 months after cessation of the drug. Men who want to conceive a child with a partner should be counseled to discontinue the medication 3 months prior. It is also important to counsel men that, although fertility may be reduced, the use of SSZ is not considered an effective form of contraception. Abnormal or reduced spermatogenesis is due to the SSZ metabolite sulfapyridine.

In contrast to its effects on men, SSZ is among the more preferred DMARDs for pregnant women who require ongoing disease-modifying therapy. Much of the data regarding pregnancy outcomes with SSZ exposure is derived from the inflammatory bowel disease (IBD) literature. A recent meta-analysis of 2200 pregnant women with IBD (642 received SSZ or related agents) found no statistically significant differences for congenital anomalies (OR, 1.16; 95% confidence interval [CI], 0.76 to 1.77) or other adverse pregnancy outcomes. These data confirm several other reports of population-based studies published over the past several decades. Congenital abnormalities discussed in isolated case reports have not been seen in systematic population-based studies.

Two additional considerations are worth noting. The first is the theoretic risk that SSZ may cause folate deficiency as it inhibits dihydrofolate reductase and the cellular uptake of folate. Pregnant women require at least 800 μg of folic acid replacement, which should be continued, and perhaps increased, during SSZ use. Folic acid deficiency has been linked to neural tube defects, which can occur as early as 5 weeks of pregnancy (just 3 weeks after conception), so this should be part of the preconception planning discussions. The second consideration is in regards to late-term use. A SSZ metabolite called sulfapyridine can cross the placenta, displace bilirubin from albumin, and possibly lead to neonatal jaundice. Although there are no reports of this occurring after in utero exposure to SSZ, one may consider holding SSZ use during lactation in a preterm or jaundiced baby for 1 to 2 months.

Except in the setting of prematurity, hyperbilirubinemia, or other acute stresses, SSZ is considered safe during lactation. Sulfapyridine levels in breast milk were found to be about 30% to 60% of those in the mother’s serum, but should not pose risk for a healthy full-term infant.

**Azathioprine**

Azathioprine (AZA) carries an FDA classification of D, but it is widely viewed as one of the safest immunosuppressants available in pregnancy. While animal studies have identified congenital anomalies after in utero exposure, several large human studies in women taking the drug for SLE, inflammatory bowel disease, or solid-organ transplants have documented no risk of congenital anomalies following exposure. The few reported congenital anomalies in pregnancies exposed to AZA seem to be sporadic without a specific organ or pathogenic pattern. Pregnancies exposed to AZA, however, are at increased risk for preterm birth and growth retardation. Whether this is more related to the drug or the underlying maternal disease remains unclear.

AZA does cross the placenta, and infants are born with low levels of the drug and its metabolites. This leads to a theoretic risk for neonatal infection that has not been borne out in prospective studies.

For some women with systemic rheumatologic disease, the benefits of disease suppression outweigh the potential risks of AZA. Significant SLE activity, for example, is associated with up to a 40% risk of miscarriage and over a 60%
risk for preterm birth. Any potential harm from AZA use in pregnancy is dwarfed by these statistics. Two small studies of AZA use in SLE pregnancies show an association between AZA therapy and pregnancy loss that is strongly confounded by SLE activity. For SLE pregnancies in which SLE activity remains quiescent with AZA therapy, there is no increase in pregnancy loss or preterm birth.

Although AZA does not appear to transfer into breast milk to a significant degree, breast-feeding while taking this medication is discouraged by the World Health Organization, Working Group on Drugs and Human Lactation.

**Methotrexate (FDA Class X)**

Methotrexate is a dihydrofolate reductase inhibitor that impairs purine metabolism, leading to abnormalities in ribonucleic acid and deoxyribonucleic acid synthesis. It is a known teratogen that can cause dysmorphic facial features, skull and limb abnormalities, and growth deficiency. In addition, developmental delay and mental retardation have been reported. The main toxicity appears to occur before 8 weeks' gestation and can occur with high-dose methotrexate for cancer therapy or pregnancy termination, and in low-dose methotrexate it is used to treat rheumatologic disease. Although case reports include descriptions of significant anomalies, prospective and retrospective cohorts of pregnancies exposed to this drug do not demonstrate a high rate of anomalies. A collection of 97 pregnancies in women with rheumatologic disease exposed to methotrexate from 6 reports found a rate of anomaly of 7% of live births. Elective termination was frequent (17%), and spontaneous miscarriage was high (23% of pregnancies not terminated).

All women taking methotrexate should be well informed about the risk for fetal anomalies and pregnancy loss associated with the drug. A reminder that methotrexate is routinely used to terminate ectopic pregnancies (at a dose about threefold the rheumatologic dose) can make this risk clearer. If a pregnancy is exposed to methotrexate, an increased dose of folic acid may be initiated; folinic acid therapy has been demonstrated to decrease the risk for fetal anomalies in some animal models. For women who have discontinued methotrexate at least 3 months before conception, as is recommended, continuation of folic acid through pregnancy is recommended.

Given the relatively low rate of fetal anomalies identified in pregnancy cohorts, routine recommendation for pregnancy termination is not required but should be offered to all women. Women who decide to continue pregnancy must understand the risks for both pregnancy loss and fetal abnormalities. In wanted pregnancies, ultrasound screening for fetal anomalies should be encouraged with consideration of termination if these are identified. Unfortunately, ultrasound may not be able to uncover all fetal anomalies, leaving some risk for anomaly despite normal evaluations.

Methotrexate has not been studied in lactation and is not recommended.

**Leflunomide**

It is known that leflunomide causes fetal anomalies in animals, primarily CNS and skeletal anomalies. When leflunomide was given to rats at only 1% of the human dose, it resulted in fetal anomalies, decreased birth weight, and increased mortality in the offspring after birth. On the basis of available animal data, the FDA has categorized leflunomide as category X; this indicates that clear associations between the drug and fetal abnormalities have been identified and the risks of use in pregnant humans outweigh any potential benefits. In his article reviewing reproductive risks of leflunomide, Brent nicely outlines managing elimination of leflunomide in a woman who has become pregnant on leflunomide or desires pregnancy largely on the basis of manufacturer and FDA recommendations. The level believed to be safe in humans is 0.02 mg/L (i.e., =10 half-lives). If a woman treated with leflunomide wants to become pregnant, the medication should be stopped and cholestyramine administered to decrease the blood levels rapidly to this level. Cholestyramine, 8 g three times daily, reduces the half-life of the active metabolite from what has been found in early pharmacokinetics analysis to be up to 96 days to approximately 1 day. Therefore, a 10- to 11-day regimen of cholestyramine, 8 g three times daily, is recommended. This approach is certainly conservative but was suggested by the FDA and is the manufacturer's recommendation. Plasma levels of the drug should be verified to be less than 0.02 mg/L by two separate tests performed 2 weeks apart. If levels still remain greater than 0.02 mg/L, then the regimen should continue to be administered and women should use effective contraception until levels are less than 0.02 mg/L. Based on the natural half-life of the drug (again, ≤96 days) it may take up to 2 years to reach plasma active metabolite levels of 0.02 mg/mL due to individual variations; therefore levels should be checked (and if necessary cholestyramine administered) in any patient who has been treated with leflunomide in the previous 2 years. This practice is also supported by a panel of 29 experts who participated in a consensus workshop on antirheumatic drugs during pregnancy and lactation.

Despite the dramatic fetal anomalies seen in rodents after leflunomide exposure, the available human data are more reassuring. Data from the OTIS Collaborative Research Group prospectively evaluated 64 women with RA treated with leflunomide (95% of which were treated with cholestyramine early in the pregnancy), 108 women with RA not treated with leflunomide, and 78 healthy pregnant women. The rate of major structural defects in the exposed group was 5.4%, which was insignificant compared with both comparison groups (4.2% and 4.2%, respectively, P = .13). Although a small study group, these results are encouraging and suggest that these defects may be preventable with appropriate and early intervention. It is unknown how leflunomide exposure may affect a young infant, and lactation is considered to be unsafe until more data become available.

**Mycophenolate Mofetil**

MMF, a reversible inhibitor of inosine monophosphate, was FDA approved in 1995 for use in solid organ transplantation. It is used with increasing frequency for the treatment of autoimmune conditions including lupus nephritis. MMF has been shown to be teratogenic in experimental animals, and increasing data through case reports and registry studies support its teratogenicity in humans.**
crosses the placenta. Exposure to MMF during embryogenesis leads to a possible increased rate of spontaneous abortions, and an estimated 22% to 26% rate of congenital malformations. A distinctive MMF embryopathy has been identified as the “EMFO tetrad: Ear (microtia and auditory canal atresia); Mouth (cleft lip and palate); Fingers (brachydactyly of fifth fingers and hypoplastic toenails; and Organs (cardiac, renal, CNS, diaphragmatic, and ocular).” Based on these data, the FDA has recently included a black box warning on the package insert discussing teratogenicity as a concern with use of MMF in women of childbearing potential. Use of reliable contraception is mandatory for women of childbearing potential. Women taking MMF who want to become pregnant should discontinue the drug at a minimum of 6 weeks before conception. In cases in which ongoing immunosuppression is required to maintain the mother’s health, azathioprine is often considered a safer alternative. There are no data on the excretion of MMF into breast milk or the effect if ingested by infants. Therefore lactation is also contraindicated while using MMF.

Cyclophosphamide

Cyclophosphamide (CYC) is an alkylating agent that alters DNA synthesis. It has an FDA pregnancy classification of D, indicating a significant and known risk with fetal exposure. First-trimester exposure to CYC carries a high risk for major congenital anomalies, particularly of the palate, limbs, and eyes, as well as miscarriage. The rate of these anomalies is unknown due to difficulties with prospective studies, but these malformations are not found in all live births after first-trimester exposure.

In the second and third trimesters, the risks for congenital anomalies is far lower. While data is limited, pregnancies exposed because of maternal breast cancer in pregnancy have resulted in healthy, full-term infants. Among SLE pregnancies reported in the literature with mid to late trimester CYC exposure, two resulted in a fetal death and one in a successful preterm birth. In these cases, the mother was very ill preceding the CYC dosing, making it likely that the pregnancy loss would have occurred without the CYC treatment, as well. When a pregnant woman has a life-threatening flare of rheumatologic disease in the latter half of pregnancy, treatment with CYC can be entertained. The significant risk of pregnancy loss in this situation, with or without CYC therapy, should be discussed with the woman prior to dosing.

CYC transfers into breast milk and has been associated with cytopenias in nursing infants. For this reason, women taking CYC should not breast-feed.

Tumor Necrosis Factor

Due to a lack of fetal anomalies or adverse pregnancy outcomes found in animals exposed to TNF inhibitors, these all carry an FDA pregnancy classification of B. This does not mean, however, that they have been satisfactorily confirmed as safe in human pregnancy. Limited prospective data on pregnancies exposed to TNF inhibitors suggest that there is not an increase in fetal anomalies, pregnancy loss, or preterm birth.

One report of the FDA drug toxicity database hypothesizes that in utero TNF inhibitor exposure is associated with Vertebral abnormalities, Anal atresia, Cardiac abnormalities, Tracheoesophageal fistula and/or Esophageal atresia, Renal agenesis and dysplasia, and Limb (VACTERL) association. VACTERL association is diagnosed when an infant is born with at least three major anomalies that are included in the previous definition. This hypothesis was prompted by the diagnosis of this association in an infant exposed to etanercept 50 mg twice a week throughout gestation for maternal PsA. In the FDA database, another 54 anomalies in 40 infants were identified. None of these infants met criteria for VACTERL association, but 56% of these anomalies fell into the categories included in the VACTERL diagnosis. Although this report brings to light potentially important data, it has several drawbacks when applied to clinical practice. Most importantly, the total number of pregnancies exposed to TNF inhibitors is unknown. In order to find an anomaly rate typical in the general population (3%), only 1200 pregnancies would require exposure to produce 40 infants with anomalies. In addition, the most common anomalies reported in the FDA database including cardiac and urogenital anomalies are also the most common seen in the general population. The more uncommon anomalies are only reported in one to two infants after TNF inhibitor exposure, leading to difficulty associating these with a specific cause.

Although immunoglobulins do not cross the placenta in the first half of pregnancy, the rate of immunoglobulin transfer increases linearly through the following weeks, leading to a large transfer in the weeks before term. At birth, the infant has a higher concentration of maternal antibodies than the mother. Based on this finding, it is hypothesized that TNF inhibitors do not cross the placenta before 16 weeks’ gestation but do cross in increasing concentrations as the pregnancy progresses. Infliximab has been documented to cross the placenta in high levels close to term with infants exposed in utero having serum levels similar to the mother at delivery. The infant serum levels slowly decrease over the following weeks to months. Etanercept, on the other hand, may cross at lower levels. A single infant born after etanercept throughout pregnancy had a serum concentration under 4% of the maternal serum concentration (Figure 39-3).

IgG immunoglobulins do not transfer into human breast milk, and they are not absorbed by the human infant gastrointestinal tract. Six women taking infliximab during lactation have been documented to have no drug in their breast milk. Two women taking etanercept have been documented to have minimal drug in their breast milk. Infants fully breast-fed by mothers taking both of these drugs have not had increasing serum TNF inhibitor concentrations.

Rituximab

Rituximab is a chimeric (mouse/human) monoclonal antibody directed against B cell surface antigen CD20 that is indicated for the treatment of B cell lymphoma and moderate to severe RA. In addition, it has been studied and is occasionally used off label for the treatment of other autoimmune diseases and hematologic disorders (SLE, multiple sclerosis, autoimmune cytopenias, thrombotic
thrombocytopenic purpura). Administration results in rapid and sustained depletion of peripherally circulating CD20⁺ B cells. Like other monoclonal antibodies, rituximab has an IgG construct and therefore crosses the placenta. Little IgG is seen in fetal circulation during the first trimester of pregnancy. Levels slowly rise during the second trimester and reach maternal serum concentrations by approximately 26 weeks' gestation. Maximum IgG transfer across the maternal-fetal interface occurs during the last 4 weeks of gestation, and fetal concentration often exceeds maternal concentration at term delivery. Rituximab carries an FDA pregnancy classification of C. Several case reports of rituximab administered during the second or third trimester of pregnancy for acute life-threatening maternal disease have been published. They have not found associations between antenatal rituximab exposure and adverse pregnancy outcomes or congenital malformations. More recently a study describing pregnancy outcomes following preconceptional or antenatal exposure to rituximab in the rituximab global safety database was published. Of 153 pregnancies with known outcomes, 90 resulted in live births. Twenty-two infants were born prematurely; with one neonatal death of unknown cause at 6 weeks. Eleven neonates had hematologic abnormalities, none with corresponding infections. Four neonatal infections were reported (fever, bronchiolitis, CMV hepatitis, and chorioamnionitis); the rate of congenital malformations was not different than the 3% to 4% expected in the general population. Two congenital malformations were identified: clubfoot in one twin and cardiac malformation in a singleton birth. In almost all cases, pregnancy was complicated by active maternal disease and concomitant use of potentially teratogenic medications (including combined chemotherapy with alkylating agents, methotrexate, MMF, and warfarin). The authors concluded that there was no apparent pattern of adverse outcomes (perinatal infections, symptomatic cytopenias, or congenital malformations) associated with rituximab exposure; however, patients should be followed closely if a pregnancy occurs within 6 months following rituximab administration or therapy is indicated to treat severe maternal disease during an established pregnancy. No data are available regarding the concentration of rituximab in breast milk or its effects on breast-feeding infants.

**Tocilizumab**

The IL-6 inhibitor tocilizumab is an IgG antibody, so it likely crosses the placenta in the latter half of pregnancy. A preliminary report of pregnancy toxicity included 33 pregnancies, largely to women enrolled in the drug studies performed for FDA approval of this drug. Of these, 48% were electively terminated and 9% had unknown outcomes. Of the remaining 16 pregnancies, 5 (31%) resulted in a spontaneous abortion, 8 (50%) resulted in a full-term infant, 1 resulted in an infant death due to lung disease, and 1 was a gestational trophoblastic tumor. This suggests a higher rate of miscarriage than would be expected in RA pregnancies, but an assessment of further risks is not possible. No lactation information about this drug is available, so at this time it is not recommended.

**Selected References**

9. ACOG Committee on Practice Bulletins—Obstetrics: ACOG practice bulletin. Diagnosis and management of preeclampsia and...
557

CHAPTER 39  PREGNANCY IN THE RHEUMATIC DISEASES


*Full references for this chapter can be found on www.expertconsult.com.*
References


HISTORY IN A PATIENT WITH MUSCULOSKELETAL DISEASE

Taking an accurate and comprehensive history of a patient’s musculoskeletal symptoms is crucial for making the correct diagnosis. This history must include a precise understanding of what the patient means by the description of symptoms. The physician must obtain a detailed account of symptom onset, location, patterns of progression, and severity, as well as exacerbating and alleviating factors and associated symptoms. The relationship of the symptoms to psychosocial stressors is important and should be determined. The impact of the symptoms on all aspects of the patient’s functioning must be assessed to guide therapy.

The effects of current or previous therapy on the course of the illness are helpful in efforts to understand current symptoms. Response to anti-inflammatory or glucocorticoid medications may suggest an inflammatory origin. Such responses are not specific to inflammatory rheumatic diseases, however, and must be considered in light of the entire history and physical examination. The physician must assess compliance with therapies for musculoskeletal diseases. Noncompliance with the recommended treatment must be differentiated from treatment failure as the explanation for the patient’s lack of improvement.

While the physician is taking the patient’s history, the patient provides verbal and nonverbal clues to the nature of the illness and how the patient has responded to it. Patients with early rheumatoid arthritis may hold their hands in a flexed posture to minimize intra-articular pressure and pain. Some patients may be overly concerned, whereas others may seem inappropriately indifferent to their symptoms. The physician must appreciate the patient’s understanding of the illness and attitudes toward it to begin effective treatment.

Pain

Pain is the most common symptom that brings a patient with musculoskeletal diseases to the physician. Pain is a subjective hurting sensation or experience described in various terms, often of actual or perceived physical damage. Pain is a complex sensation that is difficult to define, qualify, and measure. The patient’s pain may be modified by emotional factors and previous experiences.

The character of the pain usually is best defined early in the interview because this can be helpful in categorizing the patient’s complaints. Aching in a joint area suggests an arthritic disorder, whereas “burning” or “numbness” in an extremity may indicate a neuropathy. Descriptions of pain as “excruciating” or “intolerable” when the patient is otherwise able to function provide a clue that emotional or psychosocial factors are contributing to or amplifying the symptoms.

The physician must elicit the distribution of the patient’s pain and determine whether this fits with anatomic
structures. Patients describe their pain location in terms of body part names, but frequently the terms are used in a nonanatomic manner. Patients frequently complain of “hip” pain when actually referring to pain in the low back, buttock, or thigh. The interviewer must attempt to clarify this complaint by asking the patient to point to the area of pain with one finger. Pain localized in the distribution of a joint or joints likely reflects an articular disorder. Pain may localize to bursae, tendons, ligaments, or nerves, implying disorders of these structures. In contrast to superficial structures, deep structures often give rise to poorly localizing pain. Similarly, pain arising from small, peripheral joints is often more focal than pain arising from proximal, large joints, such as the shoulders and hips. Pain that is widespread, vaguely described, and not respecting anatomic distributions generally suggests a chronic pain syndrome, such as fibromyalgia or psychiatric disease.

The severity of pain should be assessed. A common approach is to ask the patient to describe the level of pain on a numeric scale of intensity from 0 (no pain) to 10 (very severe pain). For monitoring of disease activity of inflammatory arthritis, measuring pain on a visual analogue scale by having the patient mark the severity of pain over the past week on a 100-mm line can be helpful. Similar scales are used in validated instruments such as the McGill Pain Questionnaire.

The physician must determine what exacerbates and alleviates the pain. Joint pain present at rest but worse with movement suggests an inflammatory process, whereas pain occurring primarily with activity and relieved by rest generally indicates a mechanical disorder, such as degenerative arthritis. Timing of pain symptoms during the day and night also provides important information, as discussed in the next section.

Stiffness

Stiffness is a common complaint among patients with arthritis. What is meant by stiffness varies from patient to patient, however. Some patients may use the term stiffness to refer to pain, soreness, weakness, fatigue, or limitation of motion.1 Rheumatologists generally use the term stiffness to describe discomfort and limitation on attempted movement of joints after a period of inactivity. This “gel” phenomenon occurs usually after an hour or more of inactivity. The duration of stiffness related to inactivity varies, with mild stiffness lasting minutes and severe stiffness lasting hours.

Morning stiffness is an early feature of inflammatory arthropathies and is particularly noted in rheumatoid arthritis and polymyalgia rheumatica, in which morning stiffness may last for several hours. The absence of morning stiffness does not exclude inflammatory arthritis, but its absence is uncommon. A useful question to assess morning stiffness is this: “In the morning, how long does it take for your joints to limber up to as good as they are going to get for the day?” Morning stiffness associated with noninflammatory joint diseases, such as degenerative arthritis, generally is of short duration (usually <30 minutes) and is less severe than stiffness of inflammatory joint disease. Additionally, the degree of stiffness in noninflammatory joint diseases is related to the extent of use of the damaged joint—stiffness is worse after excessive use, generally improving within several days to the baseline level. Morning stiffness is not specific for inflammatory arthropathies and may be described by patients with fibromyalgia or chronic idiopathic pain syndromes, neurologic disorders such as Parkinson’s disease (although generally without limbering up), and sleep-related breathing disorders.

Limitation of Motion

Limitation of motion is a common complaint among patients with articular disorders. This complaint must be differentiated from stiffness, which usually is transient and variable, whereas limitation of motion secondary to joint disease is generally fixed and varies less over time. The interviewer should determine the extent of disability resulting from the restriction in joint motion. The duration of the restriction in joint motion frequently predicts the likelihood of improvement with interventions such as oral and intra-articular glucocorticoids or physical therapy. Determining the rapidity of onset of the limitation of motion may be helpful in the differential diagnosis; abrupt onset of the limitation of motion suggests a structural derangement, such as a tendon rupture or torn knee cartilage, whereas insidious onset of restricted joint motion is more common with inflammatory joint disease.

Swelling

Joint swelling is an important symptom in patients with rheumatic diseases. The presence of true joint swelling narrows the differential diagnosis in a patient with arthralgia. To determine whether the swelling is related to joint synovitis as opposed to soft tissue conditions, clarifying the anatomic location and distribution of the swelling is key. Diffuse soft tissue swelling can occur because of venous or lymphatic obstruction, soft tissue injury, or obesity. The description of swelling in patients with such conditions usually is ill defined or is not in a distribution of particular joints, bursae, or tendons. Obese patients may interpret normal adipose tissue over the medial aspect of the elbow, the knee, or the lateral aspect of the ankle as joint swelling. In contrast, patients with inflammatory arthritis may describe swelling of joints in a distribution typical of a specific disease—symmetric swelling of the metacarpophalangeal joints and wrists in rheumatoid arthritis, or swelling of several toes and a knee in psoriatic arthritis.

It is useful to delineate the onset and progression of swelling and the factors that influence it. Swelling of a joint resulting from synovitis or bursitis frequently is associated with discomfort with motion because of tension on the inflamed tissues. If swollen tissues are periarticular, however, no discomfort may be present with joint motion because the inflamed tissues are not stressed. Swelling of a confined structure, such as a synovial cavity or bursa, is most painful when it has developed acutely, whereas a similar degree of swelling that has developed slowly often is much more tolerable.

Weakness

Weakness is another common complaint that can be associated with myriad different subjective meanings. True
weakness is the loss of muscle power. When present, it is demonstrable on physical examination.

The temporal course of weakness is important to the differential diagnosis. Weakness of sudden onset without trauma often indicates a neurologic disorder, such as an acute cerebrovascular event, which generally results in a fixed, nonprogressive deficit. Weakness of insidious onset more often suggests a muscle disease, such as an inflammatory myopathy (i.e., polymyositis). The latter tends to be ongoing and progressive. Weakness that is intermittent suggests a disorder of the neuromuscular junction, such as myasthenia gravis. Patients with this disease may describe muscle fatigue with activity as opposed to true weakness.

The physician should determine the distribution of the patient’s weakness. Proximal weakness that is bilateral and symmetric suggests an inflammatory myopathy. In contrast, inclusion body myositis causes an asymmetric and more distal weakness. The presence of a unilateral or isolated deficit generally indicates a neurogenic origin. Distal weakness, in the absence of joint findings, generally indicates neurologic disorders, such as peripheral neuropathy. Patients with peripheral neuropathies also complain of pain and sensory symptoms, such as paresthesias. In contrast, patients with inflammatory myopathy often present with painless weakness.

Inquiring about the patient’s family history may provide valuable information. A history of other family members with similar symptoms may increase the likelihood that the patient has a hereditary disorder, such as muscular dystrophy or familial neuropathy.

It also is important to review medication taken recently or currently. Many medications, including corticosteroids and lipid-lowering agents, can cause muscle injury. Less commonly, environmental exposure can lead to symptoms of weakness. Heavy metal poisoning causes a peripheral neuropathy. Dietary exposure also should be investigated such as eating undercooked pork as a source of trichinosis. Excessive alcohol intake has been associated with neuropa thy and myopathy.

Taking a complete review of systems is helpful in evaluating a patient with weakness. Constitutional symptoms, such as weight loss and night sweats, may indicate the presence of a malignancy as the cause of generalized weakness. Rash, arthralgia, or Raynaud’s phenomenon may prompt further testing for a connective tissue disease.

Fatigue

Patients with musculoskeletal disorders frequently complain of fatigue. Fatigue can be defined as an inclination to rest even though pain and weakness are not limiting factors. Fatigue after varying degrees of activity that is relieved by rest is normal. Patients with rheumatic diseases experience fatigue even without activity. Fatigue generally improves as the systemic rheumatic disease improves. Malaise frequently occurs with, but is not synonymous with, fatigue. Malaise indicates the lack of well-being that often occurs at the onset of an illness. Fatigue and malaise may occur in the absence of identifiable disease, and psychosocial factors, anxiety, or depression may account for the symptoms.

Loss of Function

The comprehensive history should include an assessment of the patient’s ability to perform activities of daily living, as loss of function is a common manifestation of musculoskeletal disease with serious impact on health and quality of life. The extent of disability may vary from loss of the ability to use one finger joint due to arthritis to complete physical incapacitation due to severe inflammatory polyarthritis. Irrespective of the cause, loss of physical function often has a profound impact on patient social activities, exercise routine, work capacity, and even basic self-care. Assessing for the presence and degree of functional disability is important in evaluating the severity of illness and in making treatment recommendations, particularly in rheumatoid arthritis, where disability is among the best predictors of long-term outcomes and mortality.

Functional capacity is assessed first by asking general questions about the patient’s ability to perform daily activities, including grooming, dressing, bathing, eating, walking, climbing stairs, opening doors, carrying objects, and so forth. A report of a specific loss of function, such as difficulty opening a milk carton, should be investigated further to clarify why the task is difficult, which will inform the differential diagnosis and guide clinical examination. This information will also yield important information for management, such as opportunities for physical and occupational therapy, use of splints/braces, and so forth. Overall functional capacity may be evaluated with the use of an instrument such as the Health Assessment Questionnaire (see Chapter 33), which is widely used in research and in the clinic to monitor changes in physical function in response to therapy among patients with rheumatoid arthritis and other rheumatic diseases.

SYSTEMATIC METHOD OF EXAMINATION

The musculoskeletal examination should be a systematic, thorough assessment of the status of the joints, periarticular soft tissues, tendons, ligaments, bursae, and muscles. Rheumatologists commonly begin by examining the upper extremities followed by the trunk and lower extremities, but many routines may be effective, provided that a systematic, consistent approach is used. Gentle handling of tender and painful joints enhances cooperation by the patient and allows an accurate evaluation of the joints.

The general aim of the examination of the joints is to detect abnormalities in structure and function. Key signs of articular disease include swelling, tenderness, limitation of motion, crepitation, deformity, and instability.

General Observation

A general examination of the patient should be done to look for any signs of systemic illness. This should include an examination of the skin, noting signs of pallor (which may suggest anemia), nodules (which may suggest rheumatoid arthritis or gout), or rashes (which may suggest lupus, vasculitis, or dermatomyositis). The patient should be
appropriately undressed for examination. The physician should assess gait by asking the patient to walk in the examining room or down the hallway, because an antalgic gait may be seen in various musculoskeletal disorders of the spine or lower extremities, and various gait disorders may occur in neuromuscular diseases. The ability of the patient to arise and transfer to the examining table should also be evaluated, as this will provide information on pain, proximal muscle strength, and overall physical function. The patient should be assessed for appearance of the muscles, including bulk, tone, and tenderness. Muscle bulk should be compared on one side of the body with the other to look for any asymmetry, hypertrophy, or atrophy. The patient’s manner and body language may provide information on his or her mood and anxiety level, which merits consideration in evaluating pain and tenderness.

Swelling

Swelling around a joint may be caused by intra-articular effusion, synovial proliferation, periarticular subcutaneous tissue inflammation, bursitis, tendinitis, bony enlargement, or extra-articular fat pads. A keen understanding of the anatomic configuration of each joint’s synovial membrane is crucial in differentiating soft tissue swelling secondary to a joint effusion from swelling of periarticular tissues. First, the examiner should inspect the joints for visible evidence of swelling, such as loss of normal landmarks or contours. It is frequently helpful to visually compare the same joints on both sides of the body to detect subtle evidence of swelling and to appreciate symmetry.

Second, the examiner should palpate each joint. The normal synovial membrane is too thin to palpate, whereas the thickened synovial membrane in many chronic inflammatory arthritides, such as rheumatoid arthritis, may have a “doughy” or “boggy” consistency. In some joints, such as the knee, the extent of the synovial cavity can be delineated on physical examination by compressing the fluid into one of the extreme synovial recesses. The edge of the resulting bulge may be palpated more easily. If this palpable edge is within the anatomic confines of the synovial membrane and disappears on release of compression, the distention usually represents synovial effusion; if it persists, it is an indication of a thickened synovial membrane. Reliable differentiation between synovial membrane thickening and effusion is not always possible by physical examination, however. Ultrasonography is being used increasingly as an extension of the physical examination, allowing the examiner to differentiate between synovial proliferation and effusion.

Tenderness

In the musculoskeletal examination, tenderness indicates unusual discomfort on palpating and putting pressure on articular and periarticular tissues. Localizing the tenderness to palpation may assist the examiner in determining whether the pathology is intra-articular or periarticular in location, such as a fat pad, tendon attachment, ligament, bursa, or muscle, or skin. It can be useful to palpate structures that are not involved to assess the importance of tenderness. Finding tender joints in a patient who also has numerous other myofascial tender points is less of a concern for arthritis than finding tender joints in a patient with no extra-articular tenderness.

Limitation of Motion

Limitation of motion is a common manifestation of articular disease; the examiner must know the normal type and range of motion for each joint. Comparison of the affected joint with an unaffected joint of the opposite extremity is useful in evaluating individual variation. Restricted joint motion may be caused by changes in the joint itself or in periarticular structures. To distinguish these possibilities, it is crucial to compare the passive with the active range of motion. If the passive range of motion is greater than the active range of motion, the restriction may be the result of pain, weakness, or the state of articular or periarticular structures. It also is important to distinguish muscle tension from a true limitation of joint motion, emphasizing the importance of ensuring relaxation of the patient. Pain that occurs with attempts to move a joint passively to the limit of range of motion in one plane is referred to as stress pain. Pain in the joint with attempted active or passive range of motion usually indicates an abnormality in the joint.

Crepitation

Crepitation is a palpable or audible grating or crunching sensation produced by motion. This sensation may or may not be accompanied by discomfort. Crepitation occurs when roughened articular or extra-articular surfaces are rubbed together by active motion or by manual compression. Fine crepitation often is palpable over joints involved by chronic inflammatory arthritis and usually indicates roughening of the opposing cartilage surfaces as a result of erosion or the presence of granulation tissue. Coarse crepitation may be caused by inflammatory or noninflammatory arthritis. Bone-on-bone crepitus produces a higher-frequency, palpable, audible squeak. Crepitation from within a joint should be differentiated from cracking or popping sounds caused by the slipping of ligaments or tendons over bony surfaces during motion. The latter phenomena are usually less contributory to the diagnosis of joint disease and may be heard over normal joints. In scleroderma, a distinct, coarse, creaking, leathery crepitation may be palpable or audible over tendon sheaths.

Deformity

Deformity of the joints may manifest as a bony enlargement, articular subluxation, contracture, or ankylosis in nonanatomic positions. Deformed joints usually do not function normally, frequently restrict activities, and may be associated with pain, especially with overuse. Occasionally, a deformed joint may function well but is a cosmetic concern. Joint deformities may be reversible or irreversible. Multiple swan neck deformities of the fingers that can be corrected with manipulation may indicate Jaccoud’s arthropathy of lupus. In contrast, hand deformities in rheumatoid arthritis generally are not correctable.
Instability
Joint instability is present when the joint has greater than normal movement in any plane. Subluxation refers to a joint in which there is partial displacement of the articular surfaces but still some joint surface-to-surface contact. A dislocated joint has lost all cartilage surface-to-surface contact. Instability is best determined by supporting the joint between the examiner’s hands and stressing the adjacent bones in directions in which the normal joint does not move. The patient must be relaxed during the examination because muscle tension may stabilize an otherwise unstable joint. A knee with a deficient ligament might appear stable if the patient contracts the quadriceps muscles during evaluation.

Other Aspects of the Examination
Examinations of the cervical spine and low back are discussed in Chapters 45 and 47.

RECORDING THE JOINT EXAMINATION
Documentation of the joint examination is important in making decisions about therapy, monitoring the activity of arthritis, and determining the efficacy of interventions. Many different recording methods have been described. Abbreviations for each joint can be used, such as PIP for the proximal interphalangeal joints. The S-T-L system has been used historically to record the degree of swelling (S), tenderness (T), and limitation of motion (L) of each joint on the basis of a quantitative estimate of gradation. This method remains useful but is used less commonly today because of increasing reliance on electronic medical records. It is easier to describe joint findings in narrative form, for example, “there is 2+ swelling of the second and third MCP joints,” where grade 0 indicates no swelling, grade 1 indicates palpable synovial thickening, grade 2 indicates loss of normal joint contours, and grade 3 indicates frank cystic swelling of the metacarpophalangeal joint. An alternative method is to record joint examination findings using a schematic skeleton or homunculus. When accuracy is necessary, the range of motion of individual joints may be measured using a goniometer.

Joint counts are being used increasingly to monitor the activity of inflammatory arthritis in practice and in clinical trials. For monitoring disease activity of rheumatoid arthritis, a 28-joint count for tenderness and swelling has been recommended. To assess the tender joint count, the examiner documents which joints the patient indicates are painful on palpation with enough pressure to blanch the nail bed of the examiner’s thumb and index fingers. To assess the swollen joint count, the examiner documents which joints have palpable soft tissue swelling or fluctuance, excluding joints affected only by deformity or bony hypertrophy. The 28-joint count includes the shoulders, elbows, wrists, first to fifth metacarpophalangeal joints, first to fifth proximal interphalangeal joints, and knees on both sides of the body. Compared with more extensive joint counts, the 28-joint count has the advantage of being quick and easy to perform; however, it is limited by the fact that the ankles and metatarsophalangeal joints are not included, so active disease in the feet may be underestimated. The 28-joint count is used to calculate the Disease Activity Score 28 (DAS28), which is a validated instrument used to monitor disease activity.

The function of the joints in normal use is not captured by assessments of tenderness, swelling, or range of motion, so other examination techniques are necessary. Other tests are available that attempt to measure joint function by assessing the patient’s ability to perform a coordinated task (e.g., shoulder arc of motion, measuring the 50-foot walk time). The results of such functional tests may vary, however. Biologic factors, such as circadian changes in joint size and grip strength among rheumatoid patients observed during a 24-hour interval, contribute to variability.

INTERPRETING THE JOINT EXAMINATION
The physician must understand the significance of specific joint findings, both their presence and absence, to make appropriate treatment decisions. As with any diagnostic assessment, the accuracy and reliability of the joint examination are important considerations. With regard to accuracy in detecting physical signs of inflammatory synovitis, numerous studies have shown that joint examination is far less sensitive in detecting synovitis or effusions than high-resolution ultrasonography or magnetic resonance imaging. Although it prevails that swollen joints are more specific for active synovitis, recent clinical studies have suggested that joint tenderness has similar value in predicting the progression of radiographic joint damage as compared with swelling. Demonstrable physical signs of arthritis may be particularly subtle for patients with early disease. Thus, the examiner must consider the physical findings in view of the complete history of joint symptoms to make an accurate diagnosis, assess prognosis, and prescribe management.

The joint examination is also affected by variability. For observations such as joint tenderness or grip strength, interobserver variability usually is greater than intraobserver variability. Considerable intraobserver variability may be noted in observations of the same patient, even over a short interval. Interobserver reliability, in general, is higher for joint-line tenderness than for swelling and is specifically related to the underlying disease, such as higher reliability of the examination for joint swelling in rheumatoid arthritis than in psoriatic arthritis.

EXAMINATION OF SPECIFIC JOINTS
Temporomandibular Joint
The temporomandibular joint is formed by the condyle of the mandible and the fossa of the temporal bone just anterior to the external auditory canal. It is difficult to visualize swelling of this joint. The examiner may palpate the joint by placing a finger just anterior to the external auditory canal and asking the patient to open and close the mouth and to move the mandible from side to side. The presence of synovial thickness or swelling of minimal or moderate
The sternoclavicular joint may be the site of septic arthritis, although this involvement is often subclinical. Ankylosing spondylitis, rheumatoid arthritis, and degenerative arthritis are commonly associated with sternoclavicular joint involvement. Synovitis usually is visible and palpable in these joints. The sternoclavicular joints are the only articulations in this group that are always diarthrodial; the others are amphiarthroses or synchondroses. The sternoclavicular joints have the only true points of articulation of the shoulder girdle with the trunk. These joints are just beneath the skin; synovitis is most easily palpable without evidence of severe arthritis.

Many arthritides can affect the temporomandibular joints, including juvenile and adult rheumatoid arthritis. Tenderness of the manubrium and the body of the sternum is at the level of attachment of the second costal cartilage to the sternum. The third through seventh sternocostal joints are immediately caudal. The articulations of the first ribs and the sternum at its upper end to form the sternoclavicular joint, consistent with the temporomandibular joint syndrome (see Chapter 51). This syndrome is thought by some investigators to result from bruxism and is likely to be a form of myofascial pain, similar to fibromyalgia.

Cricoarytenoid Joints

The paired cricoarytenoid joints are formed by the articulation of the base of the small pyramidal arytenoid cartilage and the upper posterosalateral border of the cricoid cartilage. This joint is a hinge joint. The radiohumeral and proximal radioulnar articulations allow rotation of the forearm.

To examine the elbow joint, the examiner places the thumb between the lateral epicondyle and the olecranon process in the lateral paraolecranon groove and places one or two fingers in the corresponding groove medial to the olecranon. The examiner relaxes and passively moves the elbow through flexion, extension, and rotation. One should examine the skin around the elbow joint carefully, noting abnormalities such as psoriatic plaques, rheumatoid nodules, or tophi. It is useful to palpate the olecranon bursa carefully to exclude the presence of small nodules or tophi. Synovial swelling is most easily palpated as it bulges under the examiner's thumb when the elbow is passively extended. The synovial membrane sometimes can be palpated over the posterior aspect of the joint between the olecranon process and the distal humerus. Synovitis or effusion generally results in limitation of elbow extension.

The olecranon bursa overlies the olecranon process of the ulna. Olecranon bursitis is common after chronic local trauma and in rheumatic diseases, including rheumatoid arthritis and gout. A septic olecranon bursitis may occur. A patient who has olecranon bursitis usually presents with a swelling over the olecranon process, which is often tender and may be erythematous. Sometimes a large collection of fluid over the area is palpable as a cystic mass, often requiring aspiration and drainage. There is generally no pain with elbow movement.
Figure 40-2  Wrist or radiocarpal articulation is a biaxial ellipsoid joint between the radius, the ulna, and the carpal bones. The true wrist is a complex joint formed by many articulations.

Wrist and Carpal Joints

The wrist is a complex joint formed by many articulations between the radius, the ulna, and the carpal bones. The true wrist or radiocarpal articulation is a biaxial ellipsoid joint formed proximally by the distal end of the radius and the triangular fibrocartilage and distally by a row of three carpal bones: scaphoid (navicular), lunate, and triquetrum (triangular). The distal radioulnar joint is a uniaxial pivot joint. The midcapon joints are formed by the junction of the proximal and distal rows of the carpal bones. The midcapon and carpometacarpal articular cavities often communicate. The intercarpal joints refer to the articulations between individual carpal bones.

Movements of the wrist include flexion (palmar flexion), extension (dorsiflexion), radial deviation, ulnar deviation, and circumduction. Pronation and supination of the hand and forearm occur primarily at the proximal and distal radioulnar joints. The only carpometacarpal joint that moves to a notable degree is the carpometacarpal joint of the thumb. This joint is saddle-shaped and moves in three planes. Crepitus at this joint is common because it is frequently involved in degenerative arthritis.

The wrist normally can be extended to 70 to 80 degrees and flexed to 80 to 90 degrees. Ulnar and radial deviation should allow 50 degrees (ulnar) and 20 to 30 degrees (radial) of movement. Loss of extension is the most incapacitating functional impairment of wrist motion.

The long flexor tendons of the forearm musculature cross the volar aspect of the wrist and are enclosed in the flexor retinaculum (transverse carpal ligament). The flexor retinaculum and the underlying carpal bones form the carpal tunnel. The median nerve passes through the carpal tunnel superficial to the flexor tendons. The extensor tendons of the forearm musculature are enclosed by six synovial lined compartments.

The palmar aponeurosis (fascia) spreads out into the palm from the flexor retinaculum. Dupuytren's contracture is a fibrosing condition affecting palmar aponeurosis that becomes thickened and contracted and may draw one or more fingers into flexion at the metacarpophalangeal joint. The fourth finger frequently is affected first.

Swelling of the wrist may be caused by effusion or synovial proliferation, or both, of the tendon sheaths (tenosynovitis), the wrist joint, or a combination thereof. When swelling is attributable to tenosynovitis, swelling is localized to the distribution of a particular tendon sheath or compartment (i.e., ulnar swelling attributable to tenosynovitis of the flexor carpi ulnaris tendon), tends to be more localized, and moves with flexion and extension of the fingers. Articular swelling tends to be more diffuse and protrudes anteriorly and posteriorly from under the tendons.

Synovitis of the wrist is best detected by palpation of the dorsal aspect of the joint. Accurate localization of the synovial margins is difficult because of structures overlying the volar and dorsal aspects of the wrist. To examine the wrist, the examiner should palpate the joint gently between the thumbs dorsally and the fingers on the volar aspect. Thickening or frank synovial proliferation of the synovium should be noted. When this thickening or proliferation is severe, the range of motion of the wrist joint frequently is limited and associated with stress pain.

A ganglion is a cystic enlargement arising from a joint capsule. Ganglions characteristically occur at the volar or dorsal aspect of the wrist between the tendons. Subluxation of the ulna may develop as a result of severe chronic inflammatory arthritis (Figure 40-2). The

---

Figure 40-1  Diagram of the elbow. Posterior aspect of the elbow joint showing radius and ulna in extension and the distribution of the synovial membrane in distention. (From Polley HF, Hunder GG: Rheumatologic interviewing and physical examination of the joints, ed 2, Philadelphia, 1978, WB Saunders. Used with permission of Mayo Foundation for Medical Education and Research.)
subluxated ulna appears as a prominence on the dorsomedial wrist. Chronic irritation of the extensor tendons, primarily the fourth and fifth finger extensor tendons, may cause these tendons to rupture.

“Trigger fingers” secondary to stenosing tenosynovitis can be detected by palpating crepitus or nodules along the tendons in the palm while the patient slowly flexes and extends the fingers. The patient usually gives a history of the affected finger catching or locking with movement.

Tenosynovitis of the first extensor compartment, which encloses the abductor pollicis longus and extensor pollicis brevis muscles of the thumb, is known as de Quervain’s tenosynovitis. Patients complain of pain at the radial aspect of the wrist. Tenderness may be elicited by palpating near the radial styloid process. The Finkelstein test for de Quervain’s tenosynovitis is performed by asking the patient to make a fist with the thumb enclosed in the palm of the hand, then to move the wrist into ulnar deviation. Severe pain over the radial styloid is a positive finding, often indicating stretching of the thumb tendons in a stenosed tendon sheath.

Carpal tunnel syndrome results from pressure on the median nerve in the carpal tunnel. Carpal tunnel syndrome is discussed in detail in Chapter 50.

Muscle function of the wrist may be measured by testing flexion and extension and supination and pronation of the forearm. The principal flexors of the wrist are the flexor carpi radialis (nerve roots C6 and C7) and flexor carpi ulnaris (C8 and T1) muscles. Each of these muscles can be tested separately. This testing can be accomplished if the examiner provides resistance to flexion at the base of the second metacarpal bone in the direction of extension and ulnar deviation in the case of the flexor carpi radialis muscle and resistance at the base of the fifth metacarpal in the direction of extension and radial deviation in the case of the flexor carpi ulnaris muscle. The principal extensors of the wrist are the extensor carpi radialis longus (C6 and C7), extensor carpi radialis brevis (C6 and C7), and extensor carpi ulnaris (C7 and C8) muscles. The radial and ulnar extensor muscles can be tested separately. The principal supinators of the forearm are the biceps brachii (C5 and C6) and supinator (C6) muscles. The principal pronators of the forearm are the pronator teres (C6 and C7) and pronator quadratus (C8 and T1) muscles.

Metacarpophalangeal and Proximal and Distal Interphalangeal Joints

The metacarpophalangeal joints are hinge joints. Lateral collateral ligaments that are loose in extension tighten in flexion, preventing lateral movement of the digits. The extensor tendons that cross the dorsum of each joint strengthen the articular capsule. When the extensor tendon of the digit reaches the distal end of the metacarpal head, it is joined by fibers of the interossei and lumbricales muscles and expands over the entire dorsum of the metacarpophalangeal joint and onto the dorsum of the adjacent phalanx. This expansion of the extensor mechanism is known as the extensor hood.

The proximal and distal interphalangeal joints also are hinge joints. The ligaments of the interphalangeal joints resemble those of the metacarpophalangeal joints. When the fingers are flexed, the bases of the proximal phalanges slide toward the palmar side of the heads of the metacarpal bones. The metacarpal heads form the rounded prominences of the knuckles, with the metacarpal joint spaces situated about 1 cm distal to the apex of the prominences.

To examine the metacarpophalangeal joints, the examiner should palpate the dorsal and volar aspects of each joint in 20 to 30 degrees of flexion (Figure 40-3). The skin on the palmar surface of the hand is thick and covers a fat pad between it and the metacarpophalangeal joint. This makes palpation of the palmar surface of the joint difficult. It is especially helpful in examining the small joints to compare one with another to detect subtle synovitis. Gentle lateral compression with force applied at the base of the second and fifth metacarpophalangeal joints often elicits tenderness if synovitis is present (squeeze test).

The proximal and distal interphalangeal joints are best examined by palpating gently over the lateral and medial aspects of the joint, where the flexor and extensor tendons do not interfere with assessment of the synovial membrane. Alternatively, the joint can be compressed anteroposteriorly by the thumb and index finger of one of the examiner’s hands.
hands, while the other thumb and index finger palpate for synovial distention medially and laterally. The Bunnell test is useful in differentiating synovitis of the proximal interphalangeal joints from tightening of the intrinsic muscles (see Chapter 50).

Swelling of the fingers may result from articular or periarticular causes. Synovial swelling usually produces symmetric enlargement of the joint itself, whereas extra-articular swelling may be diffuse and may extend beyond the joint space. Asymmetric enlargement, involving only one side of the digit or joint, is less common and usually indicates an extra-articular process. Diffuse swelling of an entire digit, known by the terms dactylyitis and sausage digit, may result from tenosynovitis and is seen most commonly in the spondyloarthopathies, such as reactive arthritis or psoriatic arthritis. Rheumatoid nodules are firm periarticular swellings that frequently overlie the joints or bony prominences in patients with chronic rheumatoid disease. Chronic swelling with distention of the metacarpophalangeal joints tends to produce stretching and laxity of the articular capsule and ligaments. This laxity, combined with muscle imbalance and other forces, eventually results in the extensor tendons of the digits slipping off the metacarpal heads to the ulnar sides of the joints. The abnormal pull of the displaced tendons is one of the factors that cause ulnar deviation of the fingers in chronic inflammatory arthritis (Figure 40-4).

Swan neck deformity describes a finger with a flexion contracture of the metacarpophalangeal joint, hyperextension of the proximal interphalangeal joint, and flexion of the distal interphalangeal joint. These changes are produced by contraction of the interossei and other muscles that flex the metacarpophalangeal joints and extend the proximal interphalangeal joints. This deformity is characteristic of rheumatoid arthritis but may be seen in other chronic arthritides (Figure 40-5).

Boutonnière deformity describes a finger with a flexion contracture of the proximal interphalangeal joint associated with hyperextension of the distal interphalangeal joint. The deformity is common in rheumatoid arthritis and results when the central slip of the extensor tendon of the proximal interphalangeal joint becomes detached from the base of the middle phalanx, allowing palmar dislocation of the lateral bands. The dislocated bands cross the fulcrum of the joint and act as flexors instead of extensors of the joint.

Another abnormality is telescoping or shortening of the digits produced by resorption of the ends of the phalanges secondary to destructive arthropathy. This may be seen in the arthritis mutilans form of psoriatic arthritis. Shortening of the fingers is associated with wrinkling of the skin over involved joints and is called opera-glass hand or la main en lorgnette.

A mallet finger results from avulsion or rupture of the extensor tendon at the level of the distal interphalangeal joint. With this deformity, the patient is unable to extend the distal phalanx, which remains in a flexed position. This deformity frequently results from traumatic injuries.

The Murphy sign is a test for lunate dislocation. The patient is asked to make a fist. The third metacarpal head usually is more prominent than the second and fourth. If the third metacarpal is level with the second and fourth, the finding is positive for lunate dislocation.

Involvement of the distal interphalangeal joints in rheumatoid arthritis is uncommon. Bony hypertrophy and osteophyte formation are commonly seen, however, at the distal and proximal interphalangeal joints in patients with osteoarthritis. Enlarged, bony, hypertrophic distal interphalangeal joints are called Heberden nodes, whereas similar changes at the proximal interphalangeal joints are called Bouchard nodes. These usually are easily differentiated from the synovitis of inflammatory arthritis because, on palpation, the enlargement is hard or bony. In addition, signs of inflammation are minimal. Heberden and Bouchard nodes should be easily distinguished from rheumatoid nodules, but patients occasionally confuse these when describing swellings over joints. The examiner should be aware of other causes of nodules on the hands, including tophaceous gout (Figure 40-6) and, rarely, multicentric reticuloendotheliosis. The first carpometacarpal joint also is often affected in osteoarthritis (Figure 40-7).

The patient’s fingernails should be inspected for evidence of clubbing or other abnormalities. Often in patients with
EVALUATION OF GENERALIZED AND LOCALIZED SYMPTOMS

or oppose fingers can be demonstrated by asking the patient to pick up a small object.

Strength of the hands can be assessed crudely by asking the patient to grip firmly two or more of the examiner’s fingers. More accurate measures of grip strength can be made by using a dynamometer or by having the patient squeeze a partially inflated sphygmomanometer (at 20 mm Hg). It sometimes is useful to test the strength of the fingers separately. The prime movers of flexion of the second through fifth metacarpophalangeal joints are the dorsal and palmar interossei muscles (nerve roots C8 and T1). The lumbricales muscles (C6, C7, and C8) flex the metacarpophalangeal joints when the proximal phalangeal joints are extended. The flexors of the proximal interphalangeal joints are the flexor digitorum superficialis muscles (C7, C8, and T1), and the flexor of the distal interphalangeal joints is the flexor digitorum profundus muscle (C7, C8, and T1).

The prime extensors of the metacarpophalangeal joints and interphalangeal joints of the second through fifth fingers are the extensor digitorum communis (nerve roots C6, C7, and C8), extensor indicis proprius (C6, C7, and C8), and extensor digiti minimi (C7) muscles. The interossei and lumbricales muscles simultaneously flex the metacarpophalangeal joints and extend the interphalangeal joints. The dorsal interossei (C8 and T1) and abductor digiti minimi (C7) muscles abduct the fingers, whereas the palmar interosseous muscles adduct the fingers.

The thumb is moved by several muscles. The prime flexor of the first metacarpophalangeal joint is the flexor pollicis brevis muscle (nerve roots C6, C7, C8, and T1). The prime flexor of the interphalangeal joint is the flexor pollicis longus muscle (C8 and T1). The metacarpophalangeal joint of the thumb is extended by the extensor pollicis brevis muscle, and the prime extensor of the interphalangeal joint is the extensor pollicis longus muscle (C6, C7, C8, and C9).

The principal abductors of the thumb are the abductor pollicis longus (nerve roots C6 and C7) and the abductor pollicis brevis (C6 and C7) muscles. Motion occurs primarily at the carpometacarpal joint. The principal adductor of the thumb is the adductor pollicis muscle (C8 and T1). Motion occurs primarily at the carpometacarpal joint. The principal movers in opposition of the thumb and fifth fingers are the opponens pollicis (C6 and C7) and opponens digiti minimi (C8 and T1) muscles.

Sensation and nerve injuries in the upper extremity are discussed in Chapters 45 and 50.

Hip

The hip is a spheroidal or ball-and-socket joint formed by the rounded head of the femur and the cup-shaped acetabulum (see Chapter 48). Stability of the joint is ensured by the fibrocartilaginous rim of the glenoid labrum and the dense articular capsule and surrounding ligaments, including the iliofemoral, pubofemoral, and ischiocapsular ligaments that reinforce the capsule. Support also is provided by the powerful muscle groups that surround the hip. The principal hip flexor is the iliopsoas muscle assisted by the sartorius and rectus femoris muscles. Hip adduction is accomplished by the three adductors (longus, brevis, and magnus) plus the gracilis and pectineus muscles. The gluteus
Medius is the major hip abductor, whereas the gluteus maximus and hamstrings extend the hip. Several clinically important bursae are found around the hip joint. Anteriorly, the iliopsoas bursa lies between the psoas muscle and the joint surface. The trochanteric bursa lies between the gluteus maximus muscle and the posterolateral greater trochanter, and the ischiogluteal bursa overlies the ischial tuberosity.

Examination of the hip should begin by observing the patient’s stance and gait. The patient should stand in front of the examiner so that the anterior iliac spines are visible. Pelvic tilt or obliquity may be present and related to structural scoliosis, anatomic leg-length discrepancy, or hip disease.

Hip contractures may result in abduction or adduction deformities. To compensate for an adduction contracture, the pelvis is tilted upward on the side of the contracture. This allows the legs to be parallel during walking and weight bearing. With a fixed adduction deformity, the pelvis becomes elevated on the normal side during standing or walking. This elevation causes an apparent shortening of the normal leg and forces the patient to stand or walk on the toes of the normal side or to flex the knee on the abnormal leg. Viewed from behind with the legs parallel, the patient with hip disease and an adducted hip contracture may have asymmetric gluteal folds secondary to pelvic tilt, with the diseased side elevated. In this situation, the patient is unable to stand with the foot of the involved leg flat on the floor. In abduction contracture, the findings are reversed; with both legs extended and parallel, the uninvolved side is elevated.

A hip flexion deformity commonly occurs in diseases of the hip. Unilateral flexion of the hip in the standing position reduces weight bearing on the involved side and relaxes the joint capsule, causing less pain. This posture is best noted by observing the patient from the side. A hyperlordotic curve of the lumbar spine compensates for lack of full hip extension.

Gait should be assessed in the patient with possible hip joint disease. With a normal gait, the abductors of the weight-bearing leg contract to hold the pelvis level or to elevate the non–weight-bearing side slightly. Two abnormalities of gait may be commonly observed in patients with hip disease. The most common abnormality seen with a painful hip is the antalgic (limping) gait. With this gait, the individual leans over the diseased hip during the phase of weight bearing on that hip, placing the body weight directly over the joint to avoid painful contraction of the hip abductors. With a Trendelenburg gait, with weight bearing on the affected side, the pelvis drops and the trunk shifts to the normal side. Although the antalgic gait is frequently seen with painful hips and the Trendelenburg gait is seen in patients with weak hip abductors, these gaits are not specific, and either may occur as a result of hip pain from one of several causes. A mild Trendelenburg gait is seen often in normal individuals.

The Trendelenburg test assesses the stability of the hip, together with the ability of the hip abductor muscle to stabilize the pelvis on the femur. It is a measure of the gluteus medius hip abductor strength. The patient is asked to stand while bearing weight on only one leg. Normally, the abductors hold the pelvis level or the non–weight-bearing side slightly elevated. If the non–weight-bearing side drops, the test is positive for weakness of the weight-bearing side hip abductors, especially the gluteus medius muscle. This test is nonspecific and may be used in primary neurologic or muscle disorders and in hip diseases that lead to weakness of the hip abductors.

The motion of the hip should be assessed with the patient in the supine position. Range of motion of the hip includes flexion, extension, abduction, adduction, internal and external rotation, and circumduction. The degree of flexion permitted varies with the manner with which it is assessed. When the knee is held flexed at 90 degrees, the hip normally flexes to an angle of 120 degrees between the thigh and the long axis of the body. If the knee is held in extension, the hamstrings limit hip flexion to about 90 degrees. The presence of a hip flexion contracture is suggested by persistence of lumbar lordosis and pelvic tilt, masking the contracture by allowing the involved leg to remain in contact with the examination table. The Thomas test shows the flexion contracture. With this test, the opposite hip is fully flexed to flatten the lumbar lordosis and fix the pelvis. The involved leg should be extended toward the table as far as possible. Flexion contracture of the diseased hip becomes more obvious and can be estimated in degrees from full extension. Measurement for leg-length discrepancy is performed with the patient supine and the legs fully extended. Each leg is measured from the anterior superior iliac spine to the medial malleolus. A difference of 1 cm or less is unlikely to cause any abnormality of gait and may be considered normal. In addition to true leg-length asymmetries, apparent leg-length discrepancies may result from pelvic tilt or abduction or adduction contractures of the hip.

Abduction is measured with the patient supine and the leg in an extended position perpendicular to the pelvis. Pelvic stabilization is achieved by placing an arm across the pelvis with the hand on the opposite anterior iliac spine. With the other hand, the examiner grasps the patient’s ankle and abducts the leg until the pelvis begins to move. Abduction to about 45 degrees is normal. It is helpful to compare one side with the other because the normal range of motion may vary. Alternatively, the examiner could stand at the foot of the table, grasp both the patient’s ankles, and simultaneously abduct both legs. Abduction is commonly limited in hip joint disease. Abduction is assessed by grasping the patient’s ankle and raising the leg off the examination table by flexing the hip enough to allow the tested leg to cross over the opposite leg. Normal abduction is about 20 to 30 degrees. Hip rotation may be tested with the hip and knee flexed to 90 degrees or with the leg extended. Normal hip external rotation and internal rotation are observed to 45 degrees and 40 degrees, respectively. The difference in rotation between the flexed and extended hip is attributable to increased stabilization of the joint by surrounding ligaments in the extended position. Rotation decreases with extension. To test hip rotation, the extended leg is grasped above the ankle and rotated externally and internally from the neutral position. Limitation of internal rotation of the hip is a sensitive indicator of hip joint disease.

Extension is tested with the patient in the prone position. Estimating hip extension can be difficult because some of the apparent motion arises from hyperextension of the lumbar spine, pelvis rotation, motion of the buttock soft
tissue, and flexion of the opposite hip. The pelvis and the lumbar spine can be partially immobilized by placing an arm across the posterior iliac crest and the lower lumbar spine. The examiner places the other hand under the thigh with the knee flexed and hyperextends the thigh. Normal extension ranges from 10 to 20 degrees. Limitation of extension often occurs secondary to a hip flexion contracture.

Swelling around the hip only rarely can be discerned on examination. The flexion abduction external rotation (FABER) test, also known as the Patrick test, is a commonly used screening test for intra-articular hip pathology. To perform this test, the examiner has the patient lie supine with the foot ipsilateral to the test hip, resting on the contralateral knee. The examiner then slowly lowers the test leg toward the examining table, applying gentle pressure to the knee of the test hip and the contralateral anterior superior iliac spine. Normally, the test leg will fall at least parallel to the opposite leg. The FABER test is considered positive when the maneuver reproduces the patient’s pain. Normally, the test leg will fall at least parallel to the opposite leg. Although very sensitive for hip joint disease, this test is not specific, as a positive test may indicate iliopsoas tightness or sacroiliac joint disease.

In children, two useful screening tests for congenital hip disease are the Ortolani maneuver and the Galeazzi sign. With the Ortolani maneuver, the examiner flexes the hips and grasps the legs of a supine infant so that the examiner’s thumbs are against the inner thighs and fingers are draped across the posterior iliac crest and the lower lumbar spine. With gentle traction, the hips are abducted and laterally rotated. Resistance usually is felt at 30 to 40 degrees of lateral rotation and abduction. With a positive result of the Ortolani test, a click is felt before abduction to the normal 70 degrees. The Ortolani test should not be done repeatedly because it can lead to damage of the articular cartilage on the femoral head. The Galeazzi sign is useful for assessing unilateral congenital hip dislocation in children younger than 18 months. The child is placed supine with the knees and hips flexed to 90 degrees. Both knees should normally reside at the same level, and a positive test result is indicated if one knee is higher than the other.

The iliotibial band is a part of the fascia lata extending from the iliac crest, sacrum, and ischium over the greater trochanter to the lateral femoral condyle, tibial condyle, and fibular head, and along the lateral intermuscular system, separating the hamstring from the vastus lateralis muscle. The tensor fasciae latae muscle may produce an audible snap as it slips over the greater trochanter if the weight-bearing leg moves from hip flexion and abduction to a neutral position, as in climbing stairs. Most commonly observed in young women, the snapping hip usually does not cause severe pain. The Ober test evaluates the iliotibial band for contracture. The patient lies on the side, with the lower leg flexed at the hip and knee. The examiner abducts and extends the upper leg with the knee flexed at 90 degrees. The hips should be slightly extended to allow the iliotibial band to pass over the greater trochanter. The examiner slowly lowers the limb with the muscles relaxed. A positive test result indicative of an iliotibial band contracture occurs if the leg does not fall back to the level of the table top.

A common cause of lateral hip pain is trochanteric bursitis. Patients with this condition often complain of pain and tenderness when they attempt to lie on the affected side or climb stairs. The greater trochanter should be palpated for tenderness and compared with the opposite side. In trochanteric bursitis, this area is usually exquisitely tender. The pain of trochanteric bursitis is aggravated by actively resisted abduction of the hip. Aching and tenderness over the buttock area may be secondary to an ischial bursitis. Other causes of lateral and posterior hip (buttock) discomfort include pain at muscle and tendon insertion sites.

Anterior hip and groin pain may be secondary to hip abnormality, most commonly degenerative arthritis. Decreased range of motion should be noted in these patients. Other causes include iliopsoas bursitis, in which swelling and tenderness may be noted in the middle third of the inguinal ligament lateral to the femoral pulse. This pain is aggravated by hip extension and is reduced by flexion. The bursitis may be a localized problem or may represent extension of hip synovitis. It usually is impossible to distinguish between a localized bursitis and an extension of hip synovitis on the basis of the physical examination. If the patient is tender in the region of the iliopsoas bursa, but no swelling is palpable, the examiner should consider tendinitis of the iliopsoas muscle. The inguinal region should be palpated for other abnormalities, such as hernias, femoral aneurysms, adenopathy, tumor, and pseudoabcess or masses.

Muscle strength testing should include the hip flexors, extensors, abductors, and adductors. The primary hip flexor is the iliopsoas muscle (nerve roots L2 and L3). Flexion may be tested with the patient sitting at the edge of a table. The examiner exerts downward pressure against the thigh proximal to the knee while the patient attempts to flex the hip. The pelvis may be stabilized by the examiner’s other hand placed on the ipsilateral iliac crest. Alternatively, with the patient supine and holding the leg in 90 degrees of flexion at the hip, the examiner may attempt to straighten the hip.

Hip extension is tested with the patient lying prone. The primary hip extensor is the gluteus maximus muscle (L5 and S1). With the knee flexed to remove hamstring action, the patient is instructed to extend the hip and thigh off the surface of the table as the examiner places a forearm across the posterior iliac crest to stabilize the pelvis and applies downward pressure to prevent the lateral trunk muscles from elevating the pelvis and leg off the table.

Abduction may be tested with the patient prone or supine. The patient should abduct the thigh and leg against resistance from the examiner applied at the midhigh level. The primary adductor is the adductor longus muscle (nerve roots L3 and L4). The examiner holds the upper leg proximal to the knee in slight abduction, while the patient resists and attempts to adduct the leg. Testing for abduction and adduction also may be done in the two legs simultaneously. The patient lies supine with the legs fully extended and the hips moderately abducted. To test abduction, the patient actively pushes out against the examiner’s resistance against the lateral malleoli. Adduction is tested by movement against resistance at the medial malleoli.

**Knee**

The knee is a compound condylar joint with three articulations: the patellofemoral and the lateral and medial tibiofemoral condyles with their fibrocartilaginous menisci. The
knee is stabilized by its articular capsule, the patellar ligament, medial and lateral collateral ligaments, and anterior and posterior cruciate ligaments. The collateral ligaments provide medial and lateral stability, whereas the cruciates provide anteroposterior and rotatory stability. Normal knee motion is a combination of flexion or extension and rotation. With flexion, the tibia externally rotates, and with extension, it externally rotates on the femur. The surrounding synovial membrane is the largest of the body’s joints; it extends 6 cm proximal to the joint as the suprapatellar pouch beneath the quadriceps femoris muscle. Several important bursae are found around the knee, including the superficial prepatellar bursa, the superficial and deep infrapatellar bursae, the pes anserine bursa distal to the medial tibial plateau, and the posterior medial semimembranosus and posterolateral gastrocnemius bursae. Knee extension is primarily mediated by the quadriceps femoris muscle; knee flexion is mediated by the hamstrings. The biceps femoris muscle externally rotates the lower leg on the femur, whereas the popliteus and semitendinosus muscles mediate internal rotation.

In taking the history of a patient with knee complaints, the patient should be asked about symptoms of locking, catching, or giving way. Locking is the sudden loss of ability to extend the knee; it usually is painful and may be associated with an audible noise, such as a click or pop. It often implies extensive intra-articular abnormality, including loose bodies or cartilaginous tears. Catching refers to a subjective sensation of the patient that the knee might lock; the patient may experience a momentary interruption in the smooth range of motion of the joint but is able to continue with normal motion after this brief hesitation. Catching usually implies less abnormality than true locking and may occur in various pathologic conditions. True give-way indicates that the knee actually buckles and gives out in certain positions or with certain activities. It is important to elicit details of the history to verify this common subjective complaint. Patients often experience a sensation that the knee will give out when it actually does not. Other patients say their knees are “giving out” to describe severe pain that necessitates stopping an activity. True give-way implies severe intra-articular abnormality, such as an unstable joint from ligamentous injury or incompetence.

Examination of the knees should always include observation of the patient while standing and walking. Deviation of the knees, including genu varum (lateral deviation of the knee joint with medial deviation of the lower leg), genu valgum (medial deviation of the knee with lateral deviation of the lower leg), and genu recurvatum, is most easily appreciated with the patient standing. The patient also should be observed ambulating for evidence of gait abnormalities. Measurement of the thigh circumference should be performed at 15 cm above the knee to avoid spurious results due to suprapatellar effusions.

In taking the history of a patient with knee complaints, the patient should be asked about symptoms of locking, catching, or giving way. Locking is the sudden loss of ability to extend the knee; it usually is painful and may be associated with an audible noise, such as a click or pop. It often implies extensive intra-articular abnormality, including loose bodies or cartilaginous tears. Catching refers to a subjective sensation of the patient that the knee might lock; the patient may experience a momentary interruption in the smooth range of motion of the joint but is able to continue with normal motion after this brief hesitation. Catching usually implies less abnormality than true locking and may occur in various pathologic conditions. True give-way indicates that the knee actually buckles and gives out in certain positions or with certain activities. It is important to elicit details of the history to verify this common subjective complaint. Patients often experience a sensation that the knee will give out when it actually does not. Other patients say their knees are “giving out” to describe severe pain that necessitates stopping an activity. True give-way implies severe intra-articular abnormality, such as an unstable joint from ligamentous injury or incompetence.

Examination of the knees should always include observation of the patient while standing and walking. Deviation of the knees, including genu varum (lateral deviation of the knee joint with medial deviation of the lower leg), genu valgum (medial deviation of the knee with lateral deviation of the lower leg), and genu recurvatum, is most easily appreciated with the patient standing. The patient also should be observed ambulating for evidence of gait abnormalities.

Inspection should be done with the patient standing and supine. It is essential to compare side to side, noting any asymmetry that may be caused by swelling or muscleatrophy. Suprapatellar swelling with fullness of the distal anterior thigh that obliteratesthe normal depresed contours along the sides of the patella usually indicates knee joint effusion or synovitis. Localized swelling over the surface of the patella is generally secondary to prepatellar bursitis (Figure 40-8). Patellar alignment should be noted, including high-riding or laterally displaced patellae. The examiner also should inspect the knee from behind to identify popliteal swelling due to a popliteal or Baker cyst, most commonly caused by medial semimembranosus bursal swelling. If the calves appear asymmetric, calf circumference should be measured and compared bilaterally. Popliteal cysts may rupture and dissect down into the calf muscles, resulting in enlargement and palpable fullness. Edema may be present if the cyst causes secondary venous or lymphatic obstruction. Acute rupture and dissection of a popliteal cyst can mimic thrombophlebitis, with local pain, heat, redness, and swelling. This is probably a more common cause of unilateral calf swelling in patients with rheumatoid arthritis than is deep venous thrombosis. The two conditions may be difficult to distinguish on physical examination alone.

Quadriceps femoris muscle atrophy usually develops in chronic arthritis of the knee. Atrophy of the vastus medialis muscle is the earliest change and may be appreciated by comparing the two thighs for medial asymmetry and circumference. Measurement of the thigh circumference should be performed at 15 cm above the knee to avoid spurious results due to suprapatellar effusions.
Palpation of the knee should be performed with the joint relaxed. This usually is best accomplished with the patient supine and the knees fully extended and not touching. Palpation should begin over the anterior thigh approximately 10 cm above the patella. To identify the superior margin of the suprapatellar pouch, which is an extension of the knee joint cavity, the examiner should palpate the anterior thigh, moving distally toward the knee. Swelling, thickening, nodules, loose bodies, tenderness, and warmth should be noted. A thickened synovial membrane has a boggy, doughy consistency, which differs from the surrounding soft tissue and muscle. It usually is palpated earlier over the medial aspect of the suprapatellar pouch and the medial tibiofemoral joint. To enhance detection of knee fluid, any fluid in the suprapatellar pouch is compressed with the palm of the hand placed just proximal to the patella. The synovial fluid forced into the inferior distal articular cavity is palpated with the opposite thumb and index finger laterally and medially to the patella. If the examiner alternates compression and release of the suprapatellar pouch, the synovial thickening can be differentiated from a synovial effusion. An effusion intermittently distends the joint capsule under the thumb and index finger of the opposite hand, whereas synovial thickening does not.

The examiner should not compress the suprapatellar pouch too firmly or push the tissues distally because the patella or normal soft tissue, including the fat pads, fills the palpated space and could be misinterpreted as synovitis or joint swelling. With a large effusion, the patella can be balloted by pushing it posteriorly against the femur with the right forefinger, while maintaining suprapatellar compression with the left hand.

At the other extreme, effusions 4 to 8 mL can be detected by eliciting the bulge sign. This test is performed with the knee extended and relaxed. The examiner strokes or compresses the medial aspect of the knee proximally and laterally with the palm of a hand to move fluid from the area. The lateral aspect of the knee is tapped or stroked, and a fluid wave or bulge appears medially (Figure 40-9). A so-called spontaneous bulge sign occurs if, on compression along the medial side of the joint space, fluid reaccumulates with no pressure or compression along the lateral side of the joint.

The medial and lateral tibiofemoral joint margins are palpated for tenderness and bony lipping or exostosis, as can be seen in degenerative joint disease. Joint margins can be palpated easily with the hip flexed to 45 degrees, the knee flexed to 90 degrees, and the foot resting on the examining table. Tenderness localized over the medial or lateral joint margins may represent articular cartilage disease, medial or lateral meniscal abnormality, or medial or lateral collateral ligament injury. Other causes of tenderness include pathologic conditions in the underlying bony structures.

Bursitis is another cause of localized tenderness around the knee; the two most common sites are the pes anserine and the prepatellar bursae. Exquisite local tenderness usually can be elicited if bursitis is present. Mild swelling also may be appreciated. Occasionally, the prepatellar bursa can become quite swollen. It is important not to interpret this swelling mistakenly as knee joint synovitis. The two can be differentiated because the bursal margins can be outlined by palpation; other features of true joint effusion, such as the bulge sign, are absent.

Patellofemoral malalignment is another common cause of knee pain. It is more common in female patients because of the wider Q angle caused by the broader female pelvis. The Q angle is the angle formed between the quadriceps and the patellar tendon. Patients with patellofemoral disease may complain of stiffness in the knee after a period of flexion (the moviegoer sign) or may have particular difficulty with stair climbing. Some patients may experience a sensation of catching as the patella moves over the distal femur. Patellar palpation is best performed with the knee extended and relaxed. The patella is compressed and moved so that its entire articular surface comes into contact with the underlying femur. Slight crepitation may be observed in many normally functioning knees. Pain with crepitation may suggest patellofemoral degenerative arthritis or chondromalacia patellae.

Retropatellar pain occurring with active knee flexion and extension and secondary to patellofemoral disease may be differentiated from tibiofemoral articular pain. To test this, the examiner should attempt to lift the patella away from the knee, while passively moving the knee through the range of motion. Painless motion during this maneuver indicates that the patellofemoral joint is the likely source of the pain. In addition, the “patellar grind” test is useful in patients with extensive patellofemoral abnormality. In this test, the examiner compresses the patella distally away from the femoral condyles, while instructing the patient to contract the quadriceps isometrically. Sudden patellar pain and
quadriceps relaxation indicate a positive test result. This test has frequent false-positive results, however.

Patellar stability should be assessed. The Fairbanks apprehension test is done with the patient supine, the quadriceps relaxed, and the knee in 30 degrees of flexion. The examiner slowly pushes the patella laterally. A sudden contraction of the quadriceps and a distressed reaction from the patient constitute a positive apprehension test result. A patient who has had previous patellar dislocations usually has a positive apprehension test result. The patella also can be examined for subluxation while the knee is moved through a range of motion from full flexion to extension.

The plica syndrome occasionally causes symptoms that suggest patellofemoral disease. Plicae are bands of synovial tissue, most often located on the medial side of the knee. If present, a tender band-like structure may be palpated parallel to the medial border of the patella. During flexion and extension, palpable or audible snapping may be heard, and the patient may experience symptoms of catching. Many plicae are asymptomatic, however, and they are common, so they may be considered a normal variant.

The normal knee range of motion should be from full extension (0 degrees) to full flexion of 120 to 150 degrees. Some normal individuals may be able to hyperextend to 15 degrees. Loss of full extension that is generally reversible frequently occurs with a knee joint effusion, synovitis, or both. However, permanent loss of extension due to flexion contracture is a common finding that accompanies chronic arthritis of the knee. In advanced arthritis, such as in some cases of rheumatoid arthritis, posterior subluxation of the tibia on the femur may be observed.

Ligamentous instability is tested by applying valgus and varus stress to the knee and by using the drawer test. The knee should be extended and relaxed. The examiner performs the abduction or valgus test by stabilizing the lower femur, while placing a valgus stress on the knee by abducting the lower leg with the other hand placed proximal to the ankle. A medial joint line separation with the knee fully extended indicates a tear of the medial collateral ligament plus the posterior cruciate ligament. The test is performed with the knee in 30 degrees of flexion. If the test is negative at 0 degrees, but positive at 30 degrees, the instability represents a tear of the medial collateral ligament with the posterior cruciate ligament remaining intact. The adduction or varus test is performed with the knee extended and again at 30 degrees of flexion. Separation of the lateral joint line indicates a lateral collateral ligament tear—associated or not associated with a posterior cruciate ligament tear.

The degree of ligamentous laxity observed during testing can be graded on a scale of 1 to 3. Mild or grade 1 instability indicates that the joint surfaces separate by 5 mm or less; for moderate or grade 2 instability, a separation of 5 to 10 mm is seen. Grade 3 instability is a separation greater than 10 mm. In cases of trauma, opening of the joint space indicates ligamentous instability secondary to rupture or stretching of the ligaments. In cases of chronic arthritis of the tibiofemoral compartment, medial or lateral separation may be apparent as a result of the “pseudolaxity” created by loss of cartilage and bone. If the ligaments are intact, the resulting degree of valgus or varus displacement with stressing is not any greater than in the normal knee.

The drawer test is performed with the hip flexed to 45 degrees and the knee flexed to 90 degrees. To stabilize the knee, the examiner either sits on the foot while grasping the posterior calf with both hands or supports the lower leg between his or her lateral chest wall and forearm. The anterior drawer test is performed by pulling the tibia forward. More than 6 mm of movement is abnormal and may indicate an anterior cruciate tear or laxity. Anterior subluxation may represent more complex instability, however. Rotatory instability of the knee also may exist. A positive result of the anterior drawer test, in which the lateral tibial plateau subluxates forward while the medial stays in normal position, can represent anterolateral rotatory instability. If both plateaus subluxate, tears of the middle third of the medial collateral capsular ligaments may be present. If the subluxation is not present with the tibia internally rotated, the posterior cruciate ligament is intact. A positive result of the anterior drawer test with the leg in external rotation represents a tear of the medial capsular ligament.

The Lachman test is a modification of the anterior drawer sign and tests for one-plane anterior instability. At least six variations of this test have been described. In the test as originally described, the patient lies supine with the tested knee between full extension and 15 degrees of flexion. The femur is stabilized with a hand of the examiner while the hand pulls the proximal aspect of the tibia forward. A positive test result is indicated by a soft feel rather than a firm end point when the tibia moves forward on the femur. A positive result of the Lachman test may indicate an anterior cruciate injury or abnormality in the posterior oblique ligament or arcuate popliteus complex. A posterior drawer test may be done with the patient positioned as for an anterior drawer test, but the examiner pushes the tibia toward the patient. A positive test result suggests damage to the posterior cruciate ligament.

During the complete joint examination, tests for meniscal injury should be performed. Symptoms suggesting a meniscal tear include locking during joint extension, clicking or popping during motion, and localized tenderness along the medial or lateral joint line. To examine the menisci, the medial and lateral joint line should be palpated with the lower leg internally rotated and the knee flexed to 90 degrees. Localized tenderness over the medial or lateral joint line suggests involvement of the medial or lateral meniscus. The McMurray test evaluates for evidence of meniscal tear, especially in the posterior half of the menisci. The patient’s knee is placed in full flexion, and the examiner places a hand over the knee with the fingers along the side of the knee over the joint line and the thumb along the other side. The other hand holds the leg at the ankle and is used to rotate the lower leg medially and to apply varus stress. This test can be done repeatedly, with the knee in gradually decreasing degrees of flexion. A palpable or audible snap suggests a tear of the medial meniscus. The test can be done in a similar fashion by laterally rotating the tibia and applying valgus stress to test for a lateral meniscal injury. A positive result of a lateral test may represent a tear of the popliteus tendon, which can accompany a lateral meniscal tear.

The Apley grind test also evaluates for a torn meniscus. With the patient lying prone and the knee flexed to 90 degrees, the examiner places downward compression on the
foot, while medially and then laterally rotating the tibia on the femur. Pain elicited during this maneuver suggests a meniscal tear.

The examiner performs the distraction test by placing his or her knee on the patient's posterior thigh to stabilize the leg while applying an upward distractive force on the foot. Pain from rotating the tibia suggests ligament damage.

A simple hyperflexion test is a useful screening test for meniscal damage. If the examiner is able to hyperflex the knee to more than 135 degrees without eliciting pain, it is unlikely that serious cartilaginous injury is present. If pain occurs with hyperflexion, the patient sometimes is able to localize it medially or laterally, which often correlates with the location of the meniscal injury. Although helpful, none of these tests for meniscal injury is completely reliable when verified by arthroscopy. Tenderness along the joint line is most sensitive but is less specific than manipulative tests such as the McMurray test.

Muscle strength testing includes testing flexion supplied by the hamstrings (i.e., the biceps femoris, semitendinosus, and semimembranosus) (nerve roots L5 to S3) and extension supplied by the quadriceps femori (L2, L3, and L4). The hamstrings are tested best with the patient prone and attempting to move the knee from 90 degrees to maximal flexion. The ankle should be kept in neutral position or dorsiflexed to remove gastrocnemius action. With the leg externally rotated, the biceps femoris, which inserts on the fibula and lateral tibia, is primarily tested, whereas flexion with internal rotation tests the semitendinosus and semimembranosus muscles, which insert on the medial side of the tibia. Extension is tested with the patient sitting upright with the knee fully extended. The examiner stabilizes the thigh with downward pressure just proximal to the knee and places downward pressure at the ankle to test the knee extensors.

### Ankle

The true ankle is a hinged joint, and movement is limited to planar flexion and dorsiflexion. It is formed by the distal ends of the tibia and fibula and the proximal aspect of the body of the talus. Inversion and eversion occur at the subtalar joint (see Chapter 49). The tibia forms the weight-bearing portion of the ankle joint, whereas the fibula articulates on the side of the tibia. The malleoli of the tibia and fibula extend downward beyond the weight-bearing part of the joint and articulate with the sides of the talus. The malleoli provide medial and lateral stability by enveloping the talus in a mortise-like fashion. The articular capsule of the ankle is lax on the anterior and posterior aspects of the joint, allowing extension and flexion, but it is tightly bound bilaterally by ligaments. The synovial membrane of the ankle on the inside of the capsule usually does not communicate with any other joints, bursae, or tendon sheaths.

The medial and lateral ligaments surrounding the ankle contribute to medial and lateral stability of the joint. The deltoid ligament, the only ligament on the medial side of the ankle, is a triangle-shaped fibrous band that resists evasion of the foot. It may be torn in inversion sprains of the ankle. The lateral ligaments of the foot consist of three distinct bands forming the posterior talofibular, the calcaneofibular, and the anterior talofibular ligaments. These ligaments may be injured in inversion sprains of the ankle.

All tendons crossing the ankle joint lie superficial to the articular capsule and are enclosed in synovial sheaths for part of their course across the ankle. On the anterior aspect of the ankle, the tendons and the synovial tendon sheaths of the tibialis anterior, extensor digitorum longus, peroneus tertius, and extensor hallucis longus muscles overlie the articular capsule and synovial membrane. On the medial side of the ankle, posterior and inferior to the medial malleolus, lie the flexor tendons and tendon sheaths of the tibialis posterior, flexor digitorum longus, and flexor hallucis longus muscles (Figure 40-10). All three of these muscles plantar flex and supinate the foot. The tendon of the flexor hallucis longus is located more posteriorly than the other flexor tendons and lies beneath the Achilles tendon for part of its course. The calcaneus tendon (Achilles tendon), the common tendon of the gastrocnemius and soleus muscles, inserts into the posterior surface of the calcaneus, where it is subject to external trauma, various inflammatory

---

**Figure 40-10** Diagram of the ankle. Medial aspect of the ankle shows the relationship between tendons, ligaments, artery, and nerve. (From Polley HF, Hunder GG: Rheumatologic interviewing and physical examination of the joints, ed 2, Philadelphia, 1978, WB Saunders. Used with permission of Mayo Foundation for Medical Education and Research.)
reactions, and irritations from bone spurs beneath it. On the lateral aspect of the ankle, posterior and inferior to the lateral malleolus, a synovial sheath encloses the tendons of the peroneus longus and peroneus brevis. These muscles extend the ankle (plantar flex) and evert (pronate) the foot. Each of the tendons adjacent to the ankle may be involved separately in traumatic or disease processes.

Three sets of fibrous bands or retinacula hold down the tendons that cross the ankle in their passage to the foot. The extensor retinaculum consists of a superior part (transverse crural ligament) in the anterior and inferior portions of the leg and an inferior part in the proximal portion of the dorsum of the foot. The flexor retinaculum is a thickened fibrous band on the medial side of the ankle. On the lateral side of the ankle, the peroneal retinaculum forms a superior and an inferior fibrous band. These bands bind down tendons of the peroneus longus and peroneus brevis muscles as they cross the lateral aspect of the ankle.

Synovial swelling of the ankle joint is most likely to cause fullness over the anterior or anterolateral aspect of the joint because the capsule is more lax in this area. Mild swelling of the joint may not be apparent on inspection because of the many structures crossing the joint superficially. Efforts should be made to differentiate superficial linear swelling localized to the distribution of the tendon sheaths from more diffuse fullness and swelling attributable to involvement of the ankle joint. Swelling of the heels may be observed from behind the standing patient and may be caused by enthesitis of the Achilles tendon insertion, which can occur in spondyloarthropathies (Figure 40-11).

It is difficult to observe synovitis of the intertarsal joints. Intertarsal joint synovitis may produce an erythematous puffiness or fullness over the dorsum of the foot.

From the normal position of rest in which there is a right angle between the leg and the foot, labeled 0 degrees, the ankle normally allows about 20 degrees of dorsiflexion and about 45 degrees of plantar flexion. Inversion and eversion of the foot occur mainly at the subtalar and other intertarsal joints. From the normal position of the foot, the subtalar joint normally permits about 20 degrees of eversion and 30 degrees of inversion. To test the subtalar joint, the examiner grasps the calcaneus with a hand and attempts to invert and evert it, holding the ankle motionless.

A general assessment of muscular strength of the ankle can be obtained by asking the patient to walk on toes and on heels. If the patient can walk satisfactorily on the toes and on the heels, the muscle strength of the flexors and extendors of the ankle can be considered normal. If this cannot be accomplished, it is desirable to test the muscles individually.

The principal flexors of the ankle are the gastrocnemius (nerve roots S1 and S2) and the soleus (S1 and S2) muscles. The principal extensor (dorsiflexors) of the ankle is the tibialis anterior muscle (L4, L5, and S1). The tibialis posterior muscle (L5 and S1) is the principal inverter. To test the tibialis posterior muscle, the foot should be in plantar flexion. The examiner applies graded resistance on the medial border of the forefoot while the patient attempts to invert the foot. The principal everters of the foot are the peroneus longus (L4, L5, and S1) and peroneus brevis (L4, L5, and S1) muscles.

Foot

See Chapter 49.

References


The references for this chapter can also be found on www.expertconsult.com.
The sudden onset of monoarthritis in a synovial joint is a significant clinical event. The most serious cause is joint infection because of associated mortality (≈15%) and probability of impaired long-term function. However, the commonest of the differential diagnoses (Table 41-1) is trauma or inflammation. Each presentation of monoarthritis requires immediate investigation and treatment to limit pain and prevent joint destruction. The British Society of Rheumatology and British Orthopaedic Association generated guidelines (Figure 41-1) to assist physicians in the diagnosis of the acutely swollen joint based on history, examination, radiologic, and blood and synovial investigations.

Monoarthritis can also be a new presentation of more chronic forms of arthritis, and monoarticular pain can result from noninflammatory arthropathies, periarticular problems, and bone and soft tissue disorders. These are discussed in the context of a formal differential diagnosis.

**HISTORY**

One should examine the patient's age, occupation, and social, drug, travel, and sexual history. Pain and joint stiffness, its diurnal variation, and aggravating and relieving factors should be sought, and a history of swelling should be established. Specific features of relevance include trauma, joint locking, and presence of systemic symptoms (fevers, sweats, rigors, and weight loss). Inquiring about ocular, oral, respiratory, gastrointestinal, or skin symptoms can facilitate the diagnosis. Inquiring about ocular, oral, respiratory, gastrointestinal, or skin symptoms can facilitate the diagnosis if the monoarthritis is a presentation of a chronic arthropathy, related inflammatory disease, postinfectious phenomenon, or connective tissue disease. Table 41-2 summarizes symptoms and signs that characterize patients with inflammatory or noninflammatory joint disease and soft tissue disorders to help initial evaluation.

**EXAMINATION**

The Regional Examination of the Musculoskeletal System (REMS) can be used to identify monoarticular abnormalities. Physicians should look, feel, and move joints to assess function—comparing the affected joint and normal side and for evidence of more widespread disease. In any acutely swollen joint, examination should include looking for local signs of inflammation (pain, erythema, swelling, heat, and loss of function). Any patient with preceding arthritis should be assessed for chronic changes to joint structure and disability. Local synovial swelling and/or effusion, joint instability, limited movement, and deformity of any single joint necessitate detailed investigation. Local monoarticular tenderness without swelling could indicate enthesitis, tendinitis, bursitis, or bone disease. A complete examination should look specifically for ocular signs, skin rashes, ulcers, and nodules.

**INVESTIGATIONS**

**Blood**

Inflammatory, septic, and crystal arthritides cause elevated erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), and white cell count (WCC), often associated with anemia. Systemic disease involvement is assessed by renal, liver, muscle, or bone biochemical screening and protein electrophoresis. Raised uric acid levels suggest a diagnosis of gout. In acute hemarthrosis, a platelet count, international normalized ratio, and clotting studies are warranted. Blood cultures are mandatory in patients with suspected septic arthritis and should precede antibiotic prescription. Viral screening (IgG and IgM antibodies), antistreptolysin-O test (ASOT), and Lyme serology can be diagnostic in relevant situations.

Rheumatoid factor and antinuclear antibodies (ANAs) suggest rheumatoid arthritis (RA) or systemic lupus erythematosus (SLE), but autoantibodies can be present in infectious diseases, in other autoimmune diseases, and in normal individuals. Similar limitations apply to results for

---

**KEY POINTS**

- Inflammatory arthritis causes morning stiffness and pain with restricted joint movement.
- Pain worsening after movement usually indicates mechanical joint pain.
- Soft tissue disorders usually produce pain in a periarticular distribution.
- Acute monoarthritis presentations (developing suddenly) are usually related to trauma.
- Monoarthritis (developing over 1 to 2 days) presentation is usually inflammation- or infection-related.
- Chronic monoarthritis presentation indicates pre-existing joint disease, but causes of acute monoarthritis need to be excluded.
- Synovial fluid analysis is the most valuable laboratory test in acute monoarthritis.
antineutrophil cytoplasmic antibodies (ANCAs) and angiotensin-converting enzyme (ACE), so positive results should be applied in the clinical context. Anti–cyclic citrullinated protein (CCP) antibodies are sensitive for RA and are useful in early presentation. Other relevant blood tests include thyroid function, ferritin, and vitamin D level and analysis of human leukocyte antigen (HLA) genes.

**Urine**

The urinary tract can be a source of gram-negative bacteria in septic arthritis in the elderly. Significant proteinuria and/or hematuria and red cell casts indicate renal damage in SLE, vasculitis, or subacute bacterial endocarditis.

---

### Table 41-1 Causes of Acute Joint or Periarticular Pain

<table>
<thead>
<tr>
<th>Common Acute Monoarthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Septic arthritis (nongonococcal, gonococcal)</td>
</tr>
<tr>
<td>Crystal arthritis (gout, pseudogout)</td>
</tr>
<tr>
<td>Reactive arthropathies</td>
</tr>
<tr>
<td>Lyme disease</td>
</tr>
<tr>
<td>Plant thorn synovitis</td>
</tr>
<tr>
<td>Other infections (mycobacterial, viral, soft tissue)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trauma or Internal Derangement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loose bodies</td>
</tr>
<tr>
<td>Stress fractures</td>
</tr>
<tr>
<td>Ischemic necrosis</td>
</tr>
<tr>
<td>Hemarthrosis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Acute Monoarthritis of Polyarthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psoriatic arthritis</td>
</tr>
<tr>
<td>Enteropathic arthritis</td>
</tr>
<tr>
<td>Rheumatoid arthritis/palindromic rheumatism</td>
</tr>
<tr>
<td>Juvenile inflammatory arthropathies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Monoarthropathies from Noninflammatory Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osteoarthritis</td>
</tr>
<tr>
<td>Charcot's joints</td>
</tr>
<tr>
<td>Storage diseases (hemochromatosis, ochronosis)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Synovial Diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pigmented villonodular synovitis</td>
</tr>
<tr>
<td>Lipoma arborescens</td>
</tr>
<tr>
<td>Synovial osteochondromatosis</td>
</tr>
<tr>
<td>Reflex sympathetic dystrophy</td>
</tr>
<tr>
<td>Sarcoïdosis</td>
</tr>
<tr>
<td>Amyloid</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Acute Monoarthritis of Systemic Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>Vasculitides (antineutrophil cytoplasmic antibody positive and negative)</td>
</tr>
<tr>
<td>Henoch-Schönlein purpura</td>
</tr>
<tr>
<td>Behçet's disease</td>
</tr>
<tr>
<td>Bacterial endocarditis</td>
</tr>
<tr>
<td>Familial Mediterranean fever</td>
</tr>
<tr>
<td>Relapsing polychondritis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Soft Tissue Lesions</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Bone Diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paget's disease</td>
</tr>
<tr>
<td>Osteomyelitis (Brodie's abscess)</td>
</tr>
<tr>
<td>Osteogenic/osteoid tumors</td>
</tr>
<tr>
<td>Metastatic disease</td>
</tr>
<tr>
<td>Pulmonary hypertrophic osteoarthropathy</td>
</tr>
</tbody>
</table>

---

*This table shows the causes of inflammation in any one joint (monoarthritis) and pain around the joint that presents without inflammation (monoarthropathy).*

---

**Imaging Studies**

A range of imaging studies assist in diagnosis of acute monoarthritis:

1. Plain radiographs identify soft tissue swelling, calcium in periarticular tissues, fractures, local bone disease, and loose bodies, as well as destructive changes in long-standing arthropathies.
2. Computed tomography (CT): CT scanning better identifies fractures, bone diseases, and intra-abdominal and chest pathology. It is useful when magnetic resonance imaging (MRI) is contraindicated. In acute arthritis, CT scan can show osteomyelitis over and above acute inflammation.
3. Musculoskeletal ultrasound (US): US is increasingly reliable for use in rheumatology practice provided practitioners are appropriately trained. US is quick, efficient, and cheap and as effective as MRI and clinical examination in detecting synovitis and soft tissue damage in acute shoulder injury. In acute monarthrosis US can show loculated synovial fluid to better target aspiration and injection, and power Doppler views can demonstrate increased blood flow in active synovitis. US is recommended for detection of synovitis by the European League Against Rheumatism (EULAR) in early arthritis.
4. MRI: Although it is the best technique for soft tissue imaging, MRI is time consuming and costly. MRI can diagnose internal ligament damage and tendon enthesitis and is most effective in identifying avascular necrosis of bone. MRI is also useful when identifying the extent of inflammation in acute monarthrosis and subclinical joint involvement. EULAR also suggests MRI for diagnosis of acute early arthritis.
5. Arthrography: Imaging internal joint structure after injection of radiopaque solutions in association with CT scanning is useful for hip cartilage tears and in situations where MRI is not feasible.
6. Radionuclide scans. Bone scanning with technetium-labeled methylene dipiphosphonate identifies osteoid osteomas, bone sarcomas, bony metastases, osteomyelitis, and stress fractures not seen on plain radiographs. Bone scintigraphy is helpful when excluding bone and joint disorders in patients with chronic pain syndromes. Although not specific in acute arthropathies, bone scans show differences in the pattern of joint involvement between inflammatory conditions and osteoarthritides. Labeled white cell scans can identify areas of infection, especially when the source of infection is uncertain in patients with septic arthritis.

**Synovial Fluid**

The most useful test in acute monoarthritis is examination of synovial fluid, which should be analyzed for color and cloudiness (Table 41-3). One should use microscopy to characterize the predominant cells, Gram stain to detect bacteria, and polarized light analysis to identify uric acid or calcium pyrophosphate dehydrate (CPPD) crystals. Identification of bacteria following synovial fluid culture can provide results even when a Gram stain is negative.
**Neisseria** organisms are fastidious with low yields from culture, but polymerase chain reaction (PCR) can detect Neisseria-specific deoxyribonucleic acid (DNA).

Many centers measure synovial fluid WCC, which may differentiate inflammatory and noninflammatory presentations but crucially fails to differentiate between inflammation and infection. Low synovial fluid glucose and high lactate\(^\text{10}\) and procalcitonin\(^\text{11}\) are found in septic arthritis, but evidence that these can separate infectious from inflammatory arthritis in a clinically meaningful way is limited. Bacterial identification in synovial fluid from patients with septic arthritis can be inhibited by prior antibiotic use, and hence a careful drug history should be elicited.\(^\text{12}\) CPPD crystals are difficult to detect, so in both cases a clinical diagnosis should not be excluded if symptoms, signs, and investigations are suggestive.

**Synovial or Bone Biopsy**

Arthroscopic synovial biopsy is rarely necessary, but in tuberculosis, sarcoidosis, amyloid, pigmented villonodular synovitis, lipoma arborescens, and foreign body synovitis it

---

**Table 41-2**  Symptoms and Signs from Which Patients Complain When Having Inflammatory Joint Pain Such as Septic Arthritis, Noninflammatory Joint Pain Such as Osteoarthritis, and Soft Tissue Problems Such as Tendinitis or Bursitis*

<table>
<thead>
<tr>
<th>Feature</th>
<th>Inflammatory Joint Pain</th>
<th>Noninflammatory Joint Pain</th>
<th>Soft Tissue Injury Tendon/Bursa, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptoms</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morning stiffness</td>
<td>Usually &gt; 30 min</td>
<td>Local usually &lt; 30 min</td>
<td>Localized and brief</td>
</tr>
<tr>
<td>Constitutional symptoms</td>
<td>Present (fever, malaise)</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Time of major discomfort</td>
<td>After prolonged inactivity</td>
<td>After prolonged use</td>
<td>During and after use</td>
</tr>
<tr>
<td>Locking/instability</td>
<td>Unlikely in acute joint disease</td>
<td>Suggests internal joint derangement</td>
<td>Unusual unless tendon damage/tear present</td>
</tr>
<tr>
<td><strong>Signs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swelling</td>
<td>Common</td>
<td>Can be bone</td>
<td>Unusual</td>
</tr>
<tr>
<td>Tenderness</td>
<td>Diffuse over joint space</td>
<td>Mild over joint line</td>
<td>Localized periarticular</td>
</tr>
<tr>
<td>Inflammation</td>
<td>Common</td>
<td>Unusual</td>
<td>Over tendon/bursa</td>
</tr>
<tr>
<td>Instability</td>
<td>Uncommon</td>
<td>Occasional</td>
<td>Uncommon</td>
</tr>
<tr>
<td>Multisystem disease</td>
<td>More common</td>
<td>No</td>
<td>Unusual</td>
</tr>
</tbody>
</table>

*Adapted from 1996 ACR guidelines for evaluation of adults with acute musculoskeletal symptoms.

is often required for diagnosis.\textsuperscript{13} Proteomic and genomic assays begin to show differences in the synovium from chronic arthritides, so synovial biopsy use may increase in the future. Bone biopsy may be needed to identify tumors and relevant factors in nonresolving osteomyelitis.

**SPECIFIC DIAGNOSES**

**Acute Monoarthritis**

**Septic Arthritis (Bacterial)**

The presence of bacteria in synovial fluid or after culture is a medical emergency because the acute mortality of septic arthritis is approximately 15%.\textsuperscript{14} Large limb joints are most frequently involved,\textsuperscript{15} usually associated with underlying osteoarthritis or inflammatory arthropathies, especially RA.\textsuperscript{14,16} Patients are particularly at risk following joint surgery, arthroplasty, or intra-articular injection, as are patients with a distant infection, intravenous drug users,\textsuperscript{17} and those with underlying disease or drugs that impair the immune response.\textsuperscript{18} Recent studies have shown that the most deprived in our communities are also at risk.\textsuperscript{14}

History often elicits an acute, painful, swollen monoarthritis, but it is noteworthy that more than 10% of patients with native joint infections present with polyarticular infection. At presentation, an identifiable distant site of infection may be found, but fever occurs in less than 50% and sweats/rigors occur in approximately 30%.\textsuperscript{14} Blood cultures may identify bacteria in patients where synovial fluid culture is negative. At presentation, ESR and CRP are almost invariably raised, but approximately 35% of all septic arthritis patients do not have a raised WCC, and when infection is in the context of underlying RA this figure rises to 50%. Impaired renal and liver function at presentation can predict a poorer outcome.\textsuperscript{14}

Plain radiographs show soft tissue swelling, and ultrasound localizes synovitis and fluid collection to target aspiration. MRI may demonstrate additional osteomyelitis. Synovial fluid is usually cloudy with a high WCC, and Gram stain should be undertaken.\textsuperscript{1} Synovial fluid culture is mandatory, yielding an organism after culture in

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Cells</th>
<th>Microorganisms</th>
<th>Appearance</th>
<th>Imaging Modality</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial arthritis</td>
<td>Neutrophils</td>
<td>Gram stain</td>
<td>Turbid/pus</td>
<td>Aspiration</td>
<td>Systemic symptoms, Gram stain culture, Blood and synovial fluid</td>
</tr>
<tr>
<td></td>
<td>10,000-100,000</td>
<td>usually positive</td>
<td></td>
<td>to dryness; may</td>
<td>need ultrasound</td>
</tr>
<tr>
<td>Gonococcal arthritis</td>
<td>Neutrophils</td>
<td>Gram stain</td>
<td>Turbid/pus</td>
<td>Aspiration</td>
<td>Systemic symptoms, Gram stain culture, Blood and synovial fluid</td>
</tr>
<tr>
<td></td>
<td>10,000-100,000</td>
<td>usually positive</td>
<td></td>
<td>to dryness; may</td>
<td>need ultrasound</td>
</tr>
<tr>
<td>Crystal arthritis</td>
<td>Neutrophils</td>
<td>—</td>
<td>Turbid/pus</td>
<td>XR, CPPD</td>
<td>Presence of appropriate crystals, Acute serum urate unreliable</td>
</tr>
<tr>
<td></td>
<td>10,000-100,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuberculous arthritis</td>
<td>Mononuclear</td>
<td>Acid-fast stain</td>
<td>Turbid/pus</td>
<td></td>
<td>At-risk population, Ziehl-Neelsen stain biopsy may be necessary</td>
</tr>
<tr>
<td></td>
<td>5000-50,000</td>
<td>often negative</td>
<td></td>
<td></td>
<td>Serum autoantibodies such as RF, ACPA, ANA</td>
</tr>
<tr>
<td>Inflammatory</td>
<td>Neutrophils</td>
<td>—</td>
<td>Slightly turbid</td>
<td>Ultrasound/MRI for early synovitis and erosions</td>
<td>Systemic symptoms, Gram stay culture, Blood and synovial fluid</td>
</tr>
<tr>
<td>monoarthropathies</td>
<td>5000-50,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>Mononuclear</td>
<td>—</td>
<td>Clear</td>
<td>XR changes</td>
<td>Usually noninflammatory, CPPD may be present, Arthroscopy may be</td>
</tr>
<tr>
<td></td>
<td>0-2000</td>
<td></td>
<td></td>
<td></td>
<td>necessary, Serum autoantibodies such as RF, ACPA, ANA</td>
</tr>
<tr>
<td>Internal derangement</td>
<td>Red blood cells</td>
<td>—</td>
<td>Clear/turbid</td>
<td>MRI</td>
<td></td>
</tr>
<tr>
<td>Trauma</td>
<td>Red blood cells</td>
<td>—</td>
<td>Clear/turbid</td>
<td>XR</td>
<td>Tc bone scan may aid diagnosis if radiograph normal</td>
</tr>
<tr>
<td>Ischemic necrosis</td>
<td>—</td>
<td></td>
<td>MRI in early disease</td>
<td></td>
<td>XR abnormal only in advanced cases</td>
</tr>
</tbody>
</table>

**Table 41-3 Synovial Fluid Characteristics in the Clinical Situations, with Imaging and Investigation Techniques Best Used to Identify the Cause**

ACPA, anticitrullinated protein antibody; ANA, antinuclear antibody; CPPD, calcium pyrophosphate dehydrate deposition; CXR, chest radiograph; MRI, magnetic resonance imaging; PVNS, pigmented villonodular synovitis; RF, rheumatoid factor; SF, synovial fluid; XR, radiograph.
approximately 50% of cases—local protocols for antibiotic use should be instigated therefore before culture results are available. Organisms detected most commonly include staphylococci (Staphylococcus aureus and Staphylococcus epidermidis), streptococci, and gram-negative bacteria with an increasing prevalence of methillin-resistant S. aureus. Prompt intervention can reduce mortality hence the joint must be aspirated to dryness, but this can usually be done without recourse to surgical intervention.

Interestingly, results from history, examination, routine investigations, and outcome are identical when comparing patients with bacteria-proven septic arthritis with those where a clinical diagnosis of septic arthritis is made but where bacteria cannot be isolated. Hence absence of systemic symptoms and normal investigations should not negate a strong clinical diagnostic suspicion.

Infection with Neisseria gonorrhoeae can be asymptomatic but should be suspected in sexually active patients presenting with migratory arthralgias, tenosynovitis, skin rashes, and vesicles. Untreated Neisseria infection can also lead to destructive arthritis. Most patients will be febrile with a raised acute phase response and blood WCC, but as with other bacterial causes of septic arthritis these may be normal at presentation. Investigations should include swabs from the urethra, cervix, pharynx, and rectum, inoculated immediately on Thayer-Martin medium. In contrast with other bacterial causes of septic arthritis these may be normal.

Fever is present in 34% of patients, especially in polyarticular presentations. In acute settings, blood WCC, ESR, and CRP are raised and to this extent gout mimics septic arthritis. Serum uric acid may be raised, but levels are low in 33% during acute attacks. Renal and liver function should be assessed. Needle-shaped uric acid crystals (that are negatively birefringent under polarized light) are present in synovial fluid or tophi aspirate and confirm the diagnosis. Synovial fluid should be examined for bacteria to exclude concomitant septic arthritis.

Routine radiographs frequently show no bony abnormalities but may identify erosions after repeated or prolonged attacks. Diagnostic ultrasound and MRI can identify synovial fluid for aspiration, tophi, and erosive disease.

**Pseudogout**

Patients with pseudogout or pyrophosphate (CPPD) arthropathy present with similar symptoms, usually in the knee or wrist in older females often concurrent with osteoarthritis. Acute attacks often occur following a trigger such as infection, trauma, or surgery. Calcification can be seen in cartilage, and periarticular tissues can be seen on a radiograph. Ultrasound may assist in the diagnosis of pyrophosphate arthropathy because crystal deposits may be observed by careful practitioners. Synovial fluid microscopy demonstrates rhomboid-shaped crystals at ×400 magnification. Culture should exclude coexistent septic arthritis. Hemarthrosis necessitates review for an occult fracture. Repeat imaging (e.g., using MRI) may be necessary to formally exclude bone injury if clinical suspicion exists.

**Acute Calcific Periarthritis**

Calcium deposition in periarticular tissues is common adjacent to upper and lower limb large joints. Many patients are asymptomatic, but an acute shoulder monoarthropathy with loss of function is well recognized. Calcium crystals can be associated with subacromial bursitis, identified on routine radiographs, ultrasound, or MRI. Hypercalcemia necessitates that hyperparathyroidism should be excluded.

**Calcium Phosphate Crystal Arthritis**

Intra-articular deposits of basic calcium phosphate (BCP) are rare but present in older female patients with osteoarthritis, with a destructive shoulder arthropathy usually on the dominant side (Milwaukee shoulder) as an acute on chronic monoarthritis. The effusion is not inflammatory, but synovial fluid can be viscous and blood-stained and may contain calcium aggregates and cartilage fragments. Plain radiographs show upward shoulder dislocation, and MRI exhibits characteristic features and may be useful in differential diagnosis, which should include sepsis.

**Cholesterol Crystal Arthritis**

Cholesterol crystals have been reported in synovial fluid, albeit rarely and often in association with inflammatory arthropathies. Whether these large rhomboid-shaped crystals are truly a separate cause of synovial inflammation remains a subject for speculation.

**Reactive Arthritis**

Reactive arthritis characteristically presents as a large joint monoarthritis, a flitting polyarthritis, or enthesitis with a history of preceding throat, urogenital, or gastrointestinal tract infection. Patients can present with coexistent circinate balanitis, sterile urethritis, and keratoderma blennorrhagica. Uveitis may precede arthritis and requires urgent review. Urinary cultures should be performed. The findings of an inflammatory synovial fluid with no bacteria after culture and lack of crystals should exclude other acute arthritides. The acute phase response is usually high, and conventional radiologic investigations and ultrasound can identify synovial swelling and effusions, but MRI is particularly useful for identifying enthesitis and extra-articular soft tissue disease. The high radiation exposure associated with standard sacroiliac joint radiographs has led to the increased use of MRI to assess for active sacroilitis in patients with coexistent back pain. Autoantibody tests are usually negative, and the link with HLA-B27 may be suggestive but is of limited diagnostic use. Antibodies against Chlamydia,
Shigella, Salmonella, and Campylobacter can be found but are often negative.

Lyme Disease

Patients with Lyme disease live in, or travel through, recognized at-risk geographic areas and present with expanding erythematous rashes after a “tick bite” resulting from Borrelia infection, usually Borrelia burgdorferi. Large joint monoarthritis develops a few weeks after initial infection associated with a high ESR and CRP. Specific IgG antibodies, detected 4 weeks after infection, are diagnostic. Low levels of rheumatoid factor and ANAs can be detected. Synovial fluid contains polymorphs, and Borrelia organisms may be cultured, but PCR of synovial fluid for Borrelia DNA represents a superior test.

Plant Thorn Synovitis

Foreign bodies including plant thorns cause inflammation in intra-articular and tendon synovial tissue in hands or feet, but history of penetrating injury may be absent. Ultrasound, CT, and MRI are helpful in localizing foreign bodies, but synovial biopsy can make the diagnosis. Synovial fluid can be sparse but may be infected with Enterobacter agglomerans, a gram-negative bacillus commonly found in soil.

Other Organisms Producing Monoarthritis

Monoarthritis resulting from tuberculosis should be considered in at-risk populations and people with a relevant social history. Synovial fluid contains mononuclear cells, but synovial biopsy may be required to identify organisms.

Polyarthritis after viral infection is common. Rubella, herpes, and hepatitis B and C infections are reported to present as a monoarthritis. Arthropathies occur during infection with and treatment for human immunodeficiency virus. Other features in the presenting history and examination findings can be diagnostically useful.

Arthritis, tenosynovitis, and bone lesions are reported following fungal and parasitic infections, but most infectious lesions are extensions of underlying disease.

A reactive arthritis (Poncet’s disease) can occur after tuberculosis infection elsewhere and may also follow viral and fungal infections.

TRAUMA AND INTERNAL DERANGEMENT

Trauma, either acute or after repeated injury, is the commonest cause of acute monoarticular pain, especially in the knee and the ankle. In the knee, torn menisci or loose bodies in the synovial fluid “wedge” between articulating surfaces leading to sudden and painful locking and weakness when walking described as “giving way.” History is diagnostic and examination using REMS can elicit locking using McMurray’s test. Examination for other ligament damage is mandatory, using tests for cruciate or collateral knee ligament stability. Careful attention to inversion and eversion stability of the ankle is also essential for monoarticular pain in that area. Plain radiographs may demonstrate abnormal architecture, dislocation, or loose bodies, but MRI will usually establish the cause of a trauma-related diagnosis. Arthrography may be required to assess the hip for damage, particularly to establish tears in the acetabular labrum.

Stress fractures may be incomplete but cause monoarticular pain on weight bearing and occur after repeated minor trauma (e.g., March metatarsal fracture). Note, however, that stress fractures may occur secondary to underlying local or systemic bone disease, particularly in sedentary individuals and those on prolonged bisphosphonate therapy. These can be missed on standard radiographs, and so CT, MRI, or bone scintigraphy can be helpful in the context of persistent localizing regional articular pain.

Osteonecrosis commonly affects the hip in children (Legg-Calvé-Perthe) but can involve metatarsal heads (Freiberg), humerus, or lunate. Osteonecrosis after neck of femur fracture results from reduced vascular supply. This is also observed in recipients of systemic glucocorticoids and in patients with connective tissues diseases, especially SLE. Other causes of osteonecrosis include decompression sickness, hemoglobinopathies, and hyperlipidemias in alcoholics and patients with hyperuricemia. History is of monoarticular pain, but examination can be normal. Plain radiographs are often normal in early phases of the disease, but early MRI is diagnostic.

The presence of acute hemarthrosis suggests trauma, overanticoagulation, or inherited clotting abnormalities. Hemangiomas, lipoma arborescens, and pigmented villo nodular synovitis may be suspected in the context of recurrent effusions that are resistant to local interventions such as glucocorticoid injection. Ultrasound and particularly MRI represent the investigations of choice before proceeding to biopsy if these are suspected.

POLYARTHRITIS PRESENTING AS ACUTE MONOARTHRITIS

Patients with a personal or family history of psoriasis may develop monoarthritis, digital dactylitis, or enthesitis. Radiographs may show proliferative new bone formation along the metacarpal shaft or periarticular erosions in longstanding disease. Ultrasound and MRI can differentiate inflammation at tendon insertions from synovitis and can localize synovial fluid that, if aspirated, is usually aseptic and inflammatory in character.

Up to 25% of inflammatory bowel disease patients can present with a seronegative lower limb large-joint acute monoarthritis often during exacerbation of bowel symptoms. Synovial biopsy from Crohn’s disease patients may show granulomas. In Whipple’s disease, 60% of patients present with large migratory joint monoarthritis or oligoarthritis. Blood tests confirm a high acute phase response with leukopenia, whereas synovial fluid usually shows a high WCC. The diagnosis is based on detection of periodic acid–Schiff-positive material (probably from Tropheryma whippelii) in jejunal macrophages and in foamy cells on synovial biopsy. There are reports of short-lived peripheral joint oligoarthritis in approximately 25% of patients with celiac disease. A seronegative erosive arthritis with inflammatory synovial fluid and lymphocytic infiltrate on
biopsy occurs after jejunoileal surgery for obesity, but arthropathies are not a significant problem after less invasive gastric banding surgery.

Up to 20% of patients presenting with acute knee monoarthritis progress to develop RA. This is an important diagnosis to make with a high acute phase response, rheumatoid factor, and anticyclic citrullinated protein antibodies being useful markers in early disease. The presence of an intermittent inflammatory monoarthritis lasting a few hours or days associated with a temporary rise in acute phase response suggests palindromic rheumatism. MRI and ultrasound in these patients show synovitis, which may be transient. Patients with positive rheumatoid factor and anti-CCP antibodies are at high risk of developing RA within 5 years. Many such patients may meet the American College of Rheumatology (ACR)/EULAR/2010 diagnostic classification criteria for RA and on this basis would merit earlier therapeutic intervention.

In adolescent boys, a monoarthritis may represent the harbinger of spondyloarthropathy. Features of inflammatory back pain, duration of pain for more than 3 months, and presence of HLA-B27 should raise this suspicion and prompt urgent assessment for peripheral and axial inflammatory disease. Most juvenile inflammatory arthritis (JIA) patients with monoarthritis close to the 16-year age cutoff have lower limb arthritis, commonly knee or ankle involvement, and blood and radiologic investigations are similar to those with adult-onset disorders. Uveitis with inflammatory arthritis suggests the ANA-positive variety and requires regular ophthalmologic investigation. Systemic adult-onset JIA is characterized by polyarthritis, a transient salmon pink rash, fever, and serositis. A high serum ferritin and acute phase response is found, and abnormal liver function is common. Synovial fluid shows a high neutrophil leukocytosis.

**MONOARTHRITIS ARISING FROM OTHER JOINT DISEASES**

Osteoarthritis presents as a chronic noninflammatory polyarthropathy with a classical joint distribution. A history of intra-articular fracture or recurrent occupational-related injury (carpet fitter’s knee) or previous joint surgery (arthroscopy or arthrotomy) can lead to a more localized osteoarthritis. In younger patients with hip osteoarthritis, slipped epiphysis, congenital dislocation, or avascular necrosis may antedate disease. Blood investigations will usually be normal, and synovial fluid shows a low white cell count. Ultrasound and MRI help clarify the extent of synovitis and fluid collection, as well as guide aspiration, but plain radiographs usually confirm the diagnosis. In the presence of more inflammatory symptoms it is important to exclude crystal deposition and/or infection, but inflammatory osteoarthritis may develop and require additional therapy.

Neuropathic arthropathy (Charcot’s joint) should be suspected in patients in whom there is a monoarthritis in the distal lower limb with severe osteoarthritis on radiograph in association with demonstrable peripheral neuropathy. Any peripheral neuropathy can lead to this arthropathy, but the fall in syphilis incidence means that diabetes mellitus is the commonest cause in the Western world.

The arthropathy secondary to hemochromatosis results from iron accumulation in joint tissues with osteophytes in the second and third fingers with abnormal liver function and iron saturation tests. PCR of the HFE gene, which regulates iron transport, often demonstrates sequence variations. The arthropathy associated with alkaptonuria affects the spine and large joint degenerative arthritis seen on radiographs. Discoloration of ear cartilage and sclera is diagnostic.

**SYNOVIAL CAUSES OF NONINFLAMMATORY MONOARTICULAR PAIN**

Pigmented villonodular synovitis should be suspected in patients with recurrent monoarticular swelling with blood-stained effusions. MRI is the investigation of choice, but synovial biopsy is essential for pathologic confirmation. Primary or secondary tumors in periarticular tissues including bone can present with single joint pain and are usually diagnosed following routine radiographs or MRI, or both. Lipoma arborescens is being more widely recognized as a result of increasing use of MRI in rheumatology. Patients with this benign tumor, in which synovium is replaced by mature fat cells and villous transformation, usually present with a chronic history of acute knee joint swelling with effusions. MRI changes show villous proliferation of the synovium with characteristic features similar to subcutaneous fat. Occasionally lipoma arborescens occurs together with other arthropathies such as psoriatic arthritis.

Synovial osteochondromatosis presents with symptoms of pain and locking in the large joints, predominantly hips and knees. Synovial fluid is pale in color with few cells. Plain radiographs show calcification in synovial tissues, and synovial histology after biopsy shows the formation of osteocartilaginous bodies in the synovial membrane.

Complex regional pain syndrome type I (reflex sympathetic dystrophy or algodystrophy) usually affects upper limb extremities, often following a triggering event. Pain is diffuse with a burning quality and is worsened after a minor stimulus, and local weakness limits movement. Skin swelling and color change is common with the affected limb being warm. Thermography can identify local increased temperature, bone scintigraphy shows differences between the affected and normal sides, and MRI shows patchy osteoporosis. Synovial fluid contains high protein levels but a low cell count. Hypervascularity and hyperplasia of arteriole walls are seen on synovial biopsy with limited cellular infiltrate or proliferation.

**MONOARTICULAR ARTHRITIS IN SYSTEMIC DISEASES**

Systemic vasculitides including giant cell arteritis, ANCA-associated vasculitis, Henoch-Schönlein purpura, Behçet’s disease, and connective tissue diseases such as Sjögren’s syndrome and SLE may present with arthralgia and occasionally oligoarticular synovitis. Relevant symptoms and signs and appropriate investigations should make the diagnosis. Sudden onset of new monoarticular symptoms in
patients with SLE, particularly those on steroids, should raise the possibility of infection or trauma with tendon damage.\textsuperscript{65} In avascular necrosis MRI is the investigation of choice.\textsuperscript{66}

Sarcoidosis represents approximately 4\% of cases in early synovitis clinics being linked with ankle arthritis, erythema nodosum, and bilateral hilar lymphadenopathy.\textsuperscript{62} Patients are seronegative for rheumatoid factor and have a high acute phase response, but hypercalcemia is not common at presentation. ACE levels may be raised in patients with active widespread sarcoidosis but are often negative in acute joint disease.\textsuperscript{61} Synovial fluid contains lymphocytes, and synovial biopsy shows noncaseating granulomas.\textsuperscript{64} Chronic arthritis is uncommon, but sarcoidosis of bone may present as a dactylitis with bone cysts in the phalanges detected on plain radiograph.\textsuperscript{34}

Amyloid protein deposition (AL or AA) produces an arthropathy with noninflammatory synovial fluid, which is diagnosed on synovial biopsy.\textsuperscript{65} Amyloidosis can be primary (AL) with deposition of light chains in patients with myeloma. Secondary (AA) amyloid is often associated with chronic infectious or inflammatory diseases or rarely of familial etiology. Amyloid secondary to \(\beta_2\) microglobulin deposition was previously seen after long-term dialysis.

In hereditary periodic fever syndromes, patients, often with a Mediterranean connection, may have acute short-lived episodes of monoarthritis with serositis and high fevers. The acute phase response is high, as might be expected from the neutrophil synovial infiltrate. Many genetic associations in the \textit{HFE} gene now exist across the range of autoinflammatory disorders.\textsuperscript{65}

Relapsing polychondritis is an episodic disorder in middle-aged Caucasians that destroys synovial joint cartilage and mimics acute monoarthritis. Ear pinna, nose, larynx, bronchi, and sclera can be affected, as can the aortic valves and aortic root. Synovial fluid is noninflammatory, but the acute phase response is raised, often with a leukocytosis. Biopsy shows an active cellular infiltrate into cartilage, and antibodies can be detected against a variety of cartilage constituents. Repeated episodes lead to local damage and deformity.\textsuperscript{67}

**REGIONAL DIFFERENTIAL DIAGNOSIS OF MUSCULOSKELETAL PAIN**

Injuries to soft tissues present with pain in the region of local joints and hence are often misconstrued as “monoarthritis” by patients. Table 41-4 shows a list of the most frequent causes of soft tissue injury. These conditions are not usually associated with active synovitis in the joint. Blood investigations are usually normal, as are plain radiographic images unless there is local soft tissue swelling, calcification, or underlying joint damage. Ultrasound is the quickest method of imaging to demonstrate local injury to tendons, entheses, or bursae.\textsuperscript{66} MRI is also reliable but costly to perform.\textsuperscript{34}

**MONOARTICULAR PAIN RESULTING FROM PERIARTICULAR DISORDERS**

Local skin infections can mimic acute synovitis, especially in superficial joints. Deep intra-abdominal infections can cause pressure on femoral and sciatic nerves presenting with hip, knee, or sacroiliac joint pain.\textsuperscript{69} The limited signs can lead to delay in diagnosis with consequent increasing systemic toxicity. Investigations will be similar to those of septic arthritis, but ultrasound, CT scanning, MRI, and nuclear medicine WCC scans should localize the exact site of infection. Neuropathic pain can mimic acute arthropathy. Pain from carpal tunnel syndrome can radiate down the arm to the fingers but equally can produce symptoms in the upper arm. Nerve conduction studies are the investigation of choice before injection or surgery, but ultrasound and MRI can be used to identify pressure on the median nerve. Pressure on the lumbar (and cervical) nerve roots can cause pain radiating into the affected limb and can localize into single joints. MRI of the spine is the investigation of choice to examine for disk protrusion or other local pathology. Blood investigations are usually normal.

Patients with chronic pain syndromes have poorly localized widespread pain often with headaches, irritable bowel syndrome, fatigue, and multiple pressure point tenderness. Symptoms can start from localized pain in a single area, which may be close to joints and hence mimic monoarthritis. Multiple trigger points are found on examination, but

<table>
<thead>
<tr>
<th><strong>Table 41-4</strong> Common Soft Tissue Disorders That Arise after Local Injury by Region</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upper Limb</strong></td>
</tr>
<tr>
<td><strong>Shoulder Region</strong></td>
</tr>
<tr>
<td>Rotator cuff tendinitis</td>
</tr>
<tr>
<td>Rotator cuff tears</td>
</tr>
<tr>
<td>Subacromial bursitis</td>
</tr>
<tr>
<td>Frozen shoulder</td>
</tr>
<tr>
<td>Acromioclavicular pain</td>
</tr>
<tr>
<td>Referred pain</td>
</tr>
<tr>
<td><strong>Elbow</strong></td>
</tr>
<tr>
<td>Medial and lateral epicondylitis</td>
</tr>
<tr>
<td>Olecranon bursitis</td>
</tr>
<tr>
<td><strong>Wrist and Hand</strong></td>
</tr>
<tr>
<td>De Quervain’s tenosynovitis</td>
</tr>
<tr>
<td>Trigger finger/thumb</td>
</tr>
<tr>
<td><strong>Lower Limb</strong></td>
</tr>
<tr>
<td><strong>Hip Region</strong></td>
</tr>
<tr>
<td>Iliopsoas bursitis</td>
</tr>
<tr>
<td>Ischiogluteal bursitis</td>
</tr>
<tr>
<td>Trochanteric bursitis</td>
</tr>
<tr>
<td>Adductor tendinitis</td>
</tr>
<tr>
<td>Referred pain</td>
</tr>
<tr>
<td><strong>Knee</strong></td>
</tr>
<tr>
<td>Prepatellar bursitis</td>
</tr>
<tr>
<td>Patellar tendinitis</td>
</tr>
<tr>
<td>Anserine bursitis</td>
</tr>
<tr>
<td><strong>Elbow</strong></td>
</tr>
<tr>
<td>Achilles tendinitis</td>
</tr>
<tr>
<td>Calcaneal and retrocalcaneal bursitis</td>
</tr>
<tr>
<td><strong>Foot</strong></td>
</tr>
<tr>
<td>Plantar fasciitis</td>
</tr>
<tr>
<td>Longitudinal arch problems</td>
</tr>
<tr>
<td>Hallux valgus/rigidus</td>
</tr>
<tr>
<td>Metatarsalgia</td>
</tr>
<tr>
<td>Morton's neuroma</td>
</tr>
</tbody>
</table>
serum, radiologic investigations including MRI, bone scintigraphy, and ultrasound are normal.

Bone pain presenting as acute monoarticular pain is frequently an undiagnosed fracture or resulting from metastatic disease but may be an osteoid osteoma, Paget's disease, or osteomyelitis. Acute bone pain may be indicative of osteosarcoma when associated with a limp in adolescents. Imaging investigations may require technetium bone scans or MRI if standard radiograph is not helpful.

References

8. Cyteval C: Doppler ultrasonography and dynamic magnetic resonance imaging for assessment of synovitis in the hand and wrist of patients with rheumatoid arthritis, Semin Musculoskeletal Radiol 15:66073, 2009.


The references for this chapter can also be found on [www.expertconsult.com](http://www.expertconsult.com).
Articular disorders are the most common cause of disability in the United States. Recent estimates by the Centers for Disease Control and Prevention (CDC) reveal that approximately one in five (49.9 million) adults in the United States have reported doctor-diagnosed arthritis; of these, 21.1 million adults have reported arthritis-attributable activity limitation. A substantial number of these patients manifest polyarticular joint involvement. The presentation of polyarticular joint pain poses a diagnostic challenge to the clinician, given the magnitude of the complaint, urgency to provide relief and preserve joint function, and the uncertainty of the outcome. Although many articular complaints (e.g., tendinitis, bursitis) are self-limiting and require only time and symptomatic management, others may present with or persist as chronic polyarthritis, with the potential for joint damage or disability. Knowledge of the most prevalent or commonly misconstrued causes of polyarthritis can facilitate an accurate diagnosis and appropriate therapy. This chapter will examine the diagnostic approach to polyarthritis, leaving its management to relevant chapters found throughout this textbook (see Chapters 71, 75 through 78, and 81).

**EPIDEMIOLOGY AND SOCIETAL IMPACT**

The prevalence of arthritis is expected to rise as the U.S. population ages; an estimated 60% of the U.S. population older than 65 is affected by chronic joint symptoms. In 2002, the Behavior Risk Factor Surveillance System (BRFSS), a state-based random digit–dialed telephone survey of the U.S. civilian population, found that 69.9 million Americans had chronic joint symptoms defined as pain, aching, or stiffness in or around the joints lasting longer than 3 months. Of these, 25.1 million had not seen a physician, and more than 2 million had activity limitations. The survey also found that those who had never seen a health care provider for their joint complaints were likely to be male, to have lower educational levels, and to have no health insurance. An estimated 25% of patients will be unable to work within 7 years of disease onset. Direct and indirect costs are estimated at 1.2% of the U.S. gross domestic product, or $128 billion. These statistics encompass all forms of arthritis and exert the pervasive societal impact and cost of arthritis and the diagnostic deficit that exists. Thus a facile and accurate approach to patients with joint symptoms is sorely needed.

**DEFINING POLYARTHRITIS**

Polyarthritis is classically defined as pain (with or without signs of inflammation) that affects four or more joints. In the pediatric literature, polyarthritis applies to those with involvement of five or more joints. The distinction between these definitions is moot in that both definitions imply a significant risk of added joint involvement over time. Hence, without an early diagnosis and appropriate management, polyarthritis begets polyarthritis. Patients with chronic involvement of one (monoarthritis) or two or three joints (oligoarthritis) are at low risk of polyarthritis, even with the passage of time.

The differential diagnoses underlying polyarticular complaints can be extensive and intimidating to the inexperienced. Narrowing these to firm a diagnosis is best accomplished in four diagnostic steps:

1. Evaluate for the most prevalent causes—trauma, osteoarthritis, and fibromyalgia.
2. Distinguish between acute and chronic polyarthritis.
3. Delineate articular from periarticular disorders.
4. Elicit clues to increase the likelihood of a diagnosis.

Although many rheumatologists advocate a chronologic (acute vs. chronic), pathophysiologic (inflammatory vs. noninflammatory), or anatomic (oligoarticular vs. polyarticular, periarticular vs. articular) approach to diagnosis, such an approach yields insight into underlying processes without focusing on the most likely conditions, thereby leaving the examiner with a wide range of possibilities. For example, the differential diagnosis of a chronic inflammatory polyarthritis may include rheumatoid arthritis (RA), undifferentiated polyarthritis, psoriatic arthritis, juvenile idiopathic arthritis (JIA), viral arthritis, serum sickness, and tophaceous gout and pseudogout, to name just a few. Regardless of the preferred approach, the diagnosis of polyarthritis relies heavily on an intelligent and accurate history and examination. Hence, key factors such as onset of acuity, course of disease progression, and evidence of inflammation are more diagnostic than a battery of poorly chosen laboratory investigations. Early diagnosis and treatment is the
primary goal, particularly when inflammation is present, in which case earlier treatment can avert considerable disease-related morbidity.

**APPROACH TO DIAGNOSIS**

**Frequent Noninflammatory Polyarticular Scenarios**

It is crucial that the clinician consider the most common causes of polyarticular and widespread musculoskeletal complaints—osteoarthritis (OA) and fibromyalgia (FM). Osteoarthritis affects 12 to 27 million, and fibromyalgia and related disorders (chronic fatigue syndrome, chronic widespread pain) affects 5 to 11 million in the United States alone. Although these two conditions are easily distinguishable from other arthropathies on clinical grounds (without laboratory or radiographic investigation), they often are missed for several reasons. First, it is possible to have more than one condition; RA is frequently accompanied by some degree of degenerative arthritis that may be constitutive, postrheumatic, or secondary to antecedent inflammatory disease. Similarly, patients with systemic lupus erythematosus (SLE) or Sjögren’s syndrome may have secondary FM observed as uncontrolled pain and poor sleep. Second, patients who once had an arthropathy (e.g., an inflammatory arthritis such as psoriatic arthritis) may find that it evolves into another type of arthritis. For example, psoriatic arthritis may cause years of swelling, stiffness, inflammation, and joint damage, only to “burn out” years later. However, the pattern of pain and stiffness may persist owing to secondary OA or FM, even though evidence of inflammatory arthritis is limited or nonexistent. Last, a detailed and skillfully performed musculoskeletal examination is required to discern that the joint pain is articular (e.g., RA) rather than extra-articular (e.g., FM) in origin, or that the swelling is related to synovial (e.g., RA) rather than bony (e.g., OA) hypertrophy, or that widespread pain is accompanied by tender trigger points (e.g., FM) rather than enthesitis (e.g., ankylosing spondylitis).

**Fibromyalgia**

This condition is five times more prevalent than rheumatoid arthritis and may underlie 10% to 20% of all outpatient musculoskeletal visits. It affects women more so than men, usually between the ages of 30 and 60 years, but people of all ages can be affected. Few patients announce widespread pain and tender points at presentation. Instead, most present because of regional joint or muscle pain, profound fatigue, and problems with mentation, memory, or headaches. Only upon further questioning will the examiner discern the wider distribution of painful tender points, joints, muscles, and tendons with an associated recent or long history of poor sleep. FM is a pain amplification disorder, with lowered pain thresholds, that often is accompanied by other “spastic conditions,” including migraine, irritable bowel syndrome, primary dysmenorrhea, atypical chest pain (often mislabeled as costochondritis), temporomandibular joint pain (without dysfunction or damage), and low back or neck pain. Diagnostic testing (laboratory tests, radiographs) is not necessary. The syndrome is established by history and examination and can be confirmed by (1) a positive response to simple analgesics, (2) improved sleep with better sleep hygiene plus sleep aids, (3) improvement attained with a stretching exercise program (e.g., swimming, Pilates, yoga, tai chi, physical therapy), and (4) uncontrolled depression or anxiety.

**Osteoarthritis**

Although nearly 27 million patients in the United States have radiographic evidence of articular degeneration, less than half clinically manifest symptoms of OA. It is important to note that although pain should be accompanied by radiographic OA findings, the extent of radiographic findings and the clinical symptoms reported are often quite discordant. Osteoarthritis should be suspected in patients older than 60 years of age, in those with a history of trauma, and in those whose pain is accentuated by activity. Physical examination may disclose bony hypertrophy (e.g., Heberden’s nodes), joint crepitus, or limited range of motion.

OA and FM can be considered or excluded on the basis of clinical findings. Should these not be applicable, other, more common disorders may be considered according to the chronology of the complaint (acute vs. chronic), the age of the subject, the presence of inflammatory or noninflammatory symptoms, and clues indicating high specificity (Figure 42-1).

**History**

The diagnosis of any polyarticular process relies heavily on the history and physical examination. When faced with a patient who presents with joint complaints, the goal of the encounter is to have answered the questions, “Is joint inflammation present?” and “Will it remit or persist?” “This is important because inflammation that persists may yield significant damage or disability. In addition, it is estimated that more than half of patients who present with acute undifferentiated (often seronegative) polyarthritis will spontaneously remit. Although a 6-week history demarcates acute from chronic arthritis, persistent/chronic polyarthritis is best predicted by articular symptoms lasting 12 weeks or longer. Figure 42-1 presents an algorithm that may assist in formulating a differential diagnosis by providing answers to these questions. The algorithm considers patient demographics, symptom onset, patterns of involvement, disease course, and presence or absence of inflammation.

**Demographics.** Age, sex, and family background may provide clues to the type of arthritis. Gout is more common in men than in women; osteoarthritis affects older patients more often than younger counterparts, and spondyloarthritides have greater familial association than is seen with other inflammatory arthritides. Table 42-1 lists the differential diagnoses of patients who present with inflammatory arthritis according to age and sex.

**Symptom Onset.** Attention should be paid to how symptoms first began and under what circumstances. Eliciting a history of disease onset may be difficult if patients suffered from their illness for years, but identifying the
Disease Course. Noting how the disease progresses may assist in predicting outcome. Patients who have chronic symptoms may describe varying patterns of presentation ranging from an intermittent pattern, where attacks are punctuated by periods of complete remission, to an additive pattern, where symptoms begin with a few joints and progress to involve more joints with time, or a migratory pattern, where certain joints are affected for a time, then the disease remits, only to reappear elsewhere in other joints.

Presence of Inflammation. Evaluation for inflammation may be difficult. Studies have examined features of the history that may provide diagnostic clues to inflammation, such as morning stiffness and response to activity. Clearly, patients with all or some of the cardinal signs of inflammation (redness, warmth, swelling, pain in the morning) may have an inflammatory arthropathy. Morning stiffness (an inflammatory finding) is often confused with the “gel” phenomenon (a degenerative feature). Morning stiffness,
Table 42-1 Differential Diagnoses of Patients Presenting with Inflammatory Arthritis

<table>
<thead>
<tr>
<th>Age, yr</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-30</td>
<td>Spondyloarthropathies*</td>
<td>Infectious arthritis1</td>
</tr>
<tr>
<td></td>
<td>Gout</td>
<td>RA</td>
</tr>
<tr>
<td></td>
<td>Infectious arthritis1</td>
<td>Connective tissue disorders9</td>
</tr>
<tr>
<td></td>
<td>Rheumatic fever</td>
<td>Rheumatic fever</td>
</tr>
<tr>
<td></td>
<td>Arthritis associated with oral contraceptives</td>
<td>Sarcoïdosis</td>
</tr>
<tr>
<td>30-60</td>
<td>Gout</td>
<td>Infectious arthritis</td>
</tr>
<tr>
<td></td>
<td>Spondyloarthropathy*</td>
<td>Arthritis of thyroid</td>
</tr>
<tr>
<td></td>
<td>Infectious arthritis</td>
<td>disease</td>
</tr>
<tr>
<td></td>
<td>Vasculitis3</td>
<td>Vasculitis3</td>
</tr>
<tr>
<td></td>
<td>Hemochromatosis</td>
<td>Sarcoïdosis</td>
</tr>
<tr>
<td></td>
<td>Rheumatoid arthritis</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>&gt;60</td>
<td>Bursitis</td>
<td>Arthritis associated with oral contraceptives</td>
</tr>
<tr>
<td></td>
<td>Gout</td>
<td>Infectious arthritis</td>
</tr>
<tr>
<td></td>
<td>RA</td>
<td>RA</td>
</tr>
<tr>
<td></td>
<td>Pseudogout</td>
<td>Pseudogout</td>
</tr>
<tr>
<td></td>
<td>Vasculitis3</td>
<td>Vasculitis3</td>
</tr>
<tr>
<td></td>
<td>PMR</td>
<td>PMR</td>
</tr>
<tr>
<td></td>
<td>RS3PE</td>
<td>Infectious arthritis</td>
</tr>
<tr>
<td></td>
<td>Arthritis of thyroid</td>
<td>Erosive osteoarthritis</td>
</tr>
<tr>
<td></td>
<td>Arthritis of thyroid</td>
<td>Arthritis of thyroid</td>
</tr>
<tr>
<td></td>
<td>Amyloidosis</td>
<td>Amyloidosis</td>
</tr>
<tr>
<td></td>
<td>Lymphoma</td>
<td>Lymphoma</td>
</tr>
<tr>
<td></td>
<td>Hypertrophic osteoarthritis</td>
<td>Lymphoma</td>
</tr>
</tbody>
</table>

*Spondyloarthropathies include ankylosing spondylitis, psoriatic arthritis, reactive arthritis, and enteropathic arthritis.

1Infectious arthritis includes gonococcal arthritis, Lyme arthritis, viral arthritis, and bacterial arthritis.

2Connective tissue disorders include systemic lupus erythematosus, scleroderma, Behçet’s disease, mixed connective tissue disease, and polymyositis/dermatomyositis.

3Vasculitis includes polyarteritis nodosa, granulomatosis with polyangiitis (formerly Wegener’s granulomatosis), hypersensitivity vasculitis, and temporal arteritis.

4PMR, polymyalgia rheumatica; RA, rheumatoid arthritis; RS3PE, relapsing symmetric seronegative synovitis with pitting edema.


defined as the time to maximal improvement after an extended period of inactivity (an overnight rest), typically improves with movement.5 In patients with inflammatory arthritis, the duration of morning stiffness is typically greater than 60 minutes. The gel phenomenon refers to short-lived stiffness that is brought on by short periods of rest and is common to patients with degenerative arthritis. Patients with inflammatory arthritis may present with pain without swelling, swelling without pain, or difficulty performing their activities of daily living without pain or swelling. It is important to distinguish whether inflammation is truly present as the cause of symptoms. Both arthralgia (joint pain) and arthritis (joint swelling from synovial fluid or tissue) may be present in new-onset polyarthritis; hence the joint examination is an essential component in establishing the diagnosis.

Physical Examination

Arthralgia vs. Arthritis. Characteristics that distinguish synovitis include warmth, erythema, tenderness to palpation, and synovial effusion. Any or all of these findings may accompany arthralgia. Range of motion, muscle strength, and function may be limited around the inflamed joint. In an effort to reduce joint volume and pain, the patient often will involuntarily hold the joint in a position of partial flexion. Hence, joint contractures may indicate an underlying inflammatory process (present or past). In RA, any diarthrodial joint can be affected, but the pattern of involvement typically involves the MCPs, PIPs, wrist, metatarsophalangeal joints (MTPs), and ankle joints.6 This pattern of involvement should be distinguished from osteoarthritis (DIPs, PIPs, carpometacarpal [CMC]1, knee, hip, spine), psoriatic arthritis (DIPs, PIPs, wrist, toes), and pseudogout (knee, wrist, MTPs). The accuracy of identifying synovitis depends on the skills and experience of the physician. The reproducibility of the physical examination in identifying synovitis has been a subject of concern raised in several studies.7-11 One study looked at whether rheumatologists can agree on the diagnosis of inflammatory arthritis; 24 patients were assessed by 6 rheumatologists—complete agreement was found in only 70% of cases; disagreement in the remaining 30% of cases ranged from 1 to 4 rheumatologists recognizing synovitis when others did not.12

This is further illustrated by imaging studies. When the physical examination is compared with imaging, the joint examination has a reported sensitivity of 43%, specificity of 89%, and accuracy of 71% with magnetic resonance imaging (MRI) as the reference.13 Given the difficulty for many practitioners of discerning synovitis, Emery and colleagues have advocated the value of the MCP or MTP “squeeze test” (Figure 42-2). Their study found that pain upon transverse compression of the MCP or MTP joints (a positive squeeze test) has better specificity in identifying persistent arthritis than in finding arthritis in three or more joints, or symmetric arthritis.13 Currently, the joint examination is the gold standard for detection of synovitis; however, MRI and ultrasound may provide additional diagnostic evidence of synovitis in patients with chronic synovitis that is not evident on joint examination.

Articular vs. Periarticular. Equally important to identifying inflammation is distinguishing articular from periarticular involvement. In patients with articular pain, the joint capsule is diffusely involved; thus, pain is often deep and is associated with a global decreased range of active and passive motion in all planes. Those who have periarticular abnormalities may have point tenderness in the surrounding soft tissue, and pain occurs only with active range of motion in a few planes. The differential diagnoses in patients with periarticular pain may include FM, fracture, bursitis, tendinitis, enthesitis, carpal tunnel syndrome, sickle cell crisis, polymyalgia rheumatica, neuropathy, and Raynaud’s phenomenon.

Extra-articular Manifestations. Although joint symptoms are the primary focus of patients who present with pain, extra-articular manifestations may provide clues to the correct diagnosis. Extra-articular manifestations of RA (e.g., nodules, keratoconjunctivitis sicca) are seldom present early in the disease. This sharply contrasts with other forms of polyarthritis wherein extra-articular manifestations are prominent early and may precede the onset of synovitis. This is particularly so with SLE (malar rash, serositis),
reactive arthritis (urethritis, conjunctivitis), psoriatic arthritis (psoriasis, nail pitting), and sarcoidosis (lung, fever, uveitis, parotitis). Nodules in the seronegative patient are more likely to be tophi from gout than nodules from RA, because the latter are seen only in those with high-titer rheumatoid factor (RF) or cyclic citrullinated protein (CCP) antibodies.

LABORATORY TESTS AND RADIOLOGIC STUDIES

The diagnosis of polyarthritis is made clinically with the history and physical examination; nevertheless, laboratory tests and radiographic imaging may aid in diagnosis, guide treatment, and predict outcome. Laboratory investigation of polyarthritis is indicated with chronicity (symptoms longer than 6 weeks), failure to respond to initial therapy, and the presence of systemic (e.g., fever, rash) or neurologic symptoms.

Laboratory Tests

Acute Phase Reactants

Elevations in acute phase reactants such as erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) provide a surrogate measure of inflammation; both ESR and CRP have been correlated with poor prognosis and worse radiologic outcomes. However, they do not discriminate well between RA and other inflammatory and noninflammatory arthritides. Not all patients with inflammatory polyarthritis will have elevated ESR or CRP at presentation; some may have elevations in other acute phase reactants such as ferritin, haptoglobin, ceruloplasmin, and complement levels. Other markers of inflammation may include an anemia of chronic disease and elevated platelets and white cell count.

Serologies

Autoantibodies have been used to guide the classification of various autoimmune diseases. Their performance has been studied for diagnostic and prognostic value. Testing for serologic markers may add significant diagnostic and prognostic value to the evaluation of inflammatory polyarthritis; however, issues of cost, availability, and standardization may not allow for routine use. One of the most commonly ordered tests is the rheumatoid factor (RF). Serum RF is an autoantibody (typically immunoglobulin [Ig]M) that binds to the Fc component of IgG and may play a role in acute inflammatory arthritis. The RF is not required for the initiation of inflammation because patients with agammaglobulinemia have been described to have rheumatoid-like synovitis; indeed, approximately 20% of patients meeting American College of Rheumatology (ACR) classification criteria for RA are seronegative. The rheumatoid factor performs poorly in diagnosing early RA as reported in several studies; in patients with symptoms of less than 1 year, the sensitivity of serum RF ranges from 17% to 59%, and numbers are even lower in patients with symptoms lasting less than 12 weeks. The specificity of the rheumatoid factor for RA is also poor in that it may be present in other diseases as well, including viral hepatitis, systemic lupus erythematosus, cryoglobulinemia, Sjögren’s syndrome, paraproteinemias, bacterial endocarditis, mycobacterial diseases, syphilis, leprosy, chronic interstitial lung diseases, parasitic infections, and lymphoproliferative malignancies. These conditions may also present with synovitis and may be mistaken for RA. Screening of the population with the RF test generally yields a positive rate of 4% to 5%; hence random use of RF would likely generate more
false-positives than true-positives. The presence of the RF has been found to be predictive of persistent disease and progression with radiologic damage in patients with inflammatory arthritis.7,23 Visser and associates noted that a positive IgM RF confers a relative risk of 2.99 for erosive disease; Tunn and Bacon found that RA latex positivity combined with an ESR >50 mm/hr carries a relative risk of 4.33 for persistent synovitis, with a specificity of 94%.14,29 Thus, despite the poor performance of the RF in diagnosing RA, it may be useful in predicting aggressive, tenacious disease.

Other serologic markers that have been evaluated for use in early diagnosis of RA include the anticitrullinated protein antibodies (ACPAs or CCP Ab), which are nearly as sensitive as RF but are far more specific for RA. If the CCP Ab test is combined with the RF, one can expect a sensitivity of 58% with a specificity of 100%; positive and negative predictive values are 100% and 88%, respectively.10

The clinical utility of antinuclear antibodies (ANAs) has also been evaluated in patients with polyarthritis. The ANA is not uniquely linked to lupus but may be found in a wide variety of conditions and in normal people, often at low titers; hence the practice of ordering batteries of autoantibodies to make a laboratory diagnosis of arthritis should be avoided. Many of these antinuclear antibodies are found in up to 15% of the normal healthy or elderly population and have low specificity and poor predictive value.7 The most common cause for rheumatologic consultation is to explain a low-titer ANA result that has no associated features of lupus. ANAs are a class of autoantibodies that react with various components of the cell nucleus; they may be found in a variety of autoimmune conditions, as well as in healthy individuals. The specificity of this test is poor because ANAs may be found in patients with systemic lupus erythematosus, systemic sclerosis, chronic liver disease, chronic interstitial lung disease, drug-induced lupus, or Hashimoto’s thyroiditis. Their prevalence in patients with inflammatory polyarthritis has been well studied. The presence of ANAs in patients with early inflammatory arthritis ranges from 40% to 69%.19,21 Titters are generally lower in patients with RA than in those with SLE or systemic sclerosis (SSc). No correlation has been found between ANA positivity and joint inflammation, severity of joint destruction, or persistence of disease. In juvenile idiopathic arthritis (JIA), ANA positivity is modestly correlated with risk of uveitis in the pauciarticular subset of JIA (see Chapter 106).

### Genetic Markers

The study of genes has provided insight into many diseases. Associations have been observed between regions of the major histocompatibility complex (MHC) on chromosome 6 and rheumatic diseases. One of the most striking associations is the one between HLA-B27 and the seronegative spondyloarthritides. The gene can be found in approximately 8% of the general population, but the prevalence of HLA-B27 positivity is 90% to 95% in white patients with ankylosing spondylitis (AS).31 Table 42-2 lists the frequency of HLA-B27 in the different seronegative spondyloarthritides. Given the low predictive value of the test (0.95% to 9.5%), HLA-B27 should be ordered only when suspicion for spondyloarthritis is moderate.

### Table 42-2 Frequency of HLA-B27 Gene in Patients with Seronegative Spondyloarthritis

<table>
<thead>
<tr>
<th>Disease</th>
<th>HLA-B27 Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ankylosing spondylitis</td>
<td>95</td>
</tr>
<tr>
<td>Undifferentiated spondyloarthropathy</td>
<td>70</td>
</tr>
<tr>
<td>Reactive arthritis</td>
<td>40-80</td>
</tr>
<tr>
<td>Irritable bowel disease–associated</td>
<td>35-75</td>
</tr>
<tr>
<td>Spondyloarthritis</td>
<td></td>
</tr>
<tr>
<td>Psoriatic spondyloarthritis</td>
<td>40-50</td>
</tr>
<tr>
<td>Psoriatic peripheral disease</td>
<td>25</td>
</tr>
<tr>
<td>Iritis</td>
<td>50</td>
</tr>
<tr>
<td>Cardiac conduction defects</td>
<td>80</td>
</tr>
</tbody>
</table>


Another gene that has been studied and may aid with diagnosis is the HFE gene found in patients with hereditary hemochromatosis (HHC). HHC is an autosomal recessive disorder of iron metabolism; joint pain is the most common complaint of patients who suffer from the disease.32 HFE gene mutations are fairly common, and heterozygosity occurs in 1 of 10 white persons; the frequency of homozygotes is around 4.4 per 1000.32 Screening of the general population for HHC is not recommended, but some have advocated targeted screening in patients with suggestive symptoms and laboratory results (e.g., abnormal serum ferritin, transferrin saturation, serum aspartate, or alanine aminotransferase).

Currently, much research is being conducted to identify genetic susceptibility foci that may aid in diagnosis, predict prognosis, and guide therapy in diseases such as SLE, rheumatoid arthritis, gout, osteoarthritis, sarcoidosis, and other rheumatic diseases.

### Synovial Fluid Analysis

In patients with oligoarthritis or in those who present with joint effusions, arthrocentesis may be useful in diagnosing patients and relieving symptoms. Synovial fluid from an inflamed joint is typically yellow and turbulent from inflammatory cells. White cell counts are typically greater than 10,000 cells/mm (range, 5000 to 50,000 cells/mm), with a predominance in neutrophils.33 Evaluation for infection or crystalline arthritis is an important part of the initial assessment of inflammatory arthritis. Once synovial fluid has been aspirated, prompt evaluation under polarized microscopy will maximize the yield for identifying crystals. Joint aspiration is also therapeutic in patients, reducing the pressure imposed on the joint capsule and prolonging the benefits of intra-articular steroids. One study found that aspiration of synovial fluid before intra-articular corticosteroid injections reduces the risk for relapse (23% in the arthrocentesis group vs. 47% in the nonarthrocentesis group; p = 0.001) and extends the duration of steroid effectiveness.34 In addition, joint injections may serve as a prognostic indicator. Green and colleagues found that patients whose synovitis persisted 2 weeks after intra-articular corticosteroid placement were more likely to have persistent synovitis after 1 year.35
Imaging

Radiographic changes may occur early in any polyarticular disease, but detecting these changes by conventional x-rays may be difficult. Often such images are normal or show only evidence of soft tissue swelling. Indications for radiography include (1) history of trauma or injury (to exclude fracture), (2) persistence of joint pain and swelling longer than 6 weeks, (3) suspicion of septic or gouty arthritis, and (4) as a baseline evaluation for a newly diagnosed polyarticular condition. Characteristic findings on radiographs of inflammatory arthritis may include soft tissue swelling, chondrocalcinosis, joint effusion, juxta-articular osteopenia, symmetric loss of articular cartilage with joint space narrowing, and bony erosions. Identification of these abnormalities is important in that radiographic damage has been shown to correlate with loss of productivity and increased functional disability in RA. Erosions are important markers of progressive damage. The advantages of x-rays have been proven as they have been shown to help establish prognosis and assess joint damage longitudinally in patients with RA or inflammatory arthritis. In a prospective study evaluating 113 patients for the development of radiographic changes in the first 5 years of rheumatoid arthritis, 26% of patients were found to have signs of damage at the initial visit (mean duration of symptoms, 9.4 months); the rate of radiographic progression was most rapid during the first 2 years but then decreased significantly in the third to the fourth year (p < 0.01). Only 11% of patients had no erosions by the end of the study. Erosions often start in the feet and in 14% to 18% remain only in the feet.15 Thus ordering x-rays of the feet is an important part of the initial evaluation, even if patients do not have symptoms in their lower extremities. Radiographic progression is the single most important outcome when assessed by clinical trials because it indicates persistence of inflammation.15,16,27

Other imaging modalities currently being investigated in early arthritis include magnetic resonance imaging (MRI) and ultrasound (US); both have proved sensitive for the detection of synovitis when clinical examination and conventional radiographs have failed.23-42 Advantages of these devices include ability to detect subtle synovitis and soft tissue abnormalities as tendon rupture or tenosynovitis and to permit more accurate placement of the needle in diagnostic arthrocentesis and therapeutic injections. Before these tests are ordered, their potential benefits should be weighed against their limitations including long examination times, availability of equipment, costs, and skills of the observer in interpreting changes.

The eventual diagnosis of polyarthritis will depend on key historic features, a detailed joint examination, and, occasionally, laboratory findings. Most cases can be diagnosed on the basis of clinical grounds, after consideration of the chronology of symptoms (onset, evolution, pattern of joint involvement), background history, and physical findings. Table 42-3 and Figure 42-3 provide helpful tools in the evaluation of patients who present with inflammatory polyarthritis. Reliance on laboratory testing to establish a diagnosis is ill advised because such tests are poorly predictive when used indiscriminately or as part of broad batteries of “arthritis screening” tests. The strength of laboratory and serologic testing is greatly enhanced when they are used to confirm a reasonably strong clinical suspicion garnered by evaluating the history and examination.

### UNIQUE SITUATIONS

**Undifferentiated Inflammatory Arthritis**

Because of the inadequacies of the 1987 ACR RA Classification Criteria for those with early disease, patients with inflammatory polyarthritis are often given the label undifferentiated polyarthritis.23,25 These older criteria reflected the presence of features that often require time and severity (e.g., symptoms lasting longer than 6 weeks, erosions, nodules, RF positivity), thereby limiting their diagnostic value in new-onset polyarthritis. The newer 2010 ACR/European League Against Rheumatism (EULAR)
RA Classification Criteria were designed with the goal of identifying patients with undifferentiated inflammatory synovitis who are at risk for persistent and/or erosive disease, and who would benefit most from initiation of early effective intervention (see Chapter 70). These new criteria improve the ability to classify more patients with RA; however, the prevalence of certain musculoskeletal disorders should still be considered in evaluating patients with polyarthritis. High-prevalence diseases (e.g., RA, reactive arthritis, gout, pseudogout, spondyloarthropathy, polymyalgia rheumatica) should be considered before those that rarely present as polyarthritis (e.g., amyloidosis, sarcoidosis, lymphoma, RS3PE [remitting seronegative symmetric synovitis with pitting edema, vasculitis]). Only a minority of patients with early polyarthritis will have RA.

In a prospective study of 233 patients followed for a year (median symptom duration, 31 days), 22% of patients were diagnosed with RA, 32% had unclassified arthritis, and 46% had other diagnoses such as crystalline arthritis, sarcoidosis, reactive arthritis, and psoriatic arthritis. When the study was extended to include a total of 566 patients followed over 2 years (median symptom duration, 2.7 months), results did not change: 30% were diagnosed with RA and 26% with undifferentiated polyarthritis (UPA), and 46% were given another diagnosis. Other early arthritis clinics followed their patients with early undifferentiated polyarthritis and found that up to 65% had remission of symptoms at 1 to 2 years. The prognosis and chance for remission appeared better in this undifferentiated group, which often is more easily managed. Nevertheless, patients with undifferentiated polyarthritis may develop persistent and aggressive disease. Such patients need to be serially evaluated and treatments tailored to complement the aggressiveness of their arthritis. With chronicity comes the need to aggressively treat the patient for RA or undifferentiated polyarthritis—even when serologies are negative.

The Hospitalized Patient

The CDC, in a report derived from hospitalization data of the 1997 National Hospital Discharge Survey, stated that an estimated 744,000 (3%) hospitalizations occurred with arthritis as the principal diagnosis. The most common diagnosis was related to osteoarthritis, followed by soft tissue disorders, spondylosis/spondylitis, and other rheumatic conditions such as infection. Gout and myalgia, including FM, each account for about 1% of hospitalizations. How arthritis played a role in hospitalization was not specified, but the authors were clear to point out that hospitalization courses were complicated by the high rates of pain and limitation associated with arthritis. A U.K. study published in 1991 noted that the top hospitalization diagnoses seen by rheumatology consult services in descending order were inflammatory polyarthritis, degenerative arthritis, seronegative spondyloarthritis, soft tissue rheumatism,
skeletal abnormality, and gout. In contrast, data from a U.S. academic center published in 2001 noted that the five top reasons for rheumatology consultations at a university hospital were vasculitis, lupus, gout, rheumatoid arthritis, and soft tissue rheumatic conditions. Patient demographics played a role in the diagnosis. Data from veterans’ hospital consultations revealed that the most common diagnoses were crystalline arthritis and noninflammatory regional musculoskeletal conditions. Hence arthritis and polyarthritis often occur in the inpatient setting and frequently require rheumatologic consultation. The diagnostic challenge is to discern whether new-onset polyarthritis is a consequence of the primary diagnosis, comorbidity, changing immunocompetency or metabolic state, or current drug therapy. Osteoarthritis, gout, and fibromyalgia account for a majority of such cases.

**Infection and Polyarthritis**

Infection as a cause of polyarthritis should always be considered in patients who present acutely with symptoms. Bacteria, viruses, and atypical microorganisms may cause polyarthritis directly as a pathogen or indirectly through an immune-mediated response (see Chapters 109 through 115). Bacteria that have been associated with polyarthritis include staphylococci, streptococci, enterococci, Neisseria gonorrhoeae, Borrelia burgdorferi, and gram-negative bacilli. Certain viruses can also cause polyarthritis; these include parvovirus B19, mumps, rubella, hepatitis B and C viruses, cytomegalovirus, Epstein-Barr virus, HIV, and certain enteroviruses. In addition, arboviruses (e.g., insect-transmitted viruses) have been associated with polyarthritis. Severe cases of debilitating polyarthritis have been associated with the Chikungunya virus, for which human epidemics have been reported in Africa, Asia, and certain parts of Europe. A detailed travel history may help guide specific testing for diseases endemic to the region. Although vigilance for infection is important in instituting appropriate therapy, this fact should be balanced by thoughtful clinical assessment based on history and examination before extensive testing is ordered.

**Polyarthritis, Rash, and Fever**

The triad of fever, arthritis, and rash may pose a challenge to the clinician. Differential diagnoses to consider include autoimmune (e.g., SLE, dermatomyositis, vasculitis), infectious (e.g., disseminated gonococcal infection), reactive, or inflammatory processes (e.g., serum sickness reaction, rheumatic fever, adult-onset Still’s disease), and cryopyrin-related diseases (see specific chapters related to these individual diseases). Careful history and detailed physical examination may provide clues to the diagnosis.

**Drug-Induced Syndromes**

Despite the potential benefits of medications, certain adverse effects manifest as musculoskeletal complaints that mimic a primary rheumatologic disease. Best characterized are the agents that give rise to drug-induced lupus. Several drugs have been linked to lupus-like features, including hydralazine, procainamide, isoniazid, propylthiouracil, sulfonamides, quinidine, tumor necrosis factor inhibitors, and minocycline. Drug-induced lupus may present with myalgias, arthralgias, arthritis, or systemic complaints of fever, skin rash, pleuropulmonary disease, or cytopenias; a positive ANA test is required for diagnosis. Renal and neurologic features and double-stranded DNA (dsDNA) antibodies are usually absent. Seropositivity alone is not sufficient to make this diagnosis because only a fraction of patients who become seropositive (e.g., ANA positive) on an offending drug will develop lupus-like clinical features. Symptoms can appear weeks to months after exposure to the offending drug. The diagnosis is established by findings of lupus-specific features and ANA seropositivity while the patient is receiving an offending agent, and remission upon withdrawal of the agent.

**Malignancy-Related Polyarthritis**

Rheumatic symptoms associated with cancer may be difficult to distinguish from true rheumatologic disease. Symptoms typically are not related to direct tumor invasion or metastatic disease, but instead result from a paraneoplastic process. Patients may present with hypertrophic osteoarthropathy, carcinomatous polyarthritis, dermatomyositis/polymyositis, polymyalgia, or vasculitis. Rheumatic manifestations suggestive of an occult malignancy may include rapid onset of an unusual inflammatory arthritis, clubbing, diffuse bone pain typically in patients older than 50 years of age, chronic unexplained vasculitis, refractory fasciitis, Raynaud’s syndrome unresponsive to vasodilator therapy, rapidly progressive digital gangrene, or Lambert-Eaton myasthenic syndrome. Rheumatic symptoms may coincide with or antedate the diagnosis of malignancy, with a typical course coinciding with that of the primary tumor. Prompt recognition is important so prompt treatment for the malignancy can be instituted, but extensive searching for an occult malignancy in most rheumatic syndromes is not advised unless accompanied by suggestive findings.

**Pediatric Polyarthritis**

Young patients who present with musculoskeletal complaints may be dismissed as having growing pains or symptoms attributed to a sports injury. Rheumatic diseases of childhood can be challenging to diagnose and may include juvenile idiopathic arthritis, ankylosing spondylitis, psoriatic arthritis, arthritis of inflammatory bowel disease, reactive arthritis, SLE, systemic sclerosis, vasculitis, Kawasaki disease, Henoch-Schönlein purpura, Lyme disease, septic arthritis, acute rheumatic fever, infective endocarditis, and human parvovirus B19 infection. Tracking pediatric arthritis has been difficult. In 2007, the CDC and the ACR collaborated to publish a report on the prevalence of significant pediatric arthritis and other rheumatologic conditions (SPARC) in the United States. The calculated prevalence rate of SPARC was 403 per 100,000 (95% confidence interval [CI], 257 to 548 per 100,000); the top five diagnoses were listed as synovitis and tenosynovitis, myalgia and myositis (includes fibromyalgia), osteoarthritis and allied disorders, diffuse diseases of connective tissue, and rheumatoid arthritis and other inflammatory arthropathies.
Polyarthritis in the Elderly

Evaluating geriatric patients may be challenging because symptoms are often accompanied by nonspecific general complaints related to normal aging, medication side effects, or comorbid illnesses. Gerontorheumatologic diseases that should not be overlooked include late-onset rheumatoid arthritis, polymyalgia rheumatica, remitting seronegative symmetric synovitis with pitting edema (RS3PE) syndrome, giant cell arteritis, and paraneoplastic rheumatic syndrome. Often, patients present with polyarticular chondrocalcinosis of uncertain significance. Nonetheless, pseudogout, gout, and drug-induced disorders are common in the elderly. Laboratory testing, including synovial fluid analysis, may be helpful in diagnosis.

OUTCOME: A WINDOW OF OPPORTUNITY

Treatment of polyarthritis is tailored according to the underlying disease process; once persistent inflammatory arthritis is recognized, timely intervention is essential to halt the inflammatory process to prevent or postpone permanent joint damage. Several studies have found that a substantial number of patients who have unclassified inflammatory polyarthritis go into remission after 1 to 2 years.29-42 The difficulty lies in identifying those who will have persistent, destructive disease. Hence, early referral to a rheumatologist for prompt evaluation and initiation of appropriate therapy is essential for limiting damage. If treatment is delayed in those with poor prognostic indicators (e.g., positive RF, ACPA, evidence of radiographic damage, disease duration >6 weeks), the opportunity to improve outcome may be missed. Many studies have shown that early institution of aggressive disease-modifying antirheumatic drugs (DMARDs) in early inflammatory arthritis improved disease activity scores, pain, function, and radiologic scores, and allowed a chance at remission.35-38 Yet the potential side effects of these drugs must be reviewed when the decision is made to prescribe them to patients; no therapeutic algorithms are available for patients with unclassified inflammatory arthritis. The approach of managing inflammation and controlling symptoms is essential for these patients. Standard protocol should dictate classifying patients as having slowly progressive or aggressive destructive disease. Stratification should be based on prognostic indicators (Table 42-4) such as the presence of RF or CCP seropositivity, extended duration of disease, and radiographic evidence of joint destruction. In addition, for all patients, nonpharmacologic therapies should be initiated early; treatments include patient education, physical and occupational therapy with emphasis on joint protection, and maintenance of joint function.

CONCLUSION

Evaluating and diagnosing polyarthritis is challenging. This chapter establishes the importance of prompt evaluation, early referral, and expedient initiation of therapy. The clinician should consider common forms of polyarthritis first, and should assess patient age and sex, chronology, background history, presence of extra-articular manifestations, and pattern of joint involvement when formulating a differential diagnosis. Only after careful consideration of these factors should clinicians proceed with additional investigations to confirm their suspicions. Without an established diagnosis, the patient’s complaints should be managed in a symptom-driven fashion with frequent evaluations. As the disease evolves, additional investigative measures may be required.

Table 42-4  Factors Predicting Persistent and Erosive Arthritis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio for Persistent Arthritis</th>
<th>Odds Ratio for Erosions Given Persistence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥6 weeks, &lt;6 months, ≥6 months</td>
<td>2.49</td>
<td>0.96</td>
</tr>
<tr>
<td>Morning stiffness ≥1 hour</td>
<td>1.96</td>
<td>1.96</td>
</tr>
<tr>
<td>Arthritis in ≥3 joint groups</td>
<td>1.73</td>
<td>1.73</td>
</tr>
<tr>
<td>Pain with bilateral MTP squeeze</td>
<td>1.65</td>
<td>3.78</td>
</tr>
<tr>
<td>IgM RF ≥5 IU</td>
<td>2.99</td>
<td>2.99</td>
</tr>
<tr>
<td>Anti-CCP ≥92 IU</td>
<td>4.58</td>
<td>4.58</td>
</tr>
<tr>
<td>Erosions on hand/foot radiographs</td>
<td>2.72</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

anti-CCP, Anti-cyclic citrullinated protein antibodies; Ig, immunoglobulin; MTP, metatarsophalangeal; RF, rheumatoid factor.


References


The references for this chapter can also be found on www.expertconsult.com.
The Skin and Rheumatic Diseases
LELA A. LEE • VICTORIA P. WERTH

KEY POINTS
Many systemic rheumatologic conditions present with skin findings.
The differential diagnosis for skin findings may require a skin biopsy for clinical-pathologic correlation.
Systemic steroids should be avoided if possible in psoriasis because flaring may occur with tapering of systemic glucocorticoids.
New biologic therapies have substantially improved care of patients with severe and resistant psoriasis.
The neutrophilic dermatoses comprise a group of inflammatory diseases including pyoderma gangrenosum and Sweet’s syndrome, which can be associated with autoimmune diseases.
Patients with lupus erythematosus can have a wide variety of lupus-specific and lupus-nonspecific skin lesions.
Lupus-nonspecific skin lesions are more frequently seen in patients with systemic lupus erythematosus.
Patients with dermatomyositis frequently have pathognomonic skin lesions that may be seen in the absence of muscle disease.
Antimalariais may be of benefit for patients with morphea.

DIAGNOSIS OF SKIN LESIONS ASSOCIATED WITH RHEUMATIC DISEASES

The skin is a highly visible organ that is frequently affected in rheumatic diseases, and the presence of skin lesions may be helpful diagnostically. Certain caveats are worth noting before discussing the specifics. First, a major pitfall for the nondermatologist in evaluating skin lesions is incomplete knowledge of the entities in the differential diagnosis. For example, malar erythema occurs frequently in patients with systemic lupus erythematosus (SLE), but the differential diagnosis for malar erythema is rather extensive and includes conditions that are much more prevalent than lupus (e.g., rosacea), as well as conditions that are far less prevalent (e.g., Rothmund-Thomson syndrome). Patients frequently have more than one skin condition, which often makes diagnosis more challenging.

Another caveat for the nondermatologist concerns skin biopsy. Being able to determine when a skin biopsy is likely to be diagnostically useful in many cases requires a great deal of specialized knowledge, as does the interpretation of pathology reports. It is often the case with inflammatory skin conditions that the microscopic findings are actually less diagnostically specific than is the clinical examination. Unfortunately, it is common for a pathology report to list a diagnosis without providing context regarding how definitive were the findings. For example, a skin biopsy report may have psoriasis listed as the final diagnosis, but, depending on the specific case, the unstated additional possibilities may include nummular dermatitis, atopic dermatitis, seborrheic dermatitis, lichen simplex chronicus, dermatophytosis, or drug eruption. Particularly with regard to inflammatory skin conditions, having a working knowledge of both dermatopathology and dermatology and placing the histologic findings in the context of the clinical presentation may be necessary to arrive at the correct diagnosis.

The above considerations notwithstanding, it is clearly useful for the physician caring for patients with rheumatic diseases to be well versed in their cutaneous manifestations. In this chapter, we provide an overview of these manifestations, as well as perspective on diagnosis and differential diagnosis. Therapy is discussed briefly in conditions where treatment may be specifically directed toward the skin lesions. Etiology and pathogenesis of these diseases are covered elsewhere in this text.

PSORIASIS

Psoriasis is one of the most common inflammatory skin diseases, affecting about 2% of the general population. There is a wide range of severity of skin lesions, from a few relatively asymptomatic plaques to extensive, disabling disease. The onset may be at any time during life. Once present, it may exhibit exacerbations and remissions, but it does not tend to resolve permanently. In general, onset in childhood portends more severe disease.

Skin lesions of psoriasis characteristically are sharply demarcated plaques with silvery scale and underlying erythema, although there may be a paucity of scale if the lesions have been partially treated or if they occur in intertriginous areas. When the scale is removed, pinpoint bleeding may be observed (Auspitz sign). Lesions may occur in areas of trauma (Koebner phenomenon) such as in surgical scars. In some cases lesions contain small pustules.

General phenotypes of psoriasis are chronic plaque, guttate, localized pustular, generalized pustular, and erythroderma. Chronic plaque and guttate psoriasis are the most common, and generalized pustular psoriasis and erythroderma typically the most disabling and even life threatening. Chronic plaque psoriasis lesions are often relatively large in diameter and occur preferentially on elbows, knees, scalp, genitalia, lower back, and the gluteal cleft, although they may occur at many other locations. It is quite common
EVALUATION OF GENERALIZED AND LOCALIZED SYMPTOMS

Figure 43-1  Guttate psoriasis resembles “drops” of discrete scaly papules with erythema, often on the trunk. (Courtesy Dr. Nicole Rogers, Tulane University Department of Dermatology, New Orleans.)

for only one area of skin such as the scalp to be affected. Guttate lesions are relatively small in diameter and usually quite numerous, distributed preferentially on the trunk and proximal extremities (Figure 43-1). Guttate psoriasis occurs relatively commonly in children and young adults, often manifesting a few weeks after a streptococcal infection.

Nail changes are common, occurring in about half of patients, and are often mistaken for fungal infection. Specific changes include pitting, onycholysis (“oil spots”), dystrophy of nails, and loss of the nail plate. These changes are not specific for psoriasis. Notably, pitting may occur as a result of trauma, and the finding of a few pits in the nails may not be helpful diagnostically. Nail changes are more frequent in patients with arthritis of the distal interphalangeal joints.

Arthritis occurs more often in patients with severe cutaneous disease, but cutaneous disease need not be present at all. Remissions and exacerbations of arthritis do not correlate well with remissions and exacerbations of skin disease. The presence of psoriatic skin lesions may be helpful in establishing a diagnosis of psoriatic arthritis, although many patients with psoriasis have joint disease unrelated to psoriasis.

The diagnosis of psoriatic skin disease is usually made on clinical grounds alone, largely on the basis of the morphology and distribution of lesions. The differential diagnosis may be extensive and includes in selected cases nummular eczema, seborrheic dermatitis, candidiasis (in intertriginous areas), pityriasis rubra pilaris, Bowen’s disease or Paget’s disease (for isolated plaques), drug eruption, pityriasis rosea, pityriasis lichenoides, dermatophytosis, lichen planus, secondary syphilis, parapsoriasis, cutaneous lupus (especially subacute cutaneous lupus erythematosus [SCLE]), and dermatomyositis. In cases where the diagnosis is not clear-cut, biopsy may be helpful. The histologic findings may range from virtually diagnostic for psoriasis to merely consistent with but not diagnostic. Histologically, psoriasis cannot generally be distinguished from the skin lesions seen in reactive arthritis.

Common topical therapies include corticosteroids, tar, anthralin, calcipotriene, and tazarotene. Phototherapy using sunlight, broadband ultraviolet B (UVB), narrowband UVB, or psoralen ultraviolet A-range (PUVA) is still a mainstay of therapy for many patients. Common systemic therapies include methotrexate, acitretin, cyclosporine, and the relatively new biologic agents such as etanercept, adalimumab, infliximab, and ustekinumab. Cyclosporine may be useful to attain relatively rapid control of severe psoriasis, but it is less often used as a long-term therapy. Although topical corticosteroids are an acceptable treatment for many patients, systemic corticosteroids are avoided for the treatment of cutaneous disease, in particular due to the observation of severe flaring of psoriasis following withdrawal of systemic corticosteroids.

REACTIVE ARTHRITIS

The diagnosis of reactive arthritis may be rather straightforward in a young male who develops urethritis, conjunctivitis, and arthritis following an episode of nongonococcal urethritis. However, in many cases the clinical features are not fully expressed and cutaneous lesions may be helpful in establishing the diagnosis.

Circinate balanitis is the most common of the characteristic mucocutaneous lesions. Small erythematous papules and pustules coalesce to form serpiginous erosive or crusted plaques on the glans penis. In uncircumcised men, the appearance is more often that of erosion rather than crust because the moisture and trauma minimize the formation of crust. In circumcised men, crusting may be more obvious than erosion.

The palms and, particularly, the soles may develop lesions that are initially similar to the small erythematous papules and pustules of the genital region. With time, these lesions, termed keratoderma blennorrhagica, tend to become markedly hyperkeratotic (Figure 43-2). They may coalesce into large plaques or generalized hyperkeratosis involving the entire plantar surface, or they may remain discrete, erythematous, hyperkeratotic papules a few millimeters in diameter.

Figure 43-2  Reactive arthritis with keratoderma blennorrhagica of the feet.
Erythematous, scaly plaques indistinguishable from psoriasis may appear elsewhere on the skin including the scalp, elbows, and knees. When lesions occur around the nails, it is common for there to be hyperkeratosis underneath the nails. Pitting is not typical of reactive arthritis, but thickening, ridging, or shedding of the nail plate may occur. Erosions of the oral mucosa are relatively common on the tongue, buccal mucosa, and palate.

The diagnosis of the cutaneous lesions is usually made on a clinical basis. Skin biopsy may be helpful in excluding many entities in the differential diagnosis but generally cannot exclude psoriasis. Unfortunately, the major condition in the differential diagnosis of the skin lesions is usually psoriasis. One somewhat distinguishing histologic feature is that the older lesions of keratoderma blenorrhagica may have a considerably thickened stratum corneum, corresponding to the markedly hyperkeratotic papules seen grossly.

For the genital lesions, conditions to consider in the differential diagnosis may include candidiasis, psoriasis, dermatitis, Bowen's disease, Paget's disease, squamous cell carcinoma, Zoon's balanitis, erosive lichen planus, lichen sclerosus (balanitis xerotica obliterans), aphthosis, fixed drug eruption, and certain infectious diseases. The differential diagnosis for lesions on the soles and palms may include psoriasis, hereditary or acquired hyperkeratosis of the palms and soles, pustular eruption of the palms and soles, pompholyx, scabies, and dermatophytosis. The differential diagnosis for oral lesions may include geographic tongue, lichen planus, candidiasis, aphthae, and autoimmune bullous diseases.

The approach to treatment of skin lesions is similar to that for psoriasis, particularly in cases where the lesions are persistent. Choice of topical therapies may be somewhat limited due to the sites involved. The oral mucosa is a difficult site to deliver medication topically, and the genital area may develop irritant reactions to certain topical medications. Often, topical corticosteroids are preferred for both difficult site to deliver medication topically, and the genital area may develop irritant reactions to certain topical medications. In the genital area, suprainfection with Candida may occur and concurrent therapy with a topical or systemic anticandidal medication may be necessary on occasion.

**RHEUMATOID ARTHRITIS**

The major skin manifestations associated with rheumatoid arthritis (RA) generally fall under granulomatous lesions, exemplified by the rheumatoid nodule, and neutrophilic lesions, exemplified by vasculitis and pyoderma gangrenosum.

Rheumatoid nodules are the most common cutaneous manifestations of RA. They occur more often in seropositive patients and correlate somewhat with higher rheumatoid factor titers, more severe arthritis, and increased risk for vasculitis. Nodules are usually relatively deep, firm, and painless and tend to develop over areas of pressure and trauma such as the extensor forearms, fingers, olecranon processes, ischial tuberosities, sacrum, knees, heels, and posterior scalp (Figure 43-3). In patients who wear glasses, nodules may develop under the bridge or nosepieces. In most cases, rheumatoid nodules are in the subcutaneous tissue and/or deep dermis, but occasionally they may occur more deeply or more superficially.

Clinically, depending on the presentation, numerous entities may be in the differential diagnosis including infections, inflammatory disorders, and benign tumors. If needed, biopsy of a nodule may be quite helpful in establishing the diagnosis. Rheumatoid nodules exhibit a distinctive histologic finding called necrobiosis, a fibrinoid degeneration of the connective tissue, surrounded by palisaded histiocytes. Necrobiosis is also a characteristic feature of granuloma annulare and necrobiosis lipoidica diabeticorum. Although necrobiosis lipoidica diabeticorum is easily distinguished from rheumatoid nodule clinically, the subcutaneous variant of granuloma annulare may be difficult to distinguish, both clinically and histologically.

The term *rheumatoid nodulosis* has been used to describe an entity characterized by subcutaneous rheumatoid nodules, cystic bone lesions, rheumatoid factor positivity, and arthralgias in patients with little or no evidence of systemic manifestations of RA or erosive joint disease. Older males are preferentially affected.

The development of nodules in RA patients undergoing treatment with methotrexate has been noted by several observers and termed *accelerated rheumatoid nodulosis*. The nodules are newly appearing and occur preferentially on the hands. There are also case reports of the phenomenon in RA patients treated with etanercept.

The other major type of cutaneous lesion associated with RA is neutrophil predominant. Rheumatoid vasculitis occurs more frequently in patients who are seropositive and have rheumatoid nodules, and it often occurs relatively late in the course of the disease. Vessels of any size may be affected. In the skin, vasculitis may appear as purpuric papules and macules, nodules, ulcerations, or infarcts. Bywaters’ lesions are periungual or digital pulp purpuric papules representing a small vessel vasculitis, but not necessarily associated with vasculitic lesions elsewhere.

The differential diagnosis of purpuric or petechial lesions may include stasis dermatitis, Schambberg’s purpura, platelet dysfunction, petechial drug eruptions, viral exanthems, emboli, thromboses, and sludging. Of these, Schambberg’s purpura, a relatively common condition unassociated with systemic disease, is probably the most frequently confused
with small vessel vasculitis. Skin biopsy may be helpful in establishing the diagnosis of vasculitis, particularly if an early lesion is sampled, although rheumatoid vasculitis cannot be distinguished histologically from many other causes of small vessel vasculitis. Immunofluorescent examination of an early lesion may be helpful in ruling out IgA-predominant vasculitis. The differential diagnosis of ulcers and infarcts is extensive. Biopsy is often unrewarding because nonspecific changes present in established lesions may make interpretation difficult, but on occasion biopsy of ulcers or infarcts may result in a definitive diagnosis of vasculitis.

The neutrophilic dermatoses are a group of diseases inflammatory rather than infectious in origin, typified by pyoderma gangrenosum and Sweet's syndrome. These conditions have been associated with a variety of extracutaneous diseases including RA. The classic pyoderma gangrenosum lesion is a rapidly appearing, large, destructive ulcer in which the border is undermined. The classic lesion of Sweet's syndrome is an erythematous, edematous plaque with a surface often described as mammillated, pseudovesicular, or microvesicular. Clinical appearances intermediate between these two have been described. For pyoderma gangrenosum, the differential diagnosis is usually that of conditions causing leg ulcer, and the diagnosis is mainly clinical, with biopsy primarily serving to exclude some of the other entities under consideration. For Sweet's syndrome, the differential diagnosis may include infections, halogenoderma, and other neutrophilic dermatoses. Biopsy often provides helpful supporting evidence. The mainstay of therapy for acute lesions of both conditions is systemic corticosteroids. For more persistent lesions, a variety of options may be considered, cyclosporine and infliximab being two of the more common. Colchicine or potassium iodide may be first-line therapies for Sweet's syndrome, particularly in patients with infections or contraindications to corticosteroids.

The term *rheumatoid neutrophilic dermatitis* has been given to describe chronic, erythematous, urticarial-like plaques that occur primarily on the distal arms. Clinically and histologically, rheumatoid neutrophilic dermatitis is similar to Sweet's syndrome and may be a variant of it.

Palisaded neutrophilic and granulomatous dermatitis (PNGD) of connective tissue disease is an unusual condition or set of conditions for which consistent terminology is still evolving. As the name implies, the major bases for diagnosis of this entity are the histologic appearance and the occurrence in a patient with connective tissue disease, often RA. The clinical appearance ranges from erythematous or flesh-colored papules that appear primarily on fingers and elbows to erythematous or flesh-colored linear cords on the trunk. Some authors classify the latter as interstitial granulomatous dermatitis with cutaneous cords or interstitial granulomatous dermatitis with arthritis (IGDA). Treatment of PNGD and IGDA can be challenging. PNGD may respond to dapsone or sulafpyridine. IGDA can be treated with antimalarials or immunosuppressives, but because this is both a newly described and relatively infrequent condition, all evidence is based on case reports and small case series. Patients can progress to a severe deforming arthritis. In some cases, granuloma annulare and rheumatoid nodule may be in the differential diagnosis.

---

**JUVENILE RHEUMATOID ARTHRITIS/STILL’S DISEASE**

The majority of patients with classic Still's disease manifest an exanthematous eruption coincident with daily fever spikes. The lesions are evanescent, usually nonpruritic, erythematous macules occurring over the trunk, extremities, and face. The differential diagnosis includes viral exanthem, drug eruption, familial periodic fever syndromes, and rheumatic fever. It is not unusual for exanths of any type to be more prominent during fevers, but it is not expected that viral exanthems and drug eruptions clear completely between fever spikes. However, it should be noted that the eruption of erythema infectiosum (fifth disease) due to parvovirus B19 may resolve completely but reappear when the skin temperature rises, as with warm baths or exercise. Adult-onset Still's disease is also typified by an evanescent erythematous, sometimes salmon-colored eruption over the trunk and extremities, associated with high fever. Skin biopsy may be nonspecific. However, there has been a report of a unique histologic pattern consisting of dyskeratotic keratinocytes in the upper epidermis along with increased dermal mucin in adult-onset Still's disease. The frequency with which this histologic pattern is present in Still's disease remains to be determined.

Subcutaneous nodules may develop in both juvenile-onset and adult-onset Still's disease. The lesions tend to occur at the same sites of the body, as do rheumatoid nodules in RA, but histologically they appear similar to nodules of rheumatic fever.

**LUPUS ERYTHEMATOSUS**

The skin is involved at some time in the course of disease in the majority of patients with lupus erythematosus (LE), and skin lesions may be important in establishing the diagnosis. Some skin lesions are highly likely to be associated with “systemic” (i.e., extracutaneous) disease, whereas others may or may not be associated with extracutaneous disease. The phenomenon of lupus skin lesions occurring in the absence of systemic disease has previously been termed *discoid lupus* by some. However, dermatologists use this term to denote a specific type of skin lesion, regardless of presence or absence of systemic disease. It is the latter meaning of discoid lupus to which we refer in this chapter.

**Lupus-Specific Skin Lesions**

James Gilliam classified cutaneous lesions as being specific or nonspecific for lupus, discoid lupus lesions being an example of the former and palpable purpura being an example of the latter. Although this division of lesions is useful, sometimes a lupus-specific lesion occurs in a patient whose primary autoimmune disease is something other than LE. For example, SCLE lesions may occur in patients whose primary condition is Sjögren’s syndrome, and discoid lesions may be seen in a variety of conditions such as mixed connective tissue disease. Many of the lupus-specific skin lesions can occur in patients who have no evidence of extracutaneous disease.

The characteristic morphologies of the various lupus-specific skin lesions are in large part a function of the depth
and intensity of the inflammatory infiltrate, presence or absence of epidermal basal cell damage, involvement of hair follicles, abundance of dermal mucin, and tendency to scar. In practice, there may be some overlap of these features and there may be more than one type of lesion present in a given patient, making classification difficult. Because therapy for most of the lupus-specific lesions is similar, it is not always important to distinguish among the various types of lesions. However, it can be useful to identify conditions that are more likely to scar, in order to target more aggressive therapy, and to identify conditions that are highly likely or highly unlikely to be associated with systemic disease.

Acute cutaneous lupus (ACLE) lesions are typified by malar erythema, the classic butterfly rash (Figure 43-4). The inflammation tends to be superficial, with little propensity to scar. Precipitation or exacerbation of lesions by sun exposure is common, and lesions tend to be distributed on the sun-exposed face, neck, extensor arms, and dorsal hands, where the skin over the knuckles is relatively spared. Often the lesions are quite transient, but they may be persistent. When the face is severely affected, facial edema may be prominent. Oral lesions are often present concurrently. Acute eruptions with considerable focal basal cell damage can result in erythematous papules with dusky centers that clinically mimic erythema multiforme. The major importance of recognition of ACLE is its strong association with systemic disease. The differential diagnosis of malar rash may include several conditions. In some cases, the facial rash of ACLE may be difficult to distinguish from rosacea. Seborrheic dermatitis, atopic dermatitis, and photosensitive eruptions such as polymorphous light eruption and drug-induced photosensitivity may also be considered. Dermatomyositis may cause a photosensitive facial erythema with edema similar to ACLE, although the erythema tends to be more violaceous. Persistent lesions on the neck and arms may be indistinguishable from SCLE. Discoid lupus lesions occasionally appear in a butterfly distribution, where they can result in disfiguring scarring. Skin biopsy is usually not performed on malar erythema because of its transient character, the scar resulting from biopsy, and the availability of other means of establishing the diagnosis of SLE. If a biopsy is done, it should be noted that dermatomyositis and SCLE cannot be distinguished from ACLE by histology and also that skin biopsy findings are sometimes nonspecific.

SCLE is a photosensitive eruption usually associated with anti-Ro(SSA) autoantibodies.
Tumid lupus (TLE) skin lesions are similar to DLE lesions in that they are erythematous indurated papules and plaques with a substantial lymphocytic infiltrate. Unlike DLE, though, the lesions do not exhibit epidermal abnormalities, follicular involvement, or scarring. Considerable mucin is present in the dermis, giving the lesions a somewhat boggy look and feel. In some reports, lesions are most common on the face and may be reproduced by phototesting. The risk for SLE appears to be low, and immunoglobulin deposits are not generally present in skin biopsies. Jessner’s lymphocytic infiltrate and other lymphocytic and granulomatous infiltrative conditions (see earlier) are in the differential diagnosis. Skin biopsy for routine histology is valuable in establishing the diagnosis, with the exception of reliably distinguishing TLE from Jessner’s lymphocytic infiltrate. Some have argued that Jessner’s lymphocytic infiltrate and TLE are one and the same, and it might reasonably be argued that what is called TLE is not appropriately classed as a form of chronic cutaneous LE but rather as an independent entity. However, the presence of TLE lesions in some patients with lupus is evidence to the contrary.

Lupus panniculitis (LEP) lesions have inflammation in the subcutaneous tissue, resulting in deep indurated plaques that become disfiguring, depressed areas. Usual sites of involvement are face, upper trunk, breasts, upper arms, buttocks, and thighs. The risk for SLE is not known precisely, but clearly some patients with LEP have or will develop SLE. The differential diagnosis is that of the panniculitides, but the distribution exhibited in LEP is unusual for most other conditions that cause panniculitis. The combination of clinical presentation and skin biopsy for histology usually serves to establish the diagnosis.

Some unusual variants of cutaneous lupus are chilblain lupus (red or dusky plaques on colder areas of skin such as fingers, toes, nose, elbows, knees, and lower legs), cutaneous lupus/lichen planus overlap, and a bullous eruption due to immunofluorescence may provide additional supporting diagnostic information. Lesions are expected to have granular deposits of immunoglobulins at the dermal-epidermal junction. Unless there is concomitant systemic disease, normal skin is expected not to have immunoglobulin deposits.

Discoid lupus erythematosus (DLE) lesions are the most common of the persistent lupus-specific skin lesions. Active lesions are erythematous papules and plaques that feel indurated to palpation because of the substantial numbers of inflammatory cells infiltrating the dermis. Involvement of hair follicles may be grossly evident as follicular plugs and scarring alopecia. Dyspigmentation is common, often with hypopigmentation or even depigmentation in the center of lesions. Visible scale is common and occasionally is pronounced in hypertrophic DLE. In established lesions, scarring may be disfiguring. Lesions tend to occur on the scalp, ears, and face but may be widespread and occasionally involve mucosal surfaces. It is unusual to have lesions below the neck in the absence of lesions above the neck. Sun exposure may exacerbate DLE in some cases, but the presence of lesions in sun-protected areas of the scalp and ears and the frequent absence of a history of photosensitivity indicates that sun exposure is probably not a trigger in every instance. There are case reports of squamous cell carcinoma developing in established DLE lesions. In a patient who presents with DLE lesions, the risk for developing SLE is probably about 5% to 10%, although mild systemic symptoms such as arthralgias are relatively common. The differential diagnosis of DLE lesions is often that of conditions exhibiting intense lymphocytic or granulomatous infiltrates such as sarcoid, Jessner’s lymphocytic infiltrate, granuloma faciale, polymorphous light eruption, lymphocytoma cutis, and lymphoma cutis. In the scalp, lichen planopilaris and other scarring alopecias may be considered. Skin biopsy for routine histology often establishes the diagnosis definitively. In more difficult cases, biopsy for
autoantibodies to type VII collagen or other basement membrane zone proteins. Not all bullae related to lupus are due to autoantibodies to basement membrane proteins, however. It is not unusual to develop bullae simply from intensive destruction of the basal cell layer in ACLE, SCLE, or, rarely, DLE.

Treatment of the lupus-specific lesions is relatively similar for most of the subtypes, with some exceptions and modifications. Sun protection is critical for lesions that are initiated or exacerbated by sun exposure. Many or most patients underestimate the amount of sunscreen needed to apply, the potential damage of the seemingly minimal exposure one has in the course of day-to-day activities, and the value of protective clothing. Topical therapy is often used to avoid side effects of systemic medications or to provide adjunctive therapy, although topical agents are unlikely to be beneficial if the disease process is deep, as in panniculitis.

Topical or intralesional corticosteroids are the most often used local therapy, but there are some reports of benefit from topical calcineurin inhibitors and topical retinoids. The first-line systemic medication for cutaneous lupus is antimalarial therapy. Several reports indicate that smoking tobacco decreases the likelihood of response to antimalarials. For antimalarial-resistant skin disease, a wide variety of medications have been used but there is no clear second choice when antimalarials have not worked. Although dapsone is arguably not helpful in most types of cutaneous lupus, it may be helpful in neutrophil-predominant bullous eruptions.

Measures to keep the skin warm may be useful for chilblains.

**Nonspecific Cutaneous Lesions**

A wide variety of lupus nonspecific skin lesions has been reported. Many of these such as vasculitic lesions are cutaneous clues to the possibility of extracutaneous disease. Noteworthy in this regard is livedo reticularis. The netlike erythema of livedo reticularis is a vascular phenomenon due to lowered oxygenation at the periphery of the area supplied by a particular vessel. This can simply be due to vasoconstriction, such as occurs in a cold environment, and thus can be a benign finding. If livedo is more prominent than usual, not corrected by warming, and persistent, it can indicate lowered flow due to pathology such as vasculitis, arteriosclerotic disease, or sludging. In lupus, livedo reticularis may be a sign of the presence of antiphospholipid antibodies.

Other lupus nonspecific skin lesions include Raynaud’s phenomenon, palmar erythema, periungual telangiectasia, alopecia, erythromelalgia, papulonodular mucinosis, and anetoderma. Sclerodactyly, calcinosis, and rheumatoid nodules have been reported but may be more likely in overlap syndromes than in SLE.

**NEONATAL LUPUS SYNDROME**

Neonatal lupus erythematous (NLE) is associated with maternal IgG autoantibodies to Ro/SSA and La/SSB. Affected children may have cutaneous lesions, cardiac disease (notably complete heart block and/or cardiomyopathy), hepatobiliary disease, or hematologic cytopenias. Most children have only one or two features of the disease.

Similar to the anti-Ro/SSA-associated SCLE of adults, the skin lesions are often photosensitive, have relatively superficial inflammatory infiltrates, and do not tend to scar. The lesions usually appear at a few weeks of age but have been noted at birth in several cases. The natural history of the skin disease is that the lesions last for weeks or months and resolve spontaneously, usually leaving no residuum. In a few cases, persistent telangiectasias have been noted. Individual lesions appear as erythematous annular papules or plaques. Lesions are usually more numerous and more intensely inflamed on the face and scalp but may additionally occur on the trunk and extremities. Confluent periorbital erythema, giving the appearance of an erythematous mask, is common and diagnostically helpful. Even though the skin disease resolves and most children without extracutaneous involvement remain otherwise healthy, there is a possibility that children who have had NLE are at increased risk for the development of autoimmune disease later in childhood.

Differential diagnosis of the skin lesions may include reactive erythema, drug eruption, erythema multiforme, and urticaria. Annular NLE lesions usually have little or no scale, unlike the annular lesions of tinea. In areas where there is intense destruction of the basal cell layer, lesions may be crusted and look similar to bullous impetigo. Treatment of skin lesions consists largely of sun protection and mild topical steroids.

The pathogenesis of lupus is covered elsewhere, but it is noteworthy that SCLE-like, anti-Ro/SSA-associated skin lesions may occur in neonates, but other lupus-specific skin lesions do not appear to be maternally transmissible.

**SJÖGREN’S SYNDROME**

The most common mucocutaneous findings of Sjögren’s syndrome are related to glandular dysfunction. Lacrimal gland dysfunction causes dryness and irritation of the eyes and can lead to keratitis and corneal ulceration. Salivary gland dysfunction causes dry mouth and may result in angular cheilitis and numerous dental caries. Vaginal xerosis may cause burning and dyspareunia. The skin may be dry, cracked, and pruritic. Mildly dry mucous membranes and even severely dry skin may be present in a substantial percentage of normal individuals who live in dry climates, so the findings should be interpreted in the context of the setting. In Japanese patients with Sjögren’s syndrome, an annular erythema has been described that is somewhat reminiscent of annular SCLE or annular lesions of NLE, although more indurated.

Vasculitis is a relatively common finding. In one series of 558 patients with primary Sjögren’s syndrome, 52 had vasculitis, typically involving small vessels. In most cases, lesions were purpuric, but in some, urticarial vasculitis was the clinical presentation. Patients with cutaneous evidence of vasculitis generally had more severe systemic disease.

**DERMATOMYOSITIS**

The current ACR criteria for dermatomyositis (DM) do not recognize the existence of amyopathic dermatomyositis. This has led to a problem in diagnosing patients with predominantly skin involvement with DM. Amyopathic
can result in extensive alopecia. Hyperkeratosis of the palmar and lateral surfaces of the fingers, called mechanic’s hands, can be associated with anti-Jo-1 autoantibodies and interstitial lung disease. Rarely patients can have a panniculitis. Vasculopathy, with livedo reticularis and ulceration, can occur. Itching can result in excoriations and lichenification. Damage lesions include postinflammatory hyperpigmentation, poikiloderma, calcinosis, lipoatrophy, and depressed scars. Poikiloderma is a descriptive term for a pattern of finely mottled white areas and brown pigmentation, telangiectasia, and atrophy.

Skin biopsy from a patient with cutaneous DM is identical to that seen with cutaneous lupus erythematosus. A diagnosis of DM is made by clinical-pathologic correlation and does not need to include muscle disease. In the absence of clinical muscle findings, the workup for DM should include muscle enzyme testing, pulmonary function testing, chest radiograph, electrocardiogram, and evaluation for underlying malignancy. Patients with amyopathic DM have the same incidence of pulmonary involvement as classic DM, and about 25% of patients have evidence of pulmonary fibrosis on high-resolution CT.

There has been an increased association of dermatomyositis including ADM with underlying malignancy. The most frequent malignancies are lung, ovarian, pancreatic, stomach, colorectal, and non-Hodgkin’s lymphoma.

Patients with DM frequently experience a delay in obtaining a diagnosis, and the presence of photosensitivity, malar rash, oral ulcers, and a positive antinuclear antibody (ANA) test means that they frequently get misdiagnosed as having SLE, having met the current criteria for the disease. The patients with both skin and muscle disease frequently experience resolution of their muscle disease after aggressive treatment with glucocorticoids, with or without immunosuppressives. Patients are sometimes treated with intravenous immunoglobulin (IVIG), cyclosporine, or tacrolimus. Patients with AMD or residual skin disease after treatment often benefit from hydroxychloroquine. Patients who do not improve with a single antimalarial can benefit from the addition of quinacrine or a switch from hydroxychloroquine to chloroquine. Immunosuppressives like methotrexate, azathioprine, or mycophenolate mofetil can be of additive benefit for patients with resistant skin disease. IVIG can be of benefit for skin findings of DM, and studies related to the role of biologics in treatment are ongoing.

**SCLERODERMA AND OTHER SCLEROSING CONDITIONS**

**Morphea**

Scleroderma can occur as localized or systemic disease. The localized form of the disease occurs as localized or generalized morphea, linear scleroderma, or facial hemiatrophy, otherwise known as Parry-Romberg syndrome. Linear scleroderma can occur over the forehead, in a variant called en coup de sabre (Figure 43-11). Morphea is seen more in adults, with increased incidence with advancing age, whereas linear scleroderma occurs more frequently in children and...
salivary glands and hemiatrophy of the tongue on the same side as the facial atrophy. Any reparative surgical treatment should be timed to occur no sooner than 1 year after cessation of the ongoing atrophic process. ANAs are positive 46% of the time in localized scleroderma. A positive ANA correlates with disease severity. A peripheral eosinophilia and hypergammaglobulinemia can be seen.

Treatment of localized scleroderma includes protection from trauma and cold, antimalarials (plaquenil alone, hydroxychloroquine and quinacrine together, or chloroquine with or without quinacrine), low-dose prednisone for eosinophilic fasciitis, topical calcipotriene, methotrexate, and mycophenolate mofetil. Phototherapy with narrow band UVB, UVA1, PUVA, and topical photodynamic therapy has also been used.

**Systemic Scleroderma**

Patients with early limited cutaneous scleroderma have Raynaud’s phenomenon for many years, minimal constitutional symptoms, puffy fingers, limited skin thickening, and anticentromere antibody. Patients with early diffuse cutaneous scleroderma frequently have delayed Raynaud’s, acute onset, many constitutional symptoms, arthralgias, tendon friction rubs, swollen puffy hands, and early diffuse skin thickening. They may have anti-Scl-70 antibody, as well as anti-RNA polymerase III. The major and minor criteria for diffuse cutaneous scleroderma include many skin findings. Major criteria include proximal scleroderma. Minor criteria include sclerodactyly (Figure 43-12), digital pitting and scars of the fingertips, loss of substance of the finger pad, and bibasilar pulmonary fibrosis. Ischemia and skin changes in systemic scleroderma can result in skin ulcers. Therapies used in the cutaneous treatment of systemic scleroderma include D-penicillamine, methotrexate, cyclophosphamide, photopheresis, and bone marrow transplant. Treatment of Raynaud’s phenomenon is covered elsewhere. There are reports of bosentan, an endothelin receptor antagonist, and phosphodiesterase inhibitors such as sildenafil being helpful in the treatment of skin ulcers. Patients who have pruritus may benefit from systemic antihistamines.
Eosinophilic Fasciitis

Eosinophilic fasciitis (EF) involves inflammation of the fascia overlying muscle and results in swelling of the extremities, followed by fibrosis and contractures. The digits are typically spared. There is often a rapid onset of disease activity, particularly following physical exertion. A contaminant of L-tryptophan was associated with an EF-like disease in the early 1990s. Approximately 30% of EF patients have morphea concurrently.

Diagnosis is based on a deep, usually excisional biopsy of the skin that includes fascia. There are inflammatory cells among collagen bundles, thickening of collagen, sclerosis of the dermis and fat/fascia, and absent sweat glands and hair.

POEMS Syndrome

POEMS syndrome includes polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes. Polyneuropathy and a monoclonal plasmoproliferative disorder must be present along with one other minor criterion including sclerotic bone lesions, Castleman’s disease, organomegaly, edema, endocrinopathy, papillae, derma, or skin changes. Other associated findings may include ascites, pleural effusions, thrombocytosis, fingernail clubbing, and white nails. The sensorimotor polyneuropathy has both demyelinating and axonal features and is slowly progressive and debilitating in most cases. Organomegaly may be hepatosplenomegaly or lymph node enlargement. Endocrinopathies include diabetes, impotence, gynecomastia, and hypothyroidism. The monoclonal (M) protein abnormality consists of IgA or IgG heavy chains with light chains. The skin changes may consist of hyperpigmentation, hypertrichosis, hyperhidrosis, skin thickening, telangiectasia, and glomeruloid hemangiomas.

Treatment with corticosteroids, low-dose alkylators, and high-dose melphalan and autologous peripheral blood stem cell transplantation has been reported to be successful.

Scleromyxedema

Scleromyxedema is a rare disorder frequently characterized by dysproteinemia and widespread skin changes. The paraprotein has been shown to stimulate fibroblast production of mucin, suggesting a causal role and a rationale for lowering it. Patients have waxy papules on the face, neck, upper trunk, forearms, hands, and thighs that become confluent and occur in association with underlying sclerosis, which can lead to joint contractures, sclerodactyly, and carpal tunnel syndrome. Common extracutaneous manifestations include upper gastrointestinal dysmotility, muscle weakness, joint contractures, and neurologic symptoms such as seizures, encephalopathy, coma, and obstructive or restrictive pulmonary disease. Eighty percent of patients have a monoclonal gammopathy, most frequently IgG-λ, but occasionally IgG-κ.

Skin biopsy shows mucin deposition and a proliferation of fibroblasts in the upper dermis. Treatment has included prednisone, IVIG, PUVA, systemic retinoids, thalidomide, interferon alfa-2a, plasmapheresis, photopheresis, low-dose melphalan, and high-dose dexamethasone.

Chemotherapeutic agents such as melphalan, cyclosporine, cladribine, cyclophosphamide, methotrexate, and chlorambucil have been used. Successful remission after autologous stem cell transplantation has been reported.

Nephrogenic Systemic Fibrosis

Nephrogenic systemic fibrosis (NSF) is a relatively new illness that occurs in patients with renal disease, with most of the patients having undergone dialysis for renal failure. NSF presents as either a morphea-like disease or a more diffuse acral sclerosis. Morphea-like presentations include ill-defined indurated plaques, with islands of sparing and finger-like projections that involve lower more than upper extremities. More diffuse confluent acral sclerosis, sometimes with truncal involvement, can occur. There are often yellow plaques on the conjunctiva. Patients can experience pain, severe itching, joint contractures, fibrosis and calcification of the skin, subcutaneous tissue, fascia, muscle, myocardium, lungs, renal tubules, and testes. Patients typically do not have Raynaud’s syndrome. Skin biopsy is identical to that seen with scleromyxedema, with stellate fibroblasts, glycosaminoglycans, and thickening of collagen.

TGF-β is increased in lesional skin. There is no proven effective therapy, and prognosis depends on the extent, rapidity of skin involvement, and severity of the systemic disease. Treatments that have been reported to potentially help include plasmapheresis, IVIG, immunosuppressives, glucocorticoids, interferon-α, thalidomide, PUVA, UVA1, photopheresis, and imatinib mesylate.

Recent studies suggest a strong association of NSF with gadolinium exposure.

PRIMARY NECROTIZING VASCULITIS INVOLVING THE SKIN

In cutaneous diseases, vasculitis is a term typically reserved for lesions characterized by damage to vessel walls and a neutrophilic or granulomatous infiltrate. Classification, pathogenesis, diagnostic evaluation, and therapy are covered elsewhere in the text. In this chapter, the focus is on the skin findings and differential diagnosis. The primary determinant of the appearance is the size of the vessel affected.

Leukocytoclastic Small Vessel Vasculitis and Its Variants

Leukocytoclastic small vessel vasculitis is a rather common condition with characteristic histologic findings of fibrinoid necrosis of vessel walls, a neutrophil-predominant infiltrate, and leukocytoclasia (i.e., fragmented nuclei resulting from degeneration of neutrophils). In the skin, the vessels involved are in the dermis and damage to these vessels results in lesions of a characteristic size. The lesions are both erythematous and purpuric and are distinctly palpable if there are sufficient numbers of neutrophils in the lesion. On physical examination, it may be helpful to palpate several lesions because it is common for the majority of the lesions to be nonpalpable. The usual diameter of the purpuric papules is about 0.3 to 0.6 cm, although smaller and larger lesions may be observed (Figure 43-13). Discrete lesions
Leukocytoclastic vasculitis demonstrating nonblanching, purpuric macules.

have a round shape. The center may look dusky, pustular, or ulcerated, or it may appear as a hemorrhagic vesicle. Larger ulcerations may occur when lesions coalesce. Particularly in larger lesions, the devitalized tissue may be a focus for secondary bacterial infection. Lesions tend to occur in dependent areas. Thus for ambulatory patients, lesions are most numerous on the lower legs. Koebnerization, the appearance of lesions along lines of trauma, such as scratches, is sometimes observed.

As mentioned under the discussion of rheumatoid vasculitis, the differential diagnosis of purpuric or petechial lesions may include stasis dermatitis, Schamberg’s purpura, platelet dysfunction, petechial drug eruptions, viral exanthems, emboli, thromboses, and sludging. Skin biopsy is often helpful in establishing the diagnosis of small vessel vasculitis, especially if an early lesion is sampled. Immunofluorescence of an early lesion may establish whether the vasculitis is IgA predominant.

Small vessel vasculitis unassociated with connective tissue disease and not IgA-predominant is sometimes called hypersensitivity vasculitis. It is often apparently confined to the skin and therefore has a good prognosis. In some cases, infection or drug may be implicated, but often there is no definitive initiating event discovered. Therapy is directed first toward treating the underlying cause, if a cause is found. If there is no significant extracutaneous disease, treatment is often symptomatic. Leg elevation, compression stockings, and reduction of activity may be helpful. Nonsteroidal anti-inflammatory agents or antihistamines are sometimes used. Systemic corticosteroids are not routinely indicated for skin-limited disease. For patients with persistent disease confined to the skin, colchicine and dapsone have each been used with some success.

A common variant of small vessel vasculitis is Henoch-Schönlein purpura (HS purpura). HS purpura occurs commonly in children and is often associated with extracutaneous findings of gastrointestinal or renal involvement. The typical lesion of IgA-predominant small vessel vasculitis is an erythematous or urticarial macule or papule that evolves rapidly into palpable purpura. It has been noted that IgA-predominant vasculitis in particular may display superficial plaques of palpable purpura or a retiform configuration of lesions. The most reliable means of establishing that the vasculitis is IgA predominant is biopsy for immunofluorescence. IgA-predominant vasculitis is not unusual in adults. Compared with the presentation in children, there is a lower association with preceding upper respiratory infection and higher association with medication use.

Mixed cryoglobulinemia can present as a small vessel vasculitis. Additional skin findings that may occur include livedo reticularis, urticarial papules, cold urticaria, Raynaud’s syndrome, acrocyanosis, leg ulcers, and digital ulceration or gangrene. Hepatitis C is a frequent association. Patients with type I monoclonal cryoglobulinemia may more often have purpura due to cryogelling rather than a true vasculitis.

A distinct subset of patients with small vessel vasculitis has lesions that are primarily urticarial rather than purpuric. The main diagnosis in the clinical differential is urticaria. Individual lesions of urticaria tend to be short-lived, usually less than 24 hours, whereas individual lesions of urticarial vasculitis tend to last several days. Additional skin findings may include angioedema, livedo reticularis, nodules, and bullae. In some lesions, foci of purpura may be observed. Urticarial vasculitis has been classified into two groups: normocomplementemic and hypocomplementemic. Extracutaneous disease is more likely to occur in the hypocomplementemic group, and some of these patients have underlying SLE. It has been proposed that there is a distinctive subset of the hypocomplementemic group characterized by IgG antibodies to C1q, angioedema, ocular inflammation, arthritis, obstructive pulmonary disease, and renal disease. The pulmonary disease tends to be severe and life threatening. This subset has been termed hypocomplementemic urticarial vasculitis syndrome.

Erythema elevatum diutinum is an unusual form of small vessel vasculitis that is characterized by erythematous or violaceous papules, plaques, and nodules over the dorsal hands, ears, knees, heel, and buttocks. In the clinical differential diagnosis is Sweet’s syndrome, multicentric reticulohistiocytosis, sarcoidosis, and lymphoma, among others. With time, fibrosis often occurs. Established lesions may be disfiguring and may have an appearance somewhat reminiscent of keloids. Although significant extracutaneous involvement is not expected and many patients are otherwise well, erythema elevatum diutinum has been reported in association with various autoimmune, infectious, and hematologic conditions including streptococcal infection, paraproteinemia, inflammatory bowel disease, RA, SLE, and HIV. For active skin lesions, dapsone is the usual treatment. Intraleisonal corticosteroids are sometimes used for fibrotic lesions.

Acute hemorrhagic edema of childhood is an uncommon but generally benign and self-limited form of vasculitis usually occurring in children younger than the age of 2 years, commonly preceded by an upper respiratory infection or medication. The clinical appearance may be dramatic, with large purpuric plaques on the face, ears, and extremities. On the basis of the appearance of the skin lesions, meningococcemia is sometimes suspected but the child with acute hemorrhagic edema appears relatively healthy. Generally there is no extracutaneous involvement. Treatment is symptomatic.
Granulomatous Vasculitides

Skin lesions are occasionally the presenting feature of granulomatosis with polyangiitis (formerly Wegener's granulomatosis). The most common type of lesion is palpable purpura, with or without necrosis. Many other types of lesions have been noted including papules, ecchymoses, hemorrhagic bullae, necrotic papules, subcutaneous nodules, and ulcers. Ulcerative lesions may be similar in appearance to pyoderma gangrenosum, although lacking an undermined border. Oral ulcers are relatively common but nonspecific in appearance. A more specific oral finding is hypertrophic gingival inflammation with petechiae.

Skin biopsy is sometimes helpful diagnostically, but unfortunately, biopsies are often nonspecific or show leukocytoclastic vasculitis. True granulomatous vasculitis is not often observed in skin specimens. Extravascular granulomatous inflammation is sometimes noted and may be more likely in nonpurpuric papules or nodules than in palpable purpura.

Differential diagnosis of the skin lesions includes other small vessel vasculitides, particularly when cutaneous lesions consist of palpable purpura and the biopsy finding is leukocytoclastic vasculitis. If granulomatous inflammation is observed, the differential may include Churg-Strauss syndrome, polyarteritis nodosa, microscopic polyangiitis, RA, SLE, infection, lymphoproliferative disorders, chronic active hepatitis, erythema nodosum, granuloma annulare, and inflammatory bowel disease.

Patients with Churg-Strauss syndrome characteristically present with respiratory symptoms, but skin lesions are common during the vasculitic phase of the disease. Hemorrhagic lesions ranging from petechiae to palpable purpura to ecchymoses, cutaneous nodules with or without ulceration, subcutaneous nodules, and nonspecific erythematous eruptions are most common. As with granulomatosis with polyangiitis, hemorrhagic lesions tend to show small vessel vasculitis on biopsy, and nodules are more likely to demonstrate granulomas. The clinical and histologic differential diagnoses often include polyarteritis nodosa, granulomatosis with polyangiitis, and microscopic polyangiitis. As alluded to earlier, the histologic finding of Churg-Strauss granuloma is not specific for Churg-Strauss and may be seen in several entities. Large numbers of eosinophils in the biopsy may be diagnostically helpful but not definitive. Hypereosinophilic syndrome may share clinical and laboratory features with Churg-Strauss, but vasculitis is not characteristic.

Polyarteritis Nodosa and Related Conditions

Classic polyarteritis nodosa (PAN) and microscopic polyangiitis have historically been classified together as PAN but appear to be distinct conditions distinguishable in part by the size of vessels affected. Classic PAN involves mediumsized vessels, whereas microscopic polyangiitis primarily involves vessels ranging in size from capillaries to arterioles.

Cutaneous findings of classic PAN represent damage downstream of the affected vessel and consist of ulceration, digital gangrene, ecchymoses from vessel rupture, livedo reticularis, and subcutaneous nodules that may follow the course of arteries. Because the affected vessel is proximal to the skin, skin biopsy is likely to show nonspecific findings.

The skin findings of microscopic polyangiitis reflect the smaller size of vessels affected. Petechiae, palpable purpura, purpuric plaques, erythematous nodules, and ulcerations may be observed. There have been some reports of livedo reticularis in association. The cutaneous pathology is that of a leukocytoclastic vasculitis, but, in contrast to many of the small vessel vasculitides previously discussed, may involve small arterioles. The differential diagnosis is mainly that of other small vessel vasculitides. Size of vessels affected, antineutrophilic cytoplasmic antibody (ANCA) positivity, paucity or absence of antibody deposits in vessels, and spectrum of extracutaneous involvement aid in distinguishing microscopic polyangiitis from other vasculitides.

There is a variant of PAN termed benign cutaneous PAN or, perhaps more accurately, primarily cutaneous PAN. In this condition, arterioles in the subcutaneous fat and lower dermis are affected, and the presentation is often that of tender subcutaneous nodules and livedo reticularis. Associations with Crohn’s disease, hepatitis B, and hepatitis C have been reported. Clinically, panniculitides such as erythema nodosum and erythema induratum may be considered, although livedo reticularis is not expected. Biopsy for diagnosis should include subcutaneous fat. Even after biopsy, microscopic polyangiitis may be difficult to exclude. Although relatively little information is available concerning ANCA in cutaneous PAN, a negative ANCA is more consistent with cutaneous PAN than with microscopic polyangiitis. Although the outcome is benign, the course is often chronic and relapsing. Therapy is typically conservative and may consist of intravenous corticosteroids, nonsteroidal anti-inflammatory agents (NSAIDs), low-dose methotrexate, dapsone, or occasionally systemic corticosteroids.

Kawasaki disease is considered a polyarteritis nodosa variant due to the involvement of coronary arteries. Skin findings constitute several of the criteria for diagnosis. None of the findings is specific for Kawasaki’s disease, but the constellation of findings establishes the diagnosis. Kawasaki disease usually affects young children but has been reported in adults. Criteria for diagnosis are otherwise unexplained fever for at least 5 days and four of the following five findings: (1) bilateral nonexudative conjunctivitis; (2) injected pharynx, strawberry tongue, or injected or fissured lips; (3) erythema of palms or soles, hand and foot edema, and, in the convalescent phase, desquamation; (4) erythematous, polymorphous, generalized skin eruption; and (5) cervical lymphadenopathy. In addition, erythema of the perineal region is common and transverse lines across the fingernail beds have been noted in a few cases.

Large Vessel Vasculitis

Skin findings of temporal arteritis (giant cell arteritis) consist mainly of palpable temporal arteries, skin tenderness in the area, and scalp nodules or ulcerations. Skin lesions in patients with Takayasu’s arteritis may include Raynaud’s phenomenon, livedo reticularis, ulcerated nodules, subcutaneous nodules, and pyoderma gangrenosum-like ulcers. Skin biopsy is generally not performed in these conditions.
INFECTIONS

Many infectious diseases present with both skin and rheumatologic findings. This section will highlight a few examples.

Lyme Borreliosis

*Borrelia burgdorferi*, the causative agent of Lyme disease in North America, is associated with erythema migrans (EM). In Europe the related genospecies *Borrelia afzelii* is associated with both erythema migrans and acrodermatitis chronica atrophicans, and several European studies have found compelling evidence for *B. afzelii* infection in patients with morphea. There has been no similar association of *Borrelia* with morphea in the United States. Hematogenous dissemination from the initial skin site is believed to cause secondary skin lesions and extracutaneous manifestations, and only certain subtypes of *B. burgdorferi* are associated with dissemination.

EM is the first manifestation of Lyme disease in 60% to 80% of people and occurs at the site of the tick bite. At the time of the skin lesion, which occurs within a few days to a month after the bite, the spirochetes enter the circulation and disseminate. The skin findings may be associated with fever, chills, fatigue, headache, neck stiffness, myalgias, arthralgias, conjunctivitis, erythematous throat, and regional or generalized lymphadenopathy. The lesions of EM begin as red macules that become papular and then expand into an erythematous, annular plaque (Figure 43-14). Two forms of EM exist. In one, there is an expanding red plaque with varying intensities of redness within the plaque. In the second, there is a target, with a central red plaque surrounded by normal-appearing skin, which in turn is surrounded by another band of erythema. They can enlarge rapidly, and multiple lesions due to hematogenous spread are seen 17% of the time. As the lesion enlarges, the central erythema can fade. The central portion of the lesion may be edematous, vesicular, urticarial, or crusted.

Triangular and elongated oval lesions have been described, but circular lesions are most frequent. The most common locations are the inguina, axillae, abdomen, and behind the knees. EM lesions are usually asymptomatic but can be pruritic or painful. Untreated EM lesions resolve in a median of 28 days, with a range from 1 day to 14 months. Resolution is within a few days after treatment with antibiotics such as doxycycline or penicillin.

Acrodermatitis chronica atrophicans (ACA) is associated with late Lyme disease. It occurs mainly in women between ages 40 and 70. The lesions begin on an extremity, usually the lower leg or foot as a bluish-red edematous plaque. Fibrous bands may develop, especially on the ulnar and tibial regions, and fibrous nodules may form near joints. Regional lymphadenopathy is often present. Over many years, the skin becomes atrophic. *B. burgdorferi* has been isolated from the skin of patients with ACA.

Parvovirus

Presentation of parvovirus B19 includes an erythematous “slapped cheeks” appearance; a lacy, reticulated proximal extremity rash; a febrile petechial eruption; and papular-purpuric gloves and socks syndrome (PPGSS). The infection is self-limited and generally resolves spontaneously within 1 to 2 weeks. Laboratory findings may include mild or severe leukopenia, transient neutropenia or relative neutropenia, eosinophilia, and mild thrombocytopenia. Adults who are infected often contract the virus from infected children and commonly present with systemic disease including arthropathy and a flulike illness.

Atypical Infections: *Mycobacterium marinum*

Many types of mycobacteria, atypical mycobacteria, and deep fungal infections can affect skin and joints. *Mycobacterium marinum* is an example and can be acquired through exposure to fresh water, salt water, fish tanks, swimming pools, fish or aquatic exposures, timber cuts, or splinters. The incubation period is usually about 3 weeks, although much longer periods are possible. The disease often occurs after inoculation into abrasions or after penetrating injuries to the fingers and hands. This is an indolent disease, with nodules or ulcerated plaques, occasionally with extension to deep tissue. Common areas of involvement are the fingers, dorsum of the hands, and knees. The lesions can be localized or sporotrichoid (25%), with dissemination 2% of the time.

PANNICULITIS

Panniculitis refers to a group of diseases that manifest as inflammation or alterations in the subcutaneous fat. The complexity of etiologies for even one form of panniculitis such as erythema nodosum, the relative rarity of most forms of panniculitis, and the number of different panniculitides has slowed the scientific development. The etiologies for many panniculitides are still poorly understood.

Panniculitis may be primary without an identifiable cause or secondary. Common secondary causes of panniculitis include infection, trauma, pancreatic disease, immunodeficiency states, malignancies, and connective tissue...
Erythema nodosum remains the most common form of panniculitis, and although there is a long list of diseases and medications that have been associated, it is frequently not associated with an identifiable underlying condition. Some underlying conditions associated with erythema nodosum include inflammatory bowel disease; sarcoidosis; malignancies such as leukemia and lymphoma; infections (bacteria, Yersinia, rickettsiae, chlamydia, spirochetes, and protozoal disease); pregnancy; drugs (sulfonamides and contraceptives); and autoimmune diseases such as Behçet’s disease, Sjögren’s syndrome, reactive arthritis, and SLE.74

As the understanding of lobular panniculitis has expanded, cases that were lumped into the wastebasket term of “Weber-Christian” disease are now recognized to be clearly definable and separate entities such as lupus panniculitis, cytophagic histiocytic panniculitis, α1-antitrypsin deficiency, factitial panniculitis, traumatic panniculitis, and calciphylaxis.75-77

Infections are recognized as a trigger of panniculitis, as exemplified by erythema nodosum due to streptococcal infection, hepatitis B or C associated with polyarteritis nodosa, infectious panniculitides often in immunocompromised hosts, and most recently some cases of erythema induratum/nodular vasculitis associated with Mycobacterium tuberculosis. In addition, atypical infections can themselves cause lesions that look like panniculitis.

An understanding of the heterogeneity of lymphomas that involve the fat is still evolving, but advances in differentiating various histologic and clinical outcomes are occurring.78 Some patients thought to have lupus panniculitis on the basis of cytophenias and laboratory tests are actually diagnosed with subcutaneous lymphoma after careful review of their pathology.

Patients with panniculitis frequently have erythematous tender nodules, and the clinical presentation frequently is not specific enough to allow determination of the exact subtype of panniculitis without a biopsy. Patients with panniculitis can have associated symptoms such as low-grade fevers, fatigue, arthralgias, and myalgias.

An adequate skin biopsy, often involving an elliptical excision, is essential to properly diagnose the various entities that fall into the category of panniculitis (Table 43-1). Panniculitis is typically classified into four main subgroups: septal, lobular, mixed panniculitis, and panniculitis with vasculitis, and the exact nature of the cellular infiltrate also contributes to arriving at a proper diagnosis. There is no question that these overall categorizations help to narrow the differential in any given case, but at times there are overlapping features or reaction patterns that do not allow for a specific diagnosis. Clinical-pathologic correlation is important, as emphasized by a published review that has an expanded and useful classification of panniculitis.79

There are anecdotes about the efficacy of combination antimalarials such as hydroxychloroquine and quinacrine in treating subcutaneous sarcoid, but no studies exist to allow definitive recommendations. Reports about the use of newer therapies such as mycophenolate mofetil and thalidomide for treatment of inflammatory causes of panniculitis such as nodular panniculitis and erythema nodosum are already in the literature and indicate that these drugs will likely evolve to be useful for these conditions.80,81

### Table 43-1 Classification of Panniculitis

<table>
<thead>
<tr>
<th>I. Without prominent vasculitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Septal inflammation</td>
</tr>
<tr>
<td>1. Lymphocytic: lymphocytic vasculitis, periarteritis nodosa, thrombophlebitis, ENL</td>
</tr>
<tr>
<td>2. Granulomatous: palisaded granulomatous diseases, sarcoidosis, tuberculosis, syphilis</td>
</tr>
<tr>
<td>3. Sclerotic: scleroderma, eosinophilic fasciitis, lipodermatosclerosis, toxins</td>
</tr>
<tr>
<td>B. Lobular inflammation</td>
</tr>
<tr>
<td>1. Neutrophilic: infection, rupture of folliculitis and cysts, pancreatic fat necrosis</td>
</tr>
<tr>
<td>2. Lymphocytic: lupus panniculitis, poststeroid panniculitis, lymphoma/leukemia</td>
</tr>
<tr>
<td>3. Macrophagic: histiocytic cytophagic panniculitis</td>
</tr>
<tr>
<td>4. Granulomatous: erythema induratum/nodular vasculitis, palpated granulomatous diseases, sarcoidosis, Crohn’s disease</td>
</tr>
<tr>
<td>5. Mixed inflammation with many foam cells: α1-antitrypsin deficiency, Weber-Christian disease, traumatic fat necrosis</td>
</tr>
<tr>
<td>6. Eosinophilic: eosinophilic panniculitis, arthropod bites, parasites</td>
</tr>
<tr>
<td>7. Enzymatic fat necrosis: pancreatic enzyme panniculitis</td>
</tr>
<tr>
<td>8. Crystal deposits: subcutaneous fat necrosis of the newborn, gout, oxalosis</td>
</tr>
<tr>
<td>9. Embryonic fat pattern: lipomatosis, lipodystrophy</td>
</tr>
<tr>
<td>II. With prominent vasculitis (septal or lobular)</td>
</tr>
<tr>
<td>A. Neutrophilic: leukocytoclastic vasculitis, subcutaneous polyarteritis nodosa, thrombophlebitis, ENL</td>
</tr>
<tr>
<td>B. Lymphocytic: nodular vasculitis, perniosis, angiocentric lymphomas</td>
</tr>
<tr>
<td>C. Granulomatous: nodular vasculitis/erythema induratum, ENL, granulomatosis with polyangiitis, Churg-Strauss allergic granulomatosis</td>
</tr>
<tr>
<td>III. Mixed patterns</td>
</tr>
</tbody>
</table>

ENL, erythema nodosum leprosum.

Efficacy of these and more established drugs such as NSAIDs, antimalarials, and methotrexate need to be studied, and outcomes will hopefully be more systematically evaluated.

### RELAPSING POLYCHONDritis

The diagnosis of relapsing polychondritis (RP) is based on the typical clinical manifestations, with auricular findings seen in 90% of patients. Nasal and respiratory tract chondritis can occur, along with nonerosive inflammatory arthritis, cardiac valvular insufficiency, vasculitis, eye, and audiovestibular involvement. The estimated prevalence of 3.5 per million makes controlled trials nearly impossible. The etiology is unknown, but the pathogenesis appears to be mediated by an immune reaction to type II collagen. Clinical skin manifestations include inflammation of the ear, with sparing of the earlobe. Diagnosis includes the presence of a positive serum antibody test to type II collagen and a wedge biopsy that shows cartilage necrosis and periarticular inflammation with lymphocytes and histiocytes. Involvement of other cartilage areas including the upper airway should be assessed. Glucocorticoids are the therapeutic choice for reducing the inflammatory process in patients with RP. For patients with sustained disease, many immunosuppressive drugs have been used as steroid-sparing agents. There have been reports of response to tumor necrosis factor (TNF) inhibitors in patients otherwise refractory to therapy.82
INFLAMMATORY AND INFILTRATIVE DISEASES AND SKIN/ARTHRITIS

Amyloid

Type AL amyloidosis (primary amyloidosis) is rare, with an incidence of less than 1 per 100,000. Skin lesions may occur in up to 40% of these patients. Skin lesions can be an early sign of the disease and include purpura, petechiae, and ecchymoses due to infiltration of blood vessels by amyloid. Other skin findings include alopecia, plaques, and nodules, often found on flexor surfaces, the face, or the buccal mucosa. Bullae and nail dystrophy are occasionally seen. Diagnosis is confirmed by biopsy of lesional or nonlesional skin, along with urine and serum for immunoelectrophoresis to confirm the presence of a circulating monoclonal protein. Skin biopsy shows Congo-red positive, homogeneous, hyaline, fibrillary deposits. Treatment includes autologous stem cell transplantation, with approximately 50% of patients achieving prolonged remission with such therapy. Other effective therapies include the combination of melphalan with high-dose dexamethasone or the use of thalidomide.83 The prognosis depends on the stage at the time of diagnosis, emphasizing the importance of recognizing the disease.

Sarcoïdosis

Cutaneous involvement occurs in 20% to 25% of sarcoidosis cases and is most likely to be seen early in the disease. Cutaneous lesions can be classified as nonspecific, typically erythema nodosum, or specific granulomatous. Erythema nodosum occurs frequently as part of Lofgren’s syndrome, with bilateral hilar lymphadenopathy and acute iridocyclitis. This variant has a good prognosis and resolves in 80% within 2 years. The skin lesions of sarcoïdosis generally have no prognostic significance or correlation with disease activity. Skin involvement has no effect on the course of the disease, and the number of skin lesions does not correlate with systemic disease. Skin plaques tend to be more persistent and commonly associated with chronic forms of the disease. Lupus pernio (Figure 43-15), with violaceous plaques on the nose, ears, cheeks, lips, and fingers, is often seen in long-standing sarcoïdosis and is associated with upper airway involvement and pulmonary fibrosis.84 Other forms of cutaneous sarcoïdosis include papules, follicular papules, subcutaneous nodules, ulcerative lesions, alopecia, and ichthyosis. Cutaneous sarcoïd can arise in scars. Because of the many types of presentation, diagnosis can be challenging and a skin biopsy is necessary to confirm the clinical suspicion. Mimickers of papules include xanthelasma, rosacea, trichoepithelioma, syphilis, LE, and granuloma annulare. Plaques can resemble lupus vulgaris, necrobiosis lipoidica, morphea, leprosy, Leishmania, or lupus erythematosus. Nodules can resemble lymphoma or other types of panniculitis. Treatment of cutaneous sarcoïdosis depends on the degree of systemic involvement. Clearly patients who need prednisone for systemic disease often experience improvement of their cutaneous sarcoïd. Patients with isolated skin disease or systemic disease not requiring aggressive therapy can benefit from topical or intralesional corticosteroids, topical tacrolimus, hydroxychloroquine, combination antimalarials with hydroxychloroquine and quinacrine, or chloroquine. If antimalarials are not adequate, then methotrexate or oral retinoids may be used. There have been case reports, and small case series of successful therapy with thalidomide and TNF inhibitors, as well as laser remodeling of lupus pernio.85,86 There is some concern that interferon-α therapy may induce sarcoïdosis.

MISCELLANEOUS SKIN DISEASES AND ARTHRITIS

Behçet’s Disease

The criteria for Behçet’s disease (BD) include recurrent oral and genital ulcers, eye lesions (uveitis or retinal vasculitis), characteristic skin lesions, and a positive pathergy test.87 The pathergy test involves using a sterile needle to prick the forearm. The results are positive when the puncture causes an aseptic erythematous nodule or pustule that is more than 2 mm in diameter at 24 to 48 hours. A diagnosis is made if patients have recurrent oral ulceration plus at least two of the other findings without other clinical explanations. Skin lesions include erythema nodosum, pseudofolliculitis, or papulopustular lesions or acneiform nodules in postadolescents. Oral ulcers are painful and occur on the gingiva, tongue, and buccal and labial mucosa. Genital ulcers, usually larger and deeper than oral ulcers, are typically on the scrotum and penis in men and the vulva in women. Venous involvement including superficial thrombophlebitis and deep venous thrombosis can occur. On skin biopsy, small vessel vasculitis is common. Ulcer treatment includes topical corticosteroids, colchicine, and thalidomide. Systemic corticosteroids are prescribed for unresponsive erythema nodosum.
Familial Mediterranean Fever

Familial Mediterranean fever (FMF) is an autosomal recessive disease that tends to affect certain ethnic groups including Sephardic Jews, Arabs, Armenians, and Turks. There is a mutation on the short arm of chromosome 16, and the mutant protein pyrin likely plays an inhibitory role in the control of inflammation.\cite{88,89} It is characterized by recurrent, self-limited attacks of peritonitis, pleuritis, and synovitis. Erysipelas-like erythema (ELE) is the pathognomonic skin manifestation. This is characterized by tender erythematous, well-demarcated plaques, usually located on the lower legs.\cite{90} They may be triggered by physical effort and subside spontaneously within 48 to 72 hours of bed rest. Fever and leukocytosis may accompany this condition. Other associated skin findings include Henoch-Schönlein purpura; non-specific purpura; erythema of the face, trunk, or palm; angioneurotic edema; Raynaud’s phenomenon; pyoderma; and subcutaneous nodules. Secondary generalized amyloidosis may lead to chronic renal failure and death if not recognized. Skin biopsy shows edema of the superficial dermis and sparse perivascular infiltrate composed of a few lymphocytes, neutrophils, and nuclear dust, without vasculitis. Direct immunofluorescence shows deposits of C3 in the wall of small superficial vessels. Early treatment with colchicine, which prevents or diminishes the frequency and severity of the inflammatory episodes, can be beneficial.

Multicentric Reticulohistiocytosis

Multicentric reticulohistiocytosis (MRH) is a rare condition of unknown etiology that most frequently occurs in Caucasian women in their fifth and sixth decades. There is destructive symmetric arthritis, with arthritis mutilans developing in about 45% of cases, associated with cutaneous papulonodular lesions. Skin findings include cutaneous red-to-brown papules or nodules, typically on the face, dorsum of the fingers, and over the proximal and distal interphalangeal joints, but they can be in a more generalized distribution. A rarer presentation includes photodistributed erythema, often with targeting over joints that masquerades as DM.\cite{91} Diagnosis is made by skin biopsy, which shows infiltration of histiocytes and multinucleated giant cells. These changes can be seen in a variety of tissues including the heart, lungs, skeletal muscle, and gastrointestinal tract. The differential diagnosis of skin disease includes other infiltrative processes such as sarcoidosis and even leprosy. Treatment recommendations, based on mostly small case reports, include glucocorticoids and methotrexate, with use of cyclophosphamide and finally chlorambucil if the condition is unresponsive.\cite{92,93} Cyclosporine, TNF inhibitors, and biphosphonates have also been used with reported benefit.\cite{94,95}

In about one-third of patients, MRH may precede or follow an underlying malignancy. Reported associated malignancies include breast, cervix, colon, stomach, lung, larynx, ovary, lymphoma, leukemia, sarcoma, melanoma, mesothelioma, and metastatic cancer of unknown primary.

Chronic Infantile Neurologic Cutaneous and Articular (CINCA) Syndrome

Mutations in cryopyrin (CIAS1, NALP3, PYPAF1), associated with increased proinflammatory cytokines, have been found in about 50% of patients with CINCA syndrome.\cite{96,97} CINCA syndrome is characterized by a neonatal urticarial eruption, fever, arthritis, and leukocytosis. Other findings include chronic meningitis, papilledema, hearing loss, and growth retardation. Isolated reports suggest that TNF inhibitors, anakinra, and thalidomide can be beneficial therapeutically in these patients.

Acknowledgment

The authors want to acknowledge Dr. Nicole Rogers for her assistance with the figures. The work was supported by the Department of Veterans Affairs Veterans Health Administration, Office of Research Development, Biomedical Laboratory Research and Development and by the National Institutes of Health (NIH K24-AR 02207) to Victoria P. Werth.

References


The references for this chapter can also be found on www.expertconsult.com.
The Eye and Rheumatic Diseases

JAMES T. ROSENBAUM

KEY POINTS

Symptoms of uveitis vary widely based on the location of the inflammation within the eye and the suddenness of onset.

Ankylosing spondylitis is the systemic disease most often associated with uveitis in North America and Europe.

During a lifetime, about 40% of patients with ankylosing spondylitis develop acute anterior uveitis.

The uveitis associated with HLA-B27 tends to be unilateral, recurrent, and sudden in onset. Recurrences sometimes affect the opposite eye.

Sarcoidosis frequently manifests as a uveitis.

Most patients with retinal vasculitis do not have a systemic vasculitis.

Many patients with scleritis have a systemic disease, such as rheumatoid arthritis.

Antineutrophilic cytoplasmic antibody testing helps to identify a subset of patients with severe scleritis.

Granulomatosis with polyangiitis (formerly Wegener’s granulomatosis) is the rheumatic disease that most frequently involves the orbit.

Anterior ischemic optic neuropathy is the most common ocular manifestation of temporal arteritis.

Most patients with visual loss secondary to optic nerve ischemia do not have arteritis.

Virtually all of the systemic inflammatory diseases that require rheumatologic care tend to affect the eye or its surrounding structures. Table 44-1 presents the prototypic ocular manifestations of rheumatoid arthritis, systemic lupus erythematosus, Sjögren’s syndrome, spondyloarthropathies, vasculitides including granulomatosis with polyangiitis (formerly Wegener’s granulomatosis) and temporal arteritis, scleroderma, Behçet’s syndrome, relapsing polychondritis, and dermatomyositis. Each of these diseases is addressed elsewhere in this text; this chapter focuses on specific ocular structures—the uvea, cornea, orbit, and optic nerve—and illustrates how inflammation of each might relate to an autoimmune or inflammatory process.

**OCULAR ANATOMY AND PHYSIOLOGY**

A diagram of the eye is shown in Figure 44-1. The eye is a tiny, but elegantly complex structure. The anterior segment of the eye includes the cornea, which is avascular and transparent when healthy. The lens also is an avascular structure. The anterior chamber is filled with aqueous humor, which has homology to cerebrospinal fluid. When the blood-aqueous barrier is intact, the aqueous humor contains no leukocytes and very little protein. The blood-aqueous barrier, which resembles the blood-synovial barrier, is disrupted in anterior uveitis. In this case, a routine, noninvasive biomicroscopic or slit lamp examination would reveal leukocytes and increased protein in the anterior chamber. An ophthalmologist has the opportunity to observe two universal hallmarks of inflammation noninvasively.

The term uvea derives from the Latin word for “grape.” The anterior uvea includes the iris and the ciliary body. The aqueous humor is synthesized by the ciliary body. The posterior portion of the uvea is the choroid, which is a highly vascular tissue just posterior to the retina. Any portion of the uveal tract could become inflamed; adjacent tissue also is frequently inflamed. Anatomic subsets of uveitis include anterior uveitis, which consists of iritis or iridocyclitis (ciliary body inflammation); intermediate uveitis, in which leukocytes are present within the vitreous humor; and posterior uveitis, in which the choroid and the retina are inflamed. A panuveitis occurs when all portions of the uveal tract are inflamed. An attempt has been made to standardize the nomenclature used to describe uveitis by the Standardization of Uveitis Nomenclature Working Group, although ambiguities persist because not all ophthalmologists follow these definitions as yet.

Signs and symptoms of uveitis depend on the portion of the uveal tract that is affected. An anterior uveitis, especially if it begins suddenly, is associated with redness, pain, and photophobia. Visual loss varies and often is due to macular edema if present (Figures 44-2 and 44-3). An intermediate uveitis usually causes floaters owing to leukocytes that enter the visual axis, although most floaters are due to aging or other changes within the vitreous humor. A posterior uveitis by itself does not usually produce pain or redness. Visual loss depends on the location and extent of the inflammatory process.

The outer tunic of the eye is known as the sclera. At the front of the eye, the sclera meets the cornea at a tissue known as the limbus. The most interior layer of the eye is an extension of the brain that responds to visual signals, the retina. The eye shares some common features with the joint, including the presence of hyaluronic acid primarily in the vitreous humor and the presence of type II collagen, although ocular inflammation is not a reported accompaniment of collagen-induced arthritis. Aggrecan is a proteoglycan present in both the eye and the joint. An autoimmune response to aggrecan in BALB/c mice can produce both arthritis and uveitis.
Table 44-1  Most Characteristic Ocular Findings of Selected Rheumatic Diseases

<table>
<thead>
<tr>
<th>Disease</th>
<th>Most Characteristic Eye Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatoid arthritis</td>
<td>Sicca</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>Sicca, Scleritis</td>
</tr>
<tr>
<td>Sjögren’s syndrome</td>
<td>Sicca, Cotton-wool spots</td>
</tr>
<tr>
<td>Spondyloarthritis</td>
<td>Sicca, Acute anterior uveitis</td>
</tr>
<tr>
<td>Granulomatosis with polyangiitis</td>
<td>Scleritis, Orbital inflammation</td>
</tr>
<tr>
<td>Temporal arteritis</td>
<td>Anterior ischemic optic neuropathy</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>Sicca</td>
</tr>
<tr>
<td>Behçet’s disease</td>
<td>Uveitis, retinal arteritis</td>
</tr>
<tr>
<td>Relapsing polychondritis</td>
<td>Scleritis, episcleritis, uveitis</td>
</tr>
<tr>
<td>Dermatomyositis</td>
<td>Heliotrope eyelids</td>
</tr>
</tbody>
</table>

**OCULAR IMMUNE RESPONSE**

The eye generally is regarded as an immune privileged site.\(^3\) From a teleologic perspective, many scientists believe that the eye has evolved mechanisms to avoid becoming inflamed because of the consequences this has for visual acuity. Similar to the brain, the internal portion of the eye has no lymphatics, although the conjunctiva on the ocular surface has lymphatic drainage. Portions of the eye—the cornea and the lens—are avascular. The aqueous humor contains several factors that are known to be immunosuppressive, including transforming growth factor-β and α-melanocyte-stimulating hormone. Several tissues within the eye express ligands that promote apoptosis, including tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) and Fas ligand. If a soluble antigen is injected into the anterior chamber, a cellular immune response is suppressed. This phenomenon is known as anterior chamber–associated immune deviation (ACAID). These factors are important to consider in the effort to understand why the eye sometimes is targeted as part of an immune or inflammatory disease.

**UVEITIS**

Rheumatologists may be consulted to identify a systemic disease in a patient with uveitis, and a rheumatologist often is asked to assist in the management of immunosuppression in selected patients with uveitis. In some referral practices for patients with uveitis, 40% of patients might have an associated systemic illness. Table 44-2 lists the differential diagnoses of uveitis. The immunologic diseases most likely to be associated with uveitis are listed in Table 44-3.

The most common systemic illness associated with uveitis in most North American practices is ankylosing spondylitis. From an epidemiologic perspective, anterior uveitis is more common than posterior or intermediate uveitis.\(^4\) About 50% of individuals who develop an anterior uveitis are HLA-B27 positive.\(^5\) The uveitis associated with HLA-B27 is almost always unilateral, is recurrent, is of relatively short duration (<3 months per attack), resolves completely between attacks, and is associated with reduced intraocular pressure (in contrast to herpes simplex, which can cause recurrent anterior uveitis associated with increased intraocular pressure).\(^6\) Hypopyon or pus in the anterior chamber sometimes is present in patients with HLA-B27–associated uveitis (Figure 44-4). Recurrent episodes can affect the contralateral eye, but simultaneous bilateral
A retinal vasculitis can be a prominent feature of sarcoidosis, even though systemic vasculitis is not a typical feature of the disease; this results in part from the manner in which vasculitis is diagnosed in the retina. Histologic evidence of vessel wall destruction is rarely obtained because of the morbidity of a retinal biopsy. Instead, retinal vasculitis is diagnosed on the basis of perivascular sheathing along a vessel as seen on funduscopic examination (Figure 44-6), intraretinal hemorrhages that must be secondary to vascular injury, and fluorescein angiography indicating increased vascular permeability.

The term retinal vasculitis is misleading to most rheumatologists because the classic systemic vasculitides, such as polyarteritis nodosa and granulomatosis with polyangiitis, rarely are associated with retinal vasculitis.

Sarcoidosis frequently manifests initially as an ocular problem. An ocular symptom is the initial manifestation almost as frequently as a lung symptom. Sarcoidosis frequently involves the conjunctiva, which is an accessible tissue for biopsy confirmation of the diagnosis. In most series involvement is rare. Many studies have tried to address the question of how frequently a patient with HLA-B27–associated anterior uveitis has an associated spondyloarthropathy. A wide range of answers have been suggested, and the percentage depends on the definition of spondyloarthropathy, but one reasonable estimate is that 80% of HLA-B27 positive patients with acute anterior, unilateral uveitis have associated spondyloarthropathy.5

The uveitis associated with reactive arthritis is indistinguishable from the uveitis associated with ankylosing spondylitis. In either of these entities, about 40% of patients develop acute anterior uveitis during a lifetime. Although conjunctivitis is part of the classic triad of reactive arthritis (in association with arthritis and nongonococcal urethritis), conjunctivitis is uncommon. A genome-wide screen for susceptibility genes for acute anterior uveitis identified loci that predispose to ankylosing spondylitis and loci that seem to be unassociated with susceptibility to ankylosing spondylitis.6

Approximately 5% of patients with inflammatory bowel disease and 7% of those with psoriatic arthritis develop uveitis. Although some of these patients have disease that is unilateral, anterior, and recurrent, many have disease that is bilateral, chronic in duration, and posterior to the lens.5,6 About half of all patients with Crohn’s disease or psoriatic arthritis and uveitis are HLA-B27 positive.

Sarcoidosis is the second most common systemic disease associated with uveitis, at least in North America, and in some geographic areas it might be more common than spondyloarthritis. Sarcoidosis is promiscuous within the eye, meaning that it can affect a wide range of structures, including the orbit, lacrimal gland, anterior uvea, vitreous humor, choroid, retina, or optic nerve. Ocular inflammation with sarcoidosis frequently is termed granulomatous because large collections of cells deposit on the back of the cornea (Figure 44-5). A retinal vasculitis can be a prominent feature of sarcoidosis, even though systemic vasculitis is not a typical feature of the disease; this results in part from the manner in which vasculitis is diagnosed in the retina.

Histologic evidence of vessel wall destruction is rarely obtained because of the morbidity of a retinal biopsy. Instead, retinal vasculitis is diagnosed on the basis of perivascular sheathing along a vessel as seen on funduscopic examination (Figure 44-6), intraretinal hemorrhages that must be secondary to vascular injury, and fluorescein angiography indicating increased vascular permeability.7 The term retinal vasculitis is misleading to most rheumatologists because the classic systemic vasculitides, such as polyarteritis nodosa and granulomatosis with polyangiitis, rarely are associated with retinal vasculitis.

Sarcoidosis frequently manifests initially as an ocular problem.8 An ocular symptom is the initial manifestation almost as frequently as a lung symptom. Sarcoidosis frequently involves the conjunctiva, which is an accessible tissue for biopsy confirmation of the diagnosis. In most series

### Table 44-3 Immune-Mediated Diseases Most Often Associated with Uveitis

<table>
<thead>
<tr>
<th>Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ankylosing spondylitis</td>
</tr>
<tr>
<td>Behçet’s disease</td>
</tr>
<tr>
<td>Drug/hypersensitivity reactions</td>
</tr>
<tr>
<td>Familial granulomatous synovitis</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>Interstitial nephritis</td>
</tr>
<tr>
<td>Juvenile idiopathic arthritis</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>Neonatal-onset multisystem inflammatory disease</td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
</tr>
<tr>
<td>Reactive arthritis</td>
</tr>
<tr>
<td>Sarcoidosis</td>
</tr>
<tr>
<td>Sweet’s syndrome</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>Vasculitis, especially Cogan’s syndrome and Kawasaki’s disease</td>
</tr>
<tr>
<td>Vogt-Koyanagi-Harada syndrome</td>
</tr>
</tbody>
</table>

### Table 44-2 Differential Diagnosis of Uveitis

<table>
<thead>
<tr>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infections—toxoplasmosis, syphilis, herpes simplex, herpes zoster, and cytomegalovirus</td>
</tr>
<tr>
<td>Systemic, immune-mediated diseases</td>
</tr>
<tr>
<td>Masquerade syndromes such as lymphoma</td>
</tr>
<tr>
<td>Syndromes confined to the eye, such as pars planitis, birdshot chorioretinopathy, and serpiginous choroiditis</td>
</tr>
</tbody>
</table>

Figure 44-4 The creamy material at the bottom of the pupil is an accumulation of leukocytes, also known as a hypopyon.

Figure 44-5 Keratic precipitates. The white dots result from concretions of cells depositing against the corneal endothelium. These keratic precipitates are large and usually are described as granulomatous even though a granuloma is not present histologically.
prognosis with Behçet’s syndrome can be grim, and blindness is a frequent concomitant of untreated ocular disease.

Relapsing polychondritis can have an impact on almost any portion of the eye, including the episclera, sclera, and uveal tract. Ocular inflammation is common.

Cogan’s syndrome is classically defined as sensorineural hearing loss with corneal disease, especially an interstitial keratitis. This definition usually is broadened to include any ocular inflammatory process, such as uveitis or scleritis. Although uveitis can occur with polyarteritis or granulomatosis with polyangiitis, scleral disease is typical. In contrast, anterior uveitis in association with conjunctivitis is present in most patients with Kawasaki’s disease.

Uveitis and arthritis occasionally can result from an infection such as Whipple’s disease or Lyme disease. Uveal involvement with Lyme disease has been described but is extremely rare.

Some autoinflammatory diseases are associated with uveitis. Autoinflammatory diseases are characterized by widespread inflammation in the absence of detectable autoantibodies. Many autoinflammatory syndromes respond dramatically to inhibition of interleukin-1. Blau’s syndrome, which also is known as familial granulomatous synovitis, results from a single base change in the nucleotide-binding domain of the CARD15 gene, which is also known as NOD2 or NLRC2. Polymorphisms elsewhere in this same gene predispose to Crohn’s disease. Blau’s syndrome is characterized by childhood onset of uveitis, arthritis, and dermatitis. Inflammation in additional organ systems also has been described. The disease is autosomal dominant. The histopathology of affected skin or joint can show noncaseating granuloma as in sarcoidosis. Lung involvement has not been described in Blau’s syndrome, however. Many patients thought to have so-called early-onset sarcoid have been shown by gene sequencing to have new mutations in the NOD2 gene.

Neonatal-onset multisystem inflammatory disease (NOMID), which also is known as chronic infantile neurologic cutaneous articular syndrome (CINCA), is an autosomal dominant autoinflammatory syndrome. Ocular involvement in NOMID is more variable than in Blau’s syndrome. Characteristic findings include papilledema and uveitis.
Treatment of uveitis depends on multiple factors, such as severity, location within the eye, patient preference, and the specific diagnosis (e.g., Behçet’s syndrome might be especially responsive to infliximab or interferon alfa). For noninfectious causes of uveitis that involve the anterior portion of the eye, treatment usually begins with topical corticosteroids and often dilating drops to prevent posterior synechiae and to relieve spasm of the ciliary muscle. Periocular or intraocular corticosteroid injections, usually with triamcinolone, are given for inflammation posterior to the lens that is not responding to topical medication. Local corticosteroids can increase intraocular pressure, induce cataracts, interfere with the response to infection, and delay wound healing. A long-lasting corticosteroid implant containing fluorocinolone has been approved by the Food and Drug Administration. All patients who elect this type of therapy develop a cataract if the lens has not already been surgically removed, and most patients develop glaucoma, many requiring surgery to control the intraocular pressure.

Systemic immunosuppressive therapy generally is reserved for patients with active, noninfectious causes of inflammation. For systemic immunosuppression to be indicated, usually the inflammation is bilateral and severe enough to interfere with activities of daily living. A variety of immunomodulatory medications have been tried to treat intraocular inflammation, including azathioprine, chlorambucil, cyclophosphamide, cyclosporine, dactinumab, infliximab, methotrexate, mycophenolate mofetil, and tacrolimus. The optimal choice depends on many factors, not the least of which is the empiric result of any therapeutic approach. Maintaining a treatment for uveitis usually requires some efficacy in association with good tolerability. In this regard, at least one study found methotrexate to be superior to other antimetabolites. Because of the range of diseases being treated and the range of clinical response, there is room for all of these options in the therapeutic armamentarium of a uveitis clinic.

Biologic therapies are being used increasingly to treat patients with uveitis when the disease has proven refractory to other forms of immunomodulatory therapy. Infliximab is especially useful in the treatment of Behçet’s disease. One prospective study that assessed infliximab for a variety of forms of uveitis concluded that the drug was often effective, but that it was also surprisingly toxic. Consequently, the role of biologic therapies in most specific forms of uveitis is inadequately studied.

**SCLERITIS AND CORNEAL MELT**

Scleritis often is divided into five categories: diffuse anterior, nodular, necrotizing, scleromalacia perforans, and posterior (Figure 44-8). Each of the first three categories results in a red, painful eye. Pain is more variable in scleromalacia perforans, in which a nodule pathologically similar to a rheumatoid nodule forms in the sclera (Figure 44-9). Pain also varies with posterior scleritis, and because the sclera extends back to the optic nerve, posterior scleritis can occur in a localized fashion such that the eye is not red. Because of the risk of perforation, the sclera is not normally biopsied, but biopsy studies have shown that scleritis is often a granulomatous inflammation of scleral tissue.

Patients with scleritis can develop complications within the eye, including uveitis, glaucoma, optic nerve edema, and retinal or choroidal distortion. A corneal melt or peripheral thinning of the cornea sometimes develops in individuals with severe scleritis and represents a potentially blinding complication of the disease (Figure 44-10).

About 50% of patients with scleritis have an associated systemic illness. The most common of such illnesses are limited granulomatosis with polyangiitis and rheumatoid arthritis. Generally, the associated rheumatoid arthritis is long-standing and seropositive. Patients may have associated nodules, vasculitis, or pleuroppericarditis. They have a shortened life expectancy compared with other patients with rheumatoid arthritis. It is unusual for scleritis to be an initial manifestation of rheumatoid arthritis. Granulomatosis with polyangiitis is commonly associated with scleritis. In contrast to rheumatoid arthritis, scleritis can be the initial manifestation of granulomatosis with polyangiitis, and our routine is to obtain antineutrophilic cytoplasmic antibody serology on any patient who presents with scleritis without an obvious systemic disease association. Other systemic associations with scleritis include...
Corneal melt. The white light of the slit lamp beam narrows over the peripheral cornea, where the tissue is thin.

inflammatory bowel disease, relapsing polychondritis, other vasculitides such as temporal arteritis, and ankylosing spondylitis. Infections are a rare but possible cause of scleritis. Tophaceous gout also has been reported as an unusual cause of scleritis.

Scleritis tends to be a painful and persistent disease that often lasts for years. In contrast, episcleritis involves more superficial tissue and is usually transient. Episcleritis may be a feature of rheumatoid arthritis, although many patients with episcleritis may not have any associated systemic illness. Complications within the eye, such as glaucoma or uveitis, are absent. Mild discomfort, rather than frank pain, is the usual presenting symptom. In contrast to scleritis, patients with episcleritis have vessels that constrict completely after 2.5% phenylephrine (Neo-Synephrine) is placed on the surface of the eye.

Some patients with scleritis, especially those who do not have an associated systemic illness, are treated adequately with oral nonsteroidal anti-inflammatory drug. Some experts treat scleritis with locally injected corticosteroids, but this should be avoided if the sclera is thin (necrotizing disease). In addition, corticosteroid has the theoretical risk of promoting thinning. The usual option for patients who do not respond to nonsteroidal anti-inflammatory drugs is oral prednisone. Some patients can be maintained on low doses of prednisone, but many require the addition of an antimetabolite as a steroid-sparing drug. A substantial subset of patients with scleritis, especially patients who are positive for antineutrophilic cytoplasmic antibody, is best managed with an alkylating agent, such as cyclophosphamide. Because of the risk of this therapeutic approach, this drug is often given until the disease is in remission for several months; then therapy is switched to an antimetabolite to maintain the disease-free state.

**ORBITAL DISEASE**

Graves’ disease, the most common orbital inflammatory disease, generally results in an orbital myositis that can be identified on imaging such as CT, ultrasound, or magnetic resonance imaging (MRI). From a rheumatologic perspective, granulomatosis with polyangiitis is the disease that most commonly affects the orbit. The inflammation can be extremely painful and may result in blindness. Orbital inflammation sometimes is more recalcitrant to therapy than other aspects of granulomatosis with polyangiitis. A small series suggested that rituximab may be efficacious in patients with granulomatosis with polyangiitis, including those with orbital involvement.29

Orbital pseudotumor, or nonspecific orbital inflammatory disease, is a diagnosis of exclusion that is made on the basis of objective orbital swelling as documented by imaging and a biopsy showing an inflammatory process that cannot be ascribed to another process such as Graves’ disease. A biopsy specimen of the orbit is not always obtained, but it can be useful in ruling out lymphoma or a metastatic malignancy as the cause of the proptosis. Methotrexate is a therapeutic option to treat nonspecific orbital inflammation.30

Other systemic diseases that commonly affect the orbit include sarcoidosis.

**OPTIC NEURITIS**

Optic nerve disease can result from many insults, including toxins (some of which are medications), vascular insufficiency as occurs from atherosclerotic disease or giant cell arteritis, and immunologic attack. The immune-mediated disease that most commonly affects the optic nerve is multiple sclerosis. This demyelinating condition generally starts suddenly in one eye with pain, an afferent pupillary defect, loss of color vision, and visual field loss typical of optic nerve disease. Initially, the optic nerve may show papilledema, or it may appear normal if the inflammation is retrobulbar. Over several weeks, the affected nerve usually becomes pale.

Demyelinating disease affecting the optic nerve generally is not treated by a rheumatologist, but a rare patient with optic nerve disease has inflammation that might require long-term immunosuppression. These patients carry a diagnosis labeled variously as autoimmune optic neuropathy or sometimes steroid-sensitive optic neuropathy. This diagnosis is clinically distinct from optic neuritis associated with multiple sclerosis in that MRI of the head should not indicate a demyelinating process, the disease is often bilateral, the kinetics of the inflammation is different from that of multiple sclerosis, and the disease usually responds to oral corticosteroids. In most centers, a neuro-ophthalmologist would be involved in establishing this diagnosis. Systemic lupus erythematosus and sarcoidosis may affect the optic nerve in this way, but many patients with this diagnosis do not have an associated systemic illness. Alkylating agent therapy or an antimetabolite can be beneficial for many patients with this entity.11

Sudden blindness is arguably the most feared consequence of temporal arteritis. This disease is characterized by granulomatous inflammation of multiple vessels above the waist. These frequently include the temporal artery and the posterior ciliary arteries. Inflammation in these latter vessels leads to anterior ischemic optic neuropathy (AION), which is ischemia of the optic nerve that manifests as sudden visual loss (Figure 44-11). Temporal arteritis also can affect the central retinal artery; this may result in blindness. In this condition, the funduscopic appearance of the eye shows markedly reduced arteriolar flow and a cherry-red spot in
the macula. Temporal arteritis can cause diplopia by affecting circulation to extraocular muscles.

The visual loss associated with temporal arteritis is frequently labeled arteritic AION to distinguish it from the more common nonarteritic AION, which usually is attributable to small vessel atherosclerosis. Patients with arteritic AION typically are older than 50 years and have an erythrocyte sedimentation rate greater than 50 mm/hr. Many patients with arteritic AION have associated symptoms of polymyalgia rheumatica, jaw claudication, scalp tenderness, or temporal artery tenderness. The biopsy specimen of the temporal artery shows vasculitis in about 80% of patients with temporal arteritis if an adequate length of vessel is sampled. Twenty percent of patients with temporal arteritis might have a biopsy specimen of the artery that is negative because it has spared this vessel or has affected it in a sufficiently patchy distribution such that the biopsy specimen did not reveal the pathology. Patients with nonarteritic AION tend to have small optic nerve cups.

**SUMMARY**

In some ways, from a rheumatologist’s perspective, the eye is a microcosm of the body. Its complex structures frequently reflect inflammation elsewhere. Treatment of many forms of ocular inflammation requires collaboration between a rheumatologist and an ophthalmologist.

**References**


The references for this chapter can also be found on www.expertconsult.com.
Neck Pain

JOSEPH S. CHENG • MATTHEW J. McGIRT • CLINT DEVIN

trauma, degenerative changes, infection, and autoimmune disorders such as rheumatoid arthritis and ankylosing spondylitis.

The perception and resultant reporting of this pain vary significantly according to cultural and social circumstances. Honeyman and Jacobs noted that Australian Aborigines significantly underreport pain and are rarely disabled by pain. Social circumstances also play an important role in an individual’s ability to cope with and overcome neck pain. Studies have demonstrated worse outcomes following discectomy for patients with a workers’ compensation claim or litigation surrounding their condition.

These studies indicate nonorganic contributions to neck pain for secondary gain. Fortunately, most episodes of acute neck pain resolve with “tincture of time” and patient education.

Physicians need to be able to differentiate causes of neck pain that can be managed conservatively from causes that require more aggressive treatments. An understanding of anatomy and physiology and of their association with the pathogenesis of neck pain provides the basis for obtaining a thorough history, physical examination findings, and ancillary data with the ultimate goal of effective treatment.

ANATOMY

The cervical spine consists of seven vertebrae (C1 through C7). The bony anatomy of the atlas (C1) and axis (C2) is unique, whereas C3-C7 demonstrates fairly consistent anatomy (Figure 45-2). The atlas has no vertebral body, and its lateral masses articulate with the occipital condyles of the skull, forming the atlanto-occipital joints supported by anterior and posterior occipital membranes. The atlanto-occipital joint is responsible for approximately 50% of total flexion and extension in the neck, and functional implications are clear when this motion is lost. The axis (C2) has the dens or an odontoid peg that projects upward and anterior to articulate with the posterior aspect of the anterior arch of the atlas. The principal stabilizer of the odontoid to the anterior arch of the atlas is the transverse ligament; alar and apical ligaments act as secondary stabilizers. This true synovial joint is susceptible to inflammatory processes such as those seen with rheumatoid arthritis. No intervertebral...
The subaxial cervical spine consists of C3 through C7 vertebrae, all demonstrating fairly similar anatomy. Each vertebra consists of a body, two interconnecting pedicles, two lateral masses, two transverse processes, two laminae, and spinous processes. Transverse and spinous processes

The atlantoaxial articulation provides approximately 50% of rotatory motion of the cervical spine.

disk is present between the atlanto-occipital joint and the atlantoaxial joint, and without conferred stability from a disk, destructive inflammatory arthritides may result in instability. The atlantoaxial articulation provides approximately 50% of rotatory motion of the cervical spine.

**Figure 45-1** Algorithm of neck pain. CT, computed tomography; EMG, electromyogram; MRI, magnetic resonance imaging; NCS, nerve conduction study; PT, physical therapy; SSEP, somatosensory evoked potentials.

**Figure 45-2** Cervical spine anatomy. A, Cervical spine. Anterior view of articulated cervical vertebrae. B, Posterior view of the skull, seven cervical vertebrae, and first thoracic vertebra.
project outward, providing attachment for ligaments and muscles, and creating a moment arm to facilitate motion. The spinous processes of C3 through C6 are bifid, whereas the C7 spinous process usually is not. However, the C7 spinous process is large and is the next most prominent and most easily palpable spinous process below C2.

Five articulations are present between vertebrae from C2 through C7, including the intervertebral disk, two uncovertebral joints, and two facet (zygoapophyseal) joints. Facet joints are true apophyseal joints with hyaline cartilage articulations, intervening menisci, synovial lining, and a joint capsule. This composition makes them susceptible to degenerative changes and systemic arthritides. The cartilage and the synovial lining are aneural, whereas the joint capsule is highly innervated by the dorsal primary ramus. The facet joints are angled approximately 45 degrees from the transverse plane, articulating in concert with the uncovertebral joints and ligaments.

Intervertebral disks increase in size from C2 downward, giving the cervical spine its characteristic lordotic shape. Each disk consists of an outer annulus fibrosus and an inner nucleus pulposus, as well as cephalad and caudal end plates. The annulus fibrosus consists of type I collagen, which helps give form to the disk and provides tensile strength. The annulus fibrosus is innervated by the sinuvertebral nerve, formed by branches of the ventral nerve root and the sympathetic plexus. The nucleus pulposus consists of type II collagen and proteoglycans, which interact with water to resist compressive stress. Pressure within the disk is greatest with flexion, which may explain why those with a disk herniation find this position most uncomfortable. Disk degeneration with aging includes loss of water content and resultant loss of height, annular tears, and myxomatous changes, increasing the risk of disk herniation. This typically occurs in the posterolateral aspect of the disk, where the posterior longitudinal ligament is not present and annulus fibrosus is at its weakest, medial to the uncovertebral joints.

The spinal column is supported by the interplay of ligaments and muscles (Figure 45-3). The anterior longitudinal ligament (ALL) and the posterior longitudinal ligament (PLL) course along the anterior and posterior aspects of the vertebral bodies; the anterior longitudinal ligament resists hyperextension, and the posterior longitudinal ligament resists hyperflexion. The ligamentum flavum joins the laminae of adjacent vertebrae and may become thickened, creating stenosis in the spinal canal. In a similar manner, the interspinous ligament joins the spinous process of adjacent vertebrae. The supraspinous ligament originates as the nuchal ligament at the occiput and extends caudally as an aponeurosis until it is attached to the tip of the spinous processes of C7; it then continues to the lumbar region. Fourteen paired anterior, lateral, and posterior muscles help to orchestrate complex movements of the neck.

A brief review of the spinal cord and nerve roots is beneficial for a thorough evaluation. Grossly, the spinal cord is divided into the posterior column, the lateral column, and the anterior column. The posterior column mediates proprioceptive, vibratory, and tactile sensation; the lateral column is a conduit for motor fibers, along with pain and temperature sensation from the contralateral side of the body; the anterior column conveys crude touch sensation.

Eight total cervical nerve roots are present as the dorsal and ventral roots converge to form the spinal nerve within the vertebral foramen. Cervical nerve roots enter the intervertebral foramina by passing over the top of the corresponding pedicle, except the C8 cervical nerve, which lies between C7 and T1. Therefore, C5-C6 posterolateral disk herniation will affect the C6 nerve root. The nerve root occupies approximately one-third of the foramen (Figure 45-4). Space available for the nerve root is decreased with neck extension and degenerative changes, and is increased with neck flexion.

The anterior spinal artery arises from the vertebral arteries and supplies most of the spinal cord, excluding the posterior columns. The posterior columns receive their blood supply from the two posterior spinal arteries, which originate from the inferior cerebellar artery or the vertebral arteries. The vertebral arteries arise from the subclavian arteries and course through the C6 transverse foramen cephalad, passing anterior to the emerging cervical nerve root at each level. They pass behind the lateral mass of C1 and enter the foramen magnum, where they rejoin to form the basilar artery. Diseases such as dissection of the vessel can be associated with severe neck pain, and impairment of blood flow through the vertebral artery can result in posterior circulation signs such as nystagmus, vertigo, drop attacks, dysarthria, and visual impairment. These symptoms
are often associated with head position, and a critical reduction in blood flow can result in a cerebellar infarction.

The cervical spine is the most mobile segment of the spine; an approximate 90-degree arc of motion occurs in flexion and extension, three-quarters of which is due to extension (Table 45-1). The maximal range of motion in the sagittal plane within the subaxial spine is seen at the C5-C6 level, making this a common site of disk degeneration. Rotation encompasses approximately 80 to 90 degrees of motion, and 50% of this occurs at the atlantoaxial joint. As is seen with extension, rotation reduces the cross-sectional area of the spinal canal. The cervical spine demonstrates 30 degrees of lateral mobility in each direction; this typically occurs with some degree of rotation secondary to the orientation of the facet joints.

## AXIAL NECK PAIN

Axial neck pain may originate from any tissue that receives innervation, including zygapophyseal joints, cervical disks, vertebral periosteum, posterior neck muscles, cervical dura mater, occipito-atlantoaxial joints, and vertebral arteries. Sources may include degenerative, traumatic, malignant, infectious, or systemic inflammatory processes. Zygapophyseal joints and cervical disks have the greatest quantity of direct supporting data suggesting these sites as the origin of axial neck pain. Provocative injection into the facet joint in asymptomatic volunteers invokes a reproducible pattern of occipital or axial neck pain. This pattern of pain can be accurately diagnosed and treated for at least a short time with anesthetic injections targeted at the joint capsule itself, or with blockage of the respective dorsal primary rami. Degenerative arthritis within the upper cervical spine can manifest as suboccipital headaches denoted as cervicogenic headaches, which are thought to result from irritation of the greater occipital nerve. Typically, arthritis within the atlanto-occipital joints is made worse by provocative neck flexion and extension, whereas atlantoaxial arthritis is made worse by rotation. This is supported by investigators who injected asymptomatic volunteers at the atlanto-occipital and atlantoaxial joints, reproducing this pattern of pain. Relief of suboccipital pain may be attained with fluoroscopically guided injection of corticosteroid into the diseased joint, or fusion of these joints in recalcitrant cases.

The cervical disk is a controversial source of axial neck pain, which occurs as the result of insult to the highly innervated annulus fibrosus. This idea is based on provocative diskography, whereby a diseased disk is fluoroscopically injected to a given pressure, and pain is reproduced in reliable patterns. In a true-positive or concordant study, an adjacent normal disk should not produce pain when injected. Even with careful technique, a false-positive result can occur, and it is not unusual to have a nonconcordant study in which multiple disks elicit a pain response despite being normal. Interventions directed at treating the disk as the source of axial neck pain have been found to be unpredictable, although the cervical disk is likely to contribute to axial neck pain.

Myofascial pain due to irritation of the muscles about the neck can contribute to axial neck pain. Patients with
chronic myofascial pain have been shown to have a lower level of high-energy phosphates in involved muscle tissue.\textsuperscript{25} Myofascial pain can be the original source of neck pain or, more commonly, a manifestation of postural adaptations and compensatory overuse of normal tissue that remains after the injured structure heals. A more generalized form of this is fibromyalgia, a widespread disorder defined by diffuse pain affecting all four quadrants of the body, with at least 11 of 18 pressure points noted as positive. Patients have associated symptoms of fatigue, cognitive difficulties, and irritable bowel syndrome, and a nongeriatric pattern of dysesthesias, weakness, and paresthesias.\textsuperscript{21} Systemic inflammatory arthropathies causing neck pain typically demonstrate the classic pattern of morning stiffness, polyarticular involvement, rigidity, and associated cutaneous manifestations. Rheumatoid arthritis (RA) often involves the cervical spine, initially causing stiffness and later causing pain, potentially leading to instability. After the hands and feet, the cervical spine is the most common site of disease involvement in RA.\textsuperscript{22} The upper cervical spine is most commonly involved (occiput to C2), followed by the subaxial cervical spine (C3-C7). The likelihood of developing cervical spine pathology can be predicted by the extent of rheumatoid changes that occur in the hands and feet. Basilar invagination is one such manifestation, whereby C1 lateral masses erode, allowing the odontoid peg to settle into the foramen magnum, placing pressure on the brain stem with potential for instantaneous death. The atlantoaxial joint can demonstrate instability with potential for neurologic injury. Because of these potentially catastrophic complications, dynamic radiographs of the cervical spine should be obtained before any procedure requiring intubation is performed. Seronegative spondyloarthropathies that can manifest with neck pain include ankylosing spondylitis, psoriatic arthritis, and reactive arthritis. In 70% of patients, psoriatic arthritis will manifest with skin lesions before arthritis develops; reactive arthritis rarely involves the cervical spine.

Ankylosing spondylitis often affects the entire axial skeleton and is seen as limited lumbar motion and chest expansion with later involvement of the cervical spine. In progressive patterns, the cervical spine takes on a kyphotic deformity. As the spine fuses, it biomechanically becomes similar to a long bone; minor trauma with neck pain should be taken very seriously in these patients. Even in the face of negative plain radiographs, patients should be worked up extensively with strict spine precautions and with neutral alignment varying with the baseline spinal curvature; frequent neurologic evaluations should be performed to check for development of an epidural hematoma.

Infection and neoplasms can cause axial neck pain through bone destruction with irritation of vertebral body periosteal nerves and altered biomechanics of the facet joints and cervical disks. The onus is upon the clinician to identify these patients at the initial visit because a delay in diagnosis can have catastrophic consequences. Red flags for axial neck pain that require further workup at the initial presentation include elderly patients, patients with a history of malignancy, immunocompromised patients, and those with fever, chills, unexplained weight loss, fatigue, night-time awakening, recent antecedent bacteremia, and severe nonmechanical neck pain.\textsuperscript{21} Patients who have undergone previous cervical spine surgery should be evaluated for the presence of pseudoarthrosis or iatrogenic instability. Pseudoarthrosis is failure of an attempted arthrodesis or fracture to fully heal with bridging bone. Patients will describe a “honeymoon” period, whereby they did well for 3 to 6 months following surgery. Patients then develop worsening axial neck and interscapular pain with associated headaches. This can be diagnosed on plain radiographs with evidence of hardware loosening or movement on dynamic images. Computed tomography (CT) scanning with coronal and sagittal reconstructions should be done to more definitively diagnose a pseudoarthrosis. Once pseudoarthrosis has been diagnosed, a surgical consult should be obtained, the patient should be counseled on smoking cessation given the deleterious effects of nicotine and carbon monoxide on bone healing, and a bone metabolic workup should be undertaken. An additional cause of neck pain following surgery is iatrogenic instability, whereby the surgery itself creates pathologic motion. This requires a surgical consultation to determine whether stabilization is warranted.

**RADICULOPATHY AND MYELOPATHY**

The clinician must determine whether there is evidence of nerve root compression, termed radiculopathy, versus spinal cord compression, termed myelopathy. Cervical spondylosis with changes within the disk may cause loss of height with posterior bulging of the disk into the spinal canal and foramen. As the disk collapses, the posterior soft tissue structures, including the ligamentum flavum and the facet joint capsule, fold inward, further compromising the spinal canal and neural foramen. Pressure that once was dispersed throughout the disk is transferred to the facet joints and uncinate processes, resulting in the development of bone overgrowth or osteophytes and causing extrinsic pressure on the nerve root or spinal cord.

In radiculopathy, mechanical distortion of the nerve leads to increased vascular permeability, resulting in chronic edema and eventually fibrosis. This causes hypersensitivity of the nerve root with an inflammatory response mediated by chemicals released from the cell bodies of sensory neurons and cervical disks.\textsuperscript{21} Compression of the dorsal root ganglion is felt to be especially important in producing radicular pain.\textsuperscript{24} Clinically, this presents with pain in a dermatomal distribution; dermatomes for the higher cervical nerve roots, including C3 and C4, are found along the posterior scapula, and the pain should not be confused with isolated axial neck pain.\textsuperscript{25} Minor symptoms that are tolerable may be treated with conservative care, but persistent compression on a nerve root can lead to sensory loss and weakness. Disabling deficits should be treated operatively given that prolonged nerve compression can result in irreversible changes. In patients without a neurologic deficit, it is reasonable to expect a good outcome with conservative care.\textsuperscript{26} Myelopathy has a clinical presentation of long tract signs resulting from compression of the spinal cord. Factors that contribute to the development of myelopathy include a congenitally narrow spinal canal, dynamic cord compression, dynamic thickening of the spinal cord, and vascular changes. The anterior-posterior diameter in the subaxial spine for a normal adult measures 17 to 18 mm. The cord
increasing pain evokes concern regarding a mass effect. The presence of neurologic symptoms other than pain. In duration, quality, and aggravating factors, as well as the characteristics to note include onset, distribution, frequency, and aggravating factors, as well as the presence of neurologic symptoms other than pain. In general, pain that is present only intermittently may be indicative of instability or motion, whereas constant and increasing pain evokes concern regarding a mass effect. New-onset and relatively short duration generalized neck pain is likely related to benign pathology such as muscle strain, whereas a longer duration of symptoms indicates significant or progressive pathology. Well-localized pain indicates specific nerve root irritation, whereas poorly defined pain may derive from irritation of deep connective tissue structures such as muscle, joint, bone, or disk. Aggravating and relieving factors may help elucidate biomechanical changes in the cervical spine that are contributing to the symptoms.

Localized axial neck pain is commonly reported as originating posteriorly with extension into the shoulder or occiput. Localized pain of myofascial origin may worsen with neck flexion, whereas diskogenic neck pain will worsen with neck extension or rotation. Pain referred to the occiput usually indicates pathologic changes in the upper cervical spine and may radiate down the neck and to the ear. Shoulder girdle pain develops secondary to postural adaptations from initial neck pain symptoms. It is not uncommon for pain to be referred from the shoulder, heart, lungs, viscera, or temporomandibular joint to the neck region as the result of overlapping nerve distribution. Symptoms may arise owing to irritation or activation of receptors directly, as is seen in articular pain, pseudoarticular pain, vascular and cervicogenic headaches, pseudo–angina pectoris, eye and ear symptoms, and throat symptoms.

Articular symptoms arise from innervation to the facet and uncovertebral joints causing local pain and stiffness. Patients often state that their symptoms are made worse with inactivity and describe feelings of clicking, grating, or “sand” in the neck. Pseudoarticular pain may be felt in the shoulder and elbow, with true pathology originating from the neck. Vascular symptoms result from compression of the vertebral artery by osteophytes or a protruding disk. Symptoms may intensify with neck movement or with certain postures. Tenosynovitis and tendinitis may involve the rotator cuff and tendons about the elbow, wrist, or hand. Stenosis or fibrosis of tendon sheaths or palmar fascia may be present, along with trigger points over the affected joints, giving a false impression of local pathology.

Localization of Pain Generators

Pain may be somatic or autonomic and is not always felt in precise anatomic zones. Overlapping sensory supplies may be present, as well as radiation in spinal segments by recruitment within the spinal column, causing difficulty in localization. Somatic pain is caused by cervical nerve root irritation. It is the most common type of pain, and diabetic patients are more susceptible to this nerve root irritation. Neurologic deficits correspond with the offending disk level in 80% of patients. Neuralgic and myalgic pain describes symptoms related to compression of different areas of the nerve root. Neuralgic pain originates from irritation of the dorsal sensory root and has a “lightning” or “electric” sensation, which tends to be dermatomal and associated with numbness and paresthesia. Pain tends to present more proximally and paresthesia more distally. Myalgic pain occurs with irritation of the ventral motor root. This pain is described as a deep, boring, unpleasant sensation that tends to be poorly localized because of its referral to sclerotomal areas. These sensations conform to the areas of muscles present that are innervated by the compressed nerve root.

CLINICAL FEATURES

In terms of functional anatomic pathways, neck pain is mediated via somatic or autonomic pathways. Somatic pain is the most common, being perceived in dermatomes, myotomes, or sclerotomes. Pain originating in the autonomic pathway, or in the sympathetic nervous system, may fall into somatic segmental distributions, vascular supply distributions, peripheral nerve distributions, or nonconforming patterns. Because pain mediation pathways may have significant overlap, additional clinical information regarding characteristics of neck pain, along with diagnostic studies, complements the determination of pain origin when localization based on functional anatomy is not sufficient.

Patient History

Neck pain is the most common symptom of cervical spine pathology; correctly characterizing it helps to identify conditions requiring immediate treatment. Important characteristics to note include onset, distribution, frequency, duration, quality, and aggravating factors, as well as the presence of neurologic symptoms other than pain. In general, pain that is present only intermittently may be indicative of instability or motion, whereas constant and increasing pain evokes concern regarding a mass effect. New-onset and relatively short duration generalized neck pain is likely related to benign pathology such as muscle strain, whereas a longer duration of symptoms indicates significant or progressive pathology. Well-localized pain indicates specific nerve root irritation, whereas poorly defined pain may arise from irritation of deep connective tissue structures such as muscle, joint, bone, or disk. Aggravating and relieving factors may help elucidate biomechanical changes in the cervical spine that are contributing to the symptoms.

Localized axial neck pain is commonly reported as originating posteriorly with extension into the shoulder or occiput. Localized pain of myofascial origin may worsen with neck flexion, whereas diskogenic neck pain will worsen with neck extension or rotation. Pain referred to the occiput usually indicates pathologic changes in the upper cervical spine and may radiate down the neck and to the ear. Shoulder girdle pain develops secondary to postural adaptations from initial neck pain symptoms. It is not uncommon for pain to be referred from the shoulder, heart, lungs, viscera, or temporomandibular joint to the neck region as the result of overlapping nerve distribution. Symptoms may arise owing to irritation or activation of receptors directly, as is seen in articular pain, pseudoarticular pain, vascular and cervicogenic headaches, pseudo–angina pectoris, eye and ear symptoms, and throat symptoms.

Articular symptoms arise from innervation to the facet and uncovertebral joints causing local pain and stiffness. Patients often state that their symptoms are made worse with inactivity and describe feelings of clicking, grating, or “sand” in the neck. Pseudoarticular pain may be felt in the shoulder and elbow, with true pathology originating from the neck. Vascular symptoms result from compression of the vertebral artery by osteophytes or a protruding disk. Symptoms may intensify with neck movement or with certain postures. Tenosynovitis and tendinitis may involve the rotator cuff and tendons about the elbow, wrist, or hand. Stenosis or fibrosis of tendon sheaths or palmar fascia may be present, along with trigger points over the affected joints, giving a false impression of local pathology.

Localization of Pain Generators

Pain may be somatic or autonomic and is not always felt in precise anatomic zones. Overlapping sensory supplies may be present, as well as radiation in spinal segments by recruitment within the spinal column, causing difficulty in localization. Somatic pain is caused by cervical nerve root irritation. It is the most common type of pain, and diabetic patients are more susceptible to this nerve root irritation. Neurologic deficits correspond with the offending disk level in 80% of patients. Neuralgic and myalgic pain describes symptoms related to compression of different areas of the nerve root. Neuralgic pain originates from irritation of the dorsal sensory root and has a “lightning” or “electric” sensation, which tends to be dermatomal and associated with numbness and paresthesia. Pain tends to present more proximally and paresthesia more distally. Myalgic pain occurs with irritation of the ventral motor root. This pain is described as a deep, boring, unpleasant sensation that tends to be poorly localized because of its referral to sclerotomal areas. These sensations conform to the areas of muscles present that are innervated by the compressed nerve root.
Autonomically mediated symptoms result in dizziness, blurring of vision, tinnitus, retro-ocular pain, and facial and jaw pain.

It is important to determine whether axial neck pain is isolated, or if radiating pain, weakness, changes in sensation, or alterations in proprioception are associated. Compression of a nerve root often can be localized by identifying the distribution of pain, paresthesias, or weakness as they follow the segmental distribution of that respective nerve root (Table 45-2). Sensory loss may be characterized by the patient in precise terms with a description of numbness, or alternatively may consist of vague symptoms of swelling or bogginess to the skin. If the face, head, or tongue is involved, the upper three nerve roots of the cervical plexus may be affected. Numbness of the neck, shoulder, arm, forearm, or fingers indicates involvement of C5-T1. Weakness, as occurs with sensory changes, appears in a graded fashion, depending on the amount of compression upon the nerve root. This event will present clinically with an obvious functional deficit or more subtle findings, made obvious only by repetitive testing. The clinician must be attuned to these subtle complaints of weakness, which may be described by the patient as a feeling of heaviness of the limbs, early fatigue, or insufficient power grip. If obvious atrophy is present in a muscle, more than one nerve root is affected as a result of the evolutionary benefit of multiple levels of innervation for a given muscle. Alterations in proprioception are due to compression of the dorsal column of the spinal cord. This will be described by the patient as symptoms of clumsiness, with complaints of tripping or dropping objects. This is a matter of concern for the more ominous condition of myelopathy, which may occur secondary to spinal cord compression.

Cervical spinal disease classically causes isolated axial neck pain or radicular pain that radiates to the shoulder or down the upper extremity (Table 45-3). However, less commonly, it can be the cause of headache, pseudo–angina pectoris, and otolaryngologic sensations. Cervicogenic occipital headaches can be compounded by adaptive changes in the posterior occipital muscles, often spreading to the eye region and manifesting as dull rather than pulsating pain. These headaches are unique in that they are aggravated by neck movements. Patients typically have migraine-like symptoms such as phonophobia or photophobia.

Pseudo–angina pectoris has been reported as the result of cervical spinal disease and may be confused with angina pectoris or breast pain in women (Figure 45-5). In the presence of a C6-C7 lesion, neuralgic or myalgic pain may be noted, along with tenderness in the precordium or scapular region. Heart disease is differentiated from symptoms associated with C6-C7 dysfunction on the basis of muscle weakness, fasciculations, or sensory or reflex changes. Differentiation of these two pathologies may be difficult when true angina and pseudoangina coexist in the same patient.

Cervical spinal disease may manifest in the form of eye, ear, and throat symptoms (see Figure 45-5). Eye and ear symptoms may arise from irritation of the plexuses surrounding the vertebral and internal carotid arteries. Eye symptoms can present as blurring of vision relieved by changing neck position, increased tearing, orbital and retro-orbital pain, or descriptions of eyes being “pulled backward” or “pushed forward.” Altered equilibrium with associated gait disturbances may result from irritation of the surrounding sympathetic plexus or from vertebral insufficiency. Hearing can be affected, with tinnitus and altered auditory acuity.

### Table 45-2  Cervical Nerve Root Segments and Corresponding Clinical Signs and Symptoms

<table>
<thead>
<tr>
<th>Nerve Root</th>
<th>Symptom</th>
<th>Correlate</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>Suboccipital pain with extension to the back of the ear</td>
<td>If C3, C4, and C5 are all involved, may cause paradoxical breathing</td>
</tr>
<tr>
<td>C4</td>
<td>Pain from caudad aspect of the neck to superior aspect of the shoulder</td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>Numbness over shoulder and down lateral aspect of arm to midportion. Deltoid muscle may be weak, and biceps reflex, which is innervated by C5-C6, may be affected.</td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td>Radiating pain and numbness down the lateral aspect of arm and forearm to the thumb and IF (“six shooter”). Weakness in wrist extension, elbow flexion, and supination. Diminished brachioradialis and biceps reflex.</td>
<td>Sensory component can mimic carpal tunnel syndrome</td>
</tr>
<tr>
<td>C7</td>
<td>Numbness and pain down the posterior aspect of arm and forearm to LF. Weakness in the triceps, wrist flexion, and finger extensors.</td>
<td>Most frequent. Entrapment of posterior interosseous nerve can mimic motor component; however, no sensory deficits are present.</td>
</tr>
<tr>
<td>C8</td>
<td>Pain and numbness down the medial aspect of arm and forearm into the small ring finger. Weakness in the FDP to IF and LF and FPL.</td>
<td>AIN entrapment can mimic motor component of C8 or T1; however, sensory changes and involvement of thenar muscles will not be present. Ulnar nerve entrapment will spare short thenar muscles with exception of ADP. Does not involve FPL or FDP to IF and LF.</td>
</tr>
</tbody>
</table>

ADP, adductor pollicis; AIN, anterior interosseous nerve; FDP, flexor digitorum profundus; FPL, flexor pollicis longus; IF, index finger; LF, long finger.

### Table 45-3  Cervical Pain Referral Pathways

<table>
<thead>
<tr>
<th>Location of Pain</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper posterolateral cervical region</td>
<td>C0-C1, C1-C2, C2-C3</td>
</tr>
<tr>
<td>Occipital region</td>
<td>C2-C3, C3</td>
</tr>
<tr>
<td>Upper posterior cervical region</td>
<td>C2-C3, C3-C4, C3</td>
</tr>
<tr>
<td>Middle posterior cervical region</td>
<td>C3-C4, C4-C5, C4</td>
</tr>
<tr>
<td>Lower posterior cervical region</td>
<td>C4-C5, C5-C6, C4, C5</td>
</tr>
<tr>
<td>Suprascapular region</td>
<td>C4-C5, C5-C6, C4</td>
</tr>
<tr>
<td>Superior angle of scapula</td>
<td>C6-C7, C6, C7</td>
</tr>
<tr>
<td>Midscapular region</td>
<td>C7-T1, C7</td>
</tr>
</tbody>
</table>
reported. Throat symptoms, including dysphagia, may be related to anterior vertebral osteophytes causing direct compression, as well as to cranial nerve and sympathetic nerve communications.

Symptoms of dyspnea and cardiac arrhythmia and drop attacks may have a cervical spinal origin. Dyspnea can be related to a deficit in C3-C5 innervation of the diaphragm. Cardiac palpitations and tachycardia secondary to cervical spine pathology can be differentiated from those associated with other causes by the fact that these symptoms are associated with unusual positions or hyperextension of the neck. This is caused by irritation of C4 innervation of the diaphragm and pericardium, or by irritation of the cardiac sympathetic nerve supply. Drop attacks suggest posterior circulation insufficiency, resulting in an abrupt loss of proprioception without loss of consciousness.

Myelopathy, or spinal cord compression, initially presents with subtle complaints of hand clumsiness or difficulty with balance. Patients will report worsening handwriting in the past few months or difficulty buttoning shirts. Patients may have nausea and emesis caused by equilibrium dysfunction. Paresthesias and dysesthesias may be present, often involving bilateral upper extremities and not following a dermatomal distribution. This is often mistaken for peripheral neuropathy or carpal tunnel syndrome but should be considered when bilateral extremity symptoms are present. As the disease progresses over time, more advanced manifestations include weakness in the triceps, hand intrinsics, and hip flexors most commonly. Late manifestations include spasticity, as well as bowel and bladder dysfunction.

Finally, neck pain may present concomitant with systemic disease with varied symptoms requiring further investigation. Examples include inflammatory arthritides, infection or tumor, multiple sclerosis, subacute combined degeneration, and syringomyelia. Inflammatory arthritides often present with morning stiffness, polyarticular involvement, or cutaneous manifestations. Fever, weight loss, or night pain points to an infectious or neoplastic origin. A Pancoast tumor is a neoplastic process of the apical portion of the lung that can cause a mass effect on the caudad cervical nerve roots. This should always be considered in a person with radicular symptoms and a history of smoking, with workup including a chest x-ray. Idiopathic brachial plexus neuritis, formerly known as Parsonage-Turner syndrome, is caused by viral infection of the brachial plexus presenting with severe arm pain involving multiple nerve roots. Once the acute phase resolves, patients are left with variable deficits. Subacute combined degeneration from vitamin B12 deficiency is a consideration when sensory deficit is greater in the lower extremities.

Clinical Examination

A careful clinical examination provides additional focus to the differential diagnosis. The clinical examination begins broadly with observation of the patient’s gait, as well as head and neck posture. Further palpation, range-of-motion testing, and neurologic examination for motor signs, reflexes, sensory signs, autonomic signs, and articular signs are performed (Table 45-4).

Careful palpation with knowledge of key anatomic bony and soft tissue landmarks in the cervical spine may localize pain to a particular cervical level and location. Anteriorly or anterolaterally, the transverse process of C1 is palpated between the angle of the jaw and the styloid process. C3 is identified by palpation of the hyoid bone. C4-C5 is at the level of the thyroid cartilage, and C6 is at the level of the cricoid ring and the carotid tubercle. In the process of examining the neck, also note the sagittal balance with retention or loss of normal cervical lordosis. Posteriorly and...
posteriorly, the occiput, inion, superior nuchal line, mastoid processes, and spinous processes of C2 and C7-T1 are palpable.

Soft tissues about the anterior and posterior triangles of the neck, occipital region, and posterior paraspinal muscles are examined. The sternocleidomastoid muscle is involved with whiplash injury, whereby abrupt hyperextension of the neck occurs. The muscle may be tender to palpation, or the patient may be splinting the neck with the head turned away from the injured muscle. This posturing of the neck is termed torticollis, and the clinician should remember that the head is turned away from the side of the involved sternocleidomastoid. Flexion injury may traumatize the trapezius muscle. Midline cervical tenderness is more of a concern with ligament injury, whereas paraspinal muscle tenderness typically is a more benign process. The greater occipital nerves are located lateral to the inion and may be involved in traumatic inflammation associated with flexion or extension injury resulting in suboccipital headaches. Skin markings or visible trauma should be noted at this time.

Range-of-motion examination may reveal pain or limitations in flexion-extension, lateral bending, and rotation. Flexion limitation may be assessed by the examiner placing fingers between the patient’s chin and sternum at maximum flexion with 50% of the motion occurring at the occiput-C1 joint and the remaining 50% distributed over C2-C7. If the patient is unable to place the chin on the chest, the interval between the patient’s chin and sternum at maximum flexion with 50% of rotation occurring at C1-C2, and the remaining 50% distributed in the subaxial spine between C3 and

### Table 45-4 Nerves and Tests of Principal Muscles

<table>
<thead>
<tr>
<th>Nerve</th>
<th>Nerve Roots</th>
<th>Muscle</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessory</td>
<td>Spinal</td>
<td>Trapezius</td>
<td>Triceps</td>
</tr>
<tr>
<td>Brachial plexus</td>
<td>Spinal</td>
<td>Sternoceleidomastoid</td>
<td>Triceps</td>
</tr>
<tr>
<td>C5, C6</td>
<td>Clavicular part</td>
<td>Triceps</td>
<td></td>
</tr>
<tr>
<td>C7, C8, T1</td>
<td>Serratus anterior</td>
<td>External rotation of arm</td>
<td></td>
</tr>
<tr>
<td>C4, C5</td>
<td>Infraspinatus</td>
<td>External rotation of arm</td>
<td></td>
</tr>
<tr>
<td>C4, C5, C6</td>
<td>Supraspinatus</td>
<td>Adduction of arm</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>Latissimus dorsi</td>
<td>Adduction of horizontal, externally rotated arm, coughing</td>
<td></td>
</tr>
<tr>
<td>Axillary</td>
<td>C5, C6</td>
<td>Deltoid</td>
<td>Lateral and forward elevation of arm to horizontal</td>
</tr>
<tr>
<td>Musculocutaneous</td>
<td>C5, C6</td>
<td>Biceps</td>
<td>Flexion of supinated forearm</td>
</tr>
<tr>
<td>Radial</td>
<td>C6, C7, C8</td>
<td>Triceps</td>
<td>Extension of forearm</td>
</tr>
<tr>
<td>C5, C6</td>
<td>Brachioradialis</td>
<td>Flexion of semiprone forearm</td>
<td></td>
</tr>
<tr>
<td>C6, C7</td>
<td>Extensor carpi radialis longus</td>
<td>Extension of wrist to radial side</td>
<td></td>
</tr>
<tr>
<td>Posterior</td>
<td>C5, C6</td>
<td>Supinator</td>
<td>Supination of extended forearm</td>
</tr>
<tr>
<td>interosseous</td>
<td>C7, C8, C8</td>
<td>Extensor digitorum</td>
<td>Extension of proximal phalanges</td>
</tr>
<tr>
<td>C7, C8</td>
<td>Extensor carpi ulnaris</td>
<td>Extension of wrist to ulnar side</td>
<td></td>
</tr>
<tr>
<td>C7, C8</td>
<td>Abductor pollicis longus</td>
<td>Abduction of first metacarpal in plane at right angle to palm</td>
<td></td>
</tr>
<tr>
<td>C7, C8</td>
<td>Extensor pollicis longus</td>
<td>Extension of first interphalangeal joint</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>C6, C7</td>
<td>Pronator teres</td>
<td>Extension of first metacarpophalangeal joint</td>
</tr>
<tr>
<td>C6, C7</td>
<td>Flexor carpi radialis</td>
<td>Flexion of wrist to radial side</td>
<td></td>
</tr>
<tr>
<td>C7, C8, T1</td>
<td>Flexor digitorum superficialis</td>
<td>Flexion of middle phalanges</td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>Flexor digitorum profundus</td>
<td>Flexion of terminal phalanges, index and middle fingers</td>
<td></td>
</tr>
<tr>
<td>(lateral part)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>Flexor pollicis longus (anterior interosseous nerve)</td>
<td>Flexion of distal phalanx, thumb</td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>Flexor pollicis brevis</td>
<td>Abduction of first metacarpal in plane at right angle to palm</td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>Flexor pollicis brevis</td>
<td>Flexion of proximal phalanx, thumb</td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>Opponens pollicis</td>
<td>Opposition of thumb against 5th finger</td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>1st and 2nd lumbricals</td>
<td>Extension of middle phalanges while proximal phalanges are fixed in extension</td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>Flexor pollicis brevis</td>
<td>Flexion of distal phalanges of ring and little fingers</td>
<td></td>
</tr>
<tr>
<td>Ulnar</td>
<td>C7, C8</td>
<td>Flexor carpi ulnaris</td>
<td>Observation of tendons during testing of abductor digiti minimi</td>
</tr>
<tr>
<td>C8, T1</td>
<td>Flexor digitorum profundus</td>
<td>Flexion of distal phalanges of ring and little fingers</td>
<td></td>
</tr>
<tr>
<td>(medial part)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>Hypothenar muscles</td>
<td>Abduction and opposition of little finger</td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>3rd and 4th lumbricals</td>
<td>Extension of middle phalanges while proximal phalanges are fixed in extension</td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>Adductor pollicis</td>
<td>Adduction of thumb against palmar surface of index finger</td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>Flexor pollicis brevis</td>
<td>Flexion of proximal phalanx, thumb</td>
<td></td>
</tr>
<tr>
<td>C8, T1</td>
<td>Interossei</td>
<td>Abduction and adduction of fingers</td>
<td></td>
</tr>
</tbody>
</table>
C7. A natural decrease in range of motion occurs with age, even in the healthy individual.\(^6\)

Range-of-motion tests the ligaments, capsules, and fascia; this motion will be reduced in the presence of cervical spinal muscular spasm or pain. Patients with degenerative changes in the cervical spine will have pain with decreased range of motion of the cervical spine without resistance. The most common findings due to changes in cervical spine articulations are as follows (in order): restriction of movement with or without pain, pain on movement, and local tenderness. Lateral flexion is the earliest and most impaired movement in degenerative diseases; rotation is first impaired in rheumatoid arthritis owing to involvement of the odontoid peg. A uniformly stiff neck may be caused by diffuse idiopathic skeletal hyperostosis (DISH), present in a quarter of elderly patients, but also may be due to ankylosing spondylitis or recent trauma to the neck.\(^{41}\) If articul ar signs are found, the examiner must evaluate the entire vertebral column and peripheral joints for evidence of further arthritis and to search for extra-articular manifestations.

Motion against resistance testing is performed after active and passive range of motion is established. Muscle groups tested include the flexors and extensors of the neck. In testing flexor muscles, a hand is placed between the forehead and the chest. The primary flexor is the sternocleidomastoid muscle; secondary flexors include the three scalene muscles and the small prevertebral muscles. Extensors are tested by placing a hand on the shoulder and head for resistance. Primary extensors include the paravertebral extensor mass, splenius, semispinalis capitis, and trapezius. Secondary flexors include the small intrinsic muscles of the neck. Rotators are examined by placing a hand on the shoulder and chin for resistance. The sternocleidomastoid muscle and the intrinsic muscles of the neck provide rotational force. Motion against resistance testing should include active maximum effort strength testing to the extremes of flexion, extension, and rotation to assess muscle strength. Causes of decreased range of motion of the cervical spine include joint locking and bony ankylosis resulting from degenerative changes or arthritides, fibrous contractures, muscle spasm, splinting over painful joints, and nerve root or spinal cord compression or irritation. Decreased range of motion in the presence of pain or weakness warrants further investigation.

Sensation for light touch, pin prick, temperature, and proprioception should be performed. These tests are admittedly subjective; therefore both extremities should be compared to assess differences in sensation. Comparing an unaffected area such as the face with the area of decreased sensation can also be helpful. Pin prick can be performed by using a sterile needle and temperature by using an alcohol pad to assess the function of the spinothalamic tract that traverses the anterolateral aspect of the spinal cord. Light touch and proprioception assess the function of the posterior spinal column.

Dermatomes are anatomically distributed as noted in Figure 45-6. The lower extremities demonstrate a unique dermatomal map that correlates with embryologic development, whereby the limb starts in a supinated position and pronates with longitudinal growth. Perineal sensation and rectal tone are important to examine because an abnormality may indicate compression of the spinal cord or cauda equina, requiring immediate surgical intervention. Isolating the level of pathology at times can be challenging. Nerve roots with proximal compression are more susceptible to distal compression in a phenomenon termed double crush. The cervical spine should always be considered as the potential source in patients who present with symptoms of carpal or cubital tunnel syndrome and peripheral neuropathy. Ancillary imaging and nerve conduction studies can help elucidate the origin.

After palpation, range-of-motion testing, and assessment of sensation, muscle strength testing is continued for localization of any positive findings. Lower motor neuron disease

---

Figure 45-6 Dermatome distributions. Dermatome distribution of nerve fibers from C1 through T5, carrying senses of pain, heat, cold, vibration, and touch to the head, neck, arm, hand, and thoracic area. The sclerotomes and myotomes are similar but show some overlap. Pain arising from structures deep to the deep fascia (myotome and sclerotome) does not precisely follow the dermatome distribution.
is indicated by weakness, hypotonia, and fasciculations. Upper motor neuron disease is indicated by spasticity. Motor function should be graded using the standard 0 to 5 nomenclature, with grade 0 having no function, 1 having trace, 2 having full range of joint motion with gravity eliminated, 3 having antigravity function, 4 having function against slight resistance, and 5 having normal strength against resistance (Table 45-5). If weakness is present, a more focused examination should be performed to look at other muscles innervated by that same nerve root.

Deep tendon stretch reflexes should be performed and graded 0 to 3 with 0 being no response, 1 being hyporeflexive, 2 being normal, and 3 being hyperreflexive. C5 is tested by striking the biceps tendon; C6, brachioradialis; C7, triceps; L4, patellar tendon; and S1, Achilles tendon. To facilitate reflex testing, it may be helpful to use muscle loading or Jendrassik’s maneuver (performed by having the patient flex both sets of fingers into a hook-like form, interlocking the hands, and pulling apart). This creates a diversion to help relax the patient and better assess lower extremity reflexes. If difficulty with reflex testing persists, ensure that no peripheral neuropathy is present. In addition to deep tendon reflex testing, the abdominal reflex, the Babinski test, and the bulbocavernous test should be assessed.

Provocative tests that can be helpful in confirming compressive extradural monoradiculopathy include Spurling’s test, the arm abduction test, and the axial compression and traction test. All of these tests are meant to change the diameter of the neural foramen, thus increasing or decreasing the symptoms, respectively. Spurling’s test is performed by having the patient extend his or her neck and rotate toward the side of pain. The test is positive if radicular pain is made worse in this position; this indicates foraminal stenosis with potential compression of a nerve root. The arm abduction sign is positive if the patient’s pain is relieved by placing the hand on the affected side, on top of the head. The axial compression test is performed by pressing on top of the patient’s head with the neck in neutral position; the result is positive if radicular symptoms are exacerbated by this maneuver and are relieved by placing traction on the head and opening up the foramina.

Provocative tests that are helpful in diagnosing myelopathy include the presence of Hoffmann’s sign, the finger escape sign, the abnormal grip-release test, and Lhermitte’s sign. Hoffmann’s sign is performed by holding the middle finger extended and suddenly extending the distal interphalangeal joint, leading to flexion of the index finger and thumb if pathologic. A positive finger escape sign occurs when an individual cannot hold all fingers in an adducted and extended position without the ulnar two digits falling into flexion and abduction over time. The grip-release test is an inability to rapidly open and close a fist caused by weakness and spasticity of the hand. Lhermitte’s sign evaluates for changes in the spinal cord itself and occurs when the patient’s neck is forcefully flexed, resulting in electric-like shocks that travel down the arms and legs. This indicates changes in the white matter of the spinal cord and may occur secondary to cervical myelopathy or multiple sclerosis.

### Table 45-5  Strength Grading in Motor Examination

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No function with total paralysis</td>
</tr>
<tr>
<td>1</td>
<td>Trace movement with palpable or visible contraction</td>
</tr>
<tr>
<td>2</td>
<td>Full range of joint motion with gravity eliminated</td>
</tr>
<tr>
<td>3</td>
<td>Active movement against gravity</td>
</tr>
<tr>
<td>4</td>
<td>Active movement against slight resistance</td>
</tr>
<tr>
<td>5</td>
<td>Normal strength</td>
</tr>
</tbody>
</table>

**DIAGNOSTIC EVALUATION**

After the history and clinical examination are completed, imaging studies, neurophysiologic procedures, and laboratory studies may aid in completing the differential diagnosis and building a treatment plan for the patient’s neck pain. Cervical radiographs often show degenerative changes in asymptomatic individuals in their 60s. In the absence of trauma, constitutional symptoms, or worsening neurologic deficit, 4 to 6 weeks of conservative care is indicated before radiographs are obtained. Dynamic radiographs should also be used for screening patients with rheumatoid arthritis before endotracheal intubation, given the risk of cervical instability. One study demonstrated that 61% of patients with RA had evidence of instability defined by at least 3 mm of atlantoaxial subluxation on preoperative screening x-rays.

In the presence of significant degenerative changes and end plate osteophytes, CT myelography can be helpful in further characterizing bony involvement. CT myelography should be thought of as a complementary test to magnetic resonance imaging (MRI). It should be used as the primary test to evaluate neural involvement only when MRI is contraindicated, because MRI is superior for evaluating spinal root changes such as syringomyelia, myelomalacia, or neoplasm. MRI scanning is indicated for progressive neurologic deficit, disabling weakness, or long tract signs and is recommended for patients with persistent cervical radiculopathy after 6 weeks of conservative care. The addition of gadolinium contrast enhancement is helpful in evaluating infection and neoplasm, and in differentiating scar from recurrent disk herniation in patients who have undergone previous spinal surgery. MRI results must be correlated with physical examination findings given that asymptomatic volunteers have been found to have abnormal cervical spine MRI. Increased signal on T2 sequence can be representative of a spectrum of disease from edema to myelomalacia and syrinx formation. Therefore the presence of increased signal within the spinal cord warrants a surgical consultation, and operative intervention, or close follow-up at a minimum, is provided if examination and history correlate. If MRI findings do not correlate with the history and physical examination findings, additional studies such as CT myelography may be required. CT myelography is superior for detecting bony foraminal stenosis. Nuclear bone scanning techniques, including single-photon emission computed tomography (SPECT) scans, have been used to identify and characterize acuity in occult fracture, periosteal injury, and posttraumatic osteoarthritis in the absence of positive radiographic findings.

Neurophysiologic procedures are indicated when the clinical examination and imaging studies fail to correlate, or when there is conflicting information. Electromyography (EMG), nerve conduction studies (NCSs), and somatosensory evoked responses (SERs) help differentiate cervical
spine disorders from peripheral nerve entrapment syndromes and help in differentiating intrinsic joint pathology from a radiculopathy. These tests are complementary to plain radiographs and an MRI or CT myelogram.

Neck pain typically occurs secondary to mechanical causes, and laboratory studies generally are not helpful in its diagnosis. However, laboratory studies can be critical in ruling out infection, neoplasm, and systemic arthritides. Erythrocyte sedimentation rate (ESR) indirectly measures the acute phase response with high sensitivity but low specificity. Patients younger than age 50 should have an ESR less than 20 mm/hr, with the accepted normal range increasing as the patient ages. Values above 100 are seen with infection and neoplasm, whereas less dramatic elevations are seen with rheumatoid arthritis and following surgery. 50 C-reactive protein (CRP) is an acute phase reactant synthesized by the liver; elevations peak by day 2 of the inciting event and return to normal within 3 to 7 days of removal of the insult. 51 Complete blood count (CBC) with differential and spinal tap can be helpful if meningitis is a concern.

DIFFERENTIAL DIAGNOSIS AND TREATMENT

By using the history, physical examination findings, and diagnostic studies, it is helpful to divide the differential diagnosis into benign axial neck pain, radiculopathy, myelopathy, infection, neoplasm, systemic arthritides, and referred pain. Most axial neck pain is self-limiting and will resolve with appropriate conservative care. 52 Axial neck pain with associated radiculopathy has a fairly benign course; 75% of patients have only one recurrence or mild symptoms at 19 years’ follow-up with conservative treatment. 53 Patients with isolated neck pain and a negative radiographic and laboratory workup are best treated with a multimodal approach. During the acute phase, patients can be treated with a soft collar to reduce inflammation, but this should not be worn for longer than 2 weeks so as to avoid deconditioning. Mimetic treatments, including proprioceptive training, exercise with resisted strengthening, muscle relaxers during the acute period, and nonsteroidal anti-inflammatory drugs (NSAIDs), have been found to be more effective in treating axial neck pain. 54-59 Evidence is inconclusive as to the effectiveness of radiofrequency denervation of facet joints, acupuncture, transcutaneous electrical nerve stimulation (TENS) treatment, iontophoresis, EMG biofeedback, or local injections for treatment of axial neck pain. 60-63

Cervical traction may be prescribed; a typical regimen includes 8 to 10 pounds for sessions of 15 to 20 minutes with the device at 20 to 25 degrees of flexion. However, this has resulted in only short-term relief of radicular symptoms. 64 Fluoroscopically guided interlaminar and transformaminal epidural steroids have been shown to be effective in treating lumbar radiculopathy, but this has not been shown in the cervical spine. 65 Cervical injections carry a higher risk, with complications, including neurologic deficits, reported in up to 16% of patients. 66,67 Given these increased risks with epidural steroid injections in the cervical spine, a more conservative approach should be taken in prescribing this treatment modality. Atlantoaxial facet joint osteoarthritis may be treated successfully with a facet block and nonsteroidal anti-inflammatory medications. If conservative therapy with NSAIDs fails, a fusion may be indicated.

In addition to degenerative changes to the cervical spine, trauma, and acute disk herniation, more insidious causes of neck pain include schwannoma, Pancoast tumor, brachial plexus neuritis, and complex regional pain syndrome. Schwannomas, if intradural, may involve a sensory nerve root, causing dermatomal pain in addition to a myelopathy or radiculopathy from compression. Pancoast tumor involving the lung apex may cause caudal cervical nerve root and sympathetic changes, in addition to nerve root or brachial plexus compression. Brachial plexus neuritis (Parsonage-Turner syndrome) is of viral origin, causing severe arm pain followed by weakness and then pain resolution; return of arm strength follows. This condition may progress to complex regional pain syndrome, along with autonomic changes such as discoloration of the skin, in a small number of cases associated with diffuse burning pain.

Follow-up and vigilance are in order because a progressive neurologic deficit, segmental instability, or persistent radicular symptoms for at least 6 weeks may be indications for surgical intervention. In a prospective randomized study comparing surgery, physical therapy, and cervical collar use for long-standing cervical radiculopathy, no difference was found between the three groups at 12 months. 66 Cervical myelopathy with very mild deficits can be followed closely; however, the natural course of the illness consists of long periods of stability with episodes of deterioration. Definitive indications for surgery include the presence of myelopathy for 6 months or longer, progression of signs or symptoms, difficulty walking, or changes in bowel or bladder function. Surgery is directed at decompressing the spinal cord and preventing further deterioration, rather than improving neurologic deficits.

Systemic arthritides, infection, and tumors can affect the cervical spine with variable neurologic and constitutional symptoms. Rheumatoid arthritis typically causes atlantoaxial subluxation, atlantoaxial impaction, and subaxial subluxation (Table 45-6). Surgical stabilization is indicated

### Table 45-6  Rheumatologic Disorders Causing Neck Pain

<table>
<thead>
<tr>
<th>Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Without disease of the C1-C2 joint</td>
</tr>
<tr>
<td>With structural cervical abnormalities</td>
</tr>
<tr>
<td>C1-C2 subluxation</td>
</tr>
<tr>
<td>C1-C2 facet involvement</td>
</tr>
<tr>
<td>Spondyloarthropathies</td>
</tr>
<tr>
<td>Ankylosing spondylitis</td>
</tr>
<tr>
<td>Reactive arthritis</td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
</tr>
<tr>
<td>Enteropathic arthritis</td>
</tr>
<tr>
<td>Polymyalgia rheumatic</td>
</tr>
<tr>
<td>Osteoarthritis</td>
</tr>
<tr>
<td>Fibromyalgia</td>
</tr>
<tr>
<td>Nonspecific musculoskeletal pain</td>
</tr>
<tr>
<td>Miscellaneous spondyloarthropathies</td>
</tr>
<tr>
<td>Whipple’s disease</td>
</tr>
<tr>
<td>Behçet’s disease</td>
</tr>
<tr>
<td>Paget’s disease</td>
</tr>
<tr>
<td>Acromegaly</td>
</tr>
<tr>
<td>Ossification of the posterior longitudinal ligament</td>
</tr>
<tr>
<td>Diffuse idiopathic skeletal hyperostosis</td>
</tr>
</tbody>
</table>
with progressive neurologic deficit, persistent axial neck pain with radiographic evidence of instability, canal diameter of 14 mm or less (posterior atlanto-dens interval), and odontoid migration of 5 mm or greater above McGregor’s line. In the setting of atlantoaxial subluxation or subaxial subluxation, the involved levels are fused posteriorly, and atlantoaxial impaction should be treated with occipitocervical fusion.\(^7\) Rheumatoid patients with evidence of instability demonstrate radiographic progression over time, but this correlates poorly with neurologic outcome.\(^69\) Regardless, these patients require close follow-up given that once myelopathy develops, most patients die within 1 year.\(^70\)

Ankylosing spondylitis commonly affects the cervical spine, resulting in a kyphotic deformity over time with altered biomechanics; this has implications in the setting of trauma. The kyphotic deformity can have significant functional implications because the person’s gaze moves toward the floor, making interaction with the surrounding world difficult. Corrective osteotomies are available, but risks include neurologic deficit and intraoperative bleeding. Hopefully, earlier diagnosis with the use of MRI and treatment with tumor necrosis factor–blocking agents will help make cervical deformity a thing of the past.\(^81\)

Both infection and neoplastic processes can cause destruction with mechanical or nonmechanical neck pain, constitutional symptoms, and variable neurologic deficit. The goals of treatment are similar in both with eradication of the infection or tumor, decompression if a neurologic deficit exists, and stabilization of the spinal column.

In summary, a careful history and physical examination and ancillary studies will help one arrive at a fairly narrow differential diagnosis. In the absence of spinal instability, neurologic deficit, and infectious or neoplastic processes, the patient may benefit from conservative treatment with expectant recovery with a “tincture of time.”

References


The references for this chapter can also be found on www.expertconsult.com.
Shoulder pain is one of the most common musculoskeletal complaints that may arise from diverse causes. Accurate diagnosis of shoulder pain is made difficult by the unique anatomy and position of the shoulder, which serves as a link between the upper extremity and the thorax. One of the most complex and mobile joints of the body, the shoulder is traversed by muscle, tendon, and bone, and is surrounded by major neurovascular structures, all of which may serve as potential sources of local and referred pain.

Determining the source of shoulder pain is essential in recommending the proper method of treatment. The examining physician must be able to differentiate the occurrence of shoulder pain caused by intrinsic, or local factors, and extrinsic, or remote factors, or a combination of the two. Intrinsic factors originate from the shoulder girdle and include glenohumeral and periarticular disorders, whereas extrinsic factors occur outside of the shoulder girdle with secondary referral of pain to the shoulder (Table 46-1). Examples of extrinsic factors include left shoulder pain as the initial presentation of coronary artery disease; hepatic, gallbladder, and splenic disease also may initially manifest as shoulder pain.

Accurate evaluation, diagnosis, and treatment require a thorough understanding of shoulder anatomy, including pain referral patterns. A complete and systematic physical examination is crucial for an accurate diagnosis. During the initial evaluation, care must be taken to discern all possible causes of shoulder pain. Final diagnosis may require repeated office examinations and correlation of diagnostic tests with symptoms and response to selective injections. Improvements in diagnostic tests, such as magnetic resonance imaging (MRI), computed tomography (CT)-arthrography, ultrasonography, and electromyography (EMG), have facilitated early diagnosis of shoulder pain and have provided a better understanding of shoulder pathology.

This chapter provides practical guidelines for the diagnosis and treatment of painful shoulder disorders that may be encountered in a rheumatology or general practice. A detailed analysis of shoulder problems and information on the treatment of major trauma are beyond the scope of this chapter and have been addressed by other authors.

ANATOMY AND FUNCTION

Because of its complexity, an understanding of the structural and functional anatomy of the shoulder is required for the clinician who is treating shoulder pain. The shoulder joint is the most mobile joint of the body, although mobility is gained at the sacrifice of stability. Only 25% of the humeral head surface has contact with the glenoid at any time. The labrum increases the contact area of the articular surface and confers stability to the joint. Lesions of the labrum may result from instability, and the type of lesion may indicate the type of instability. Labral tears also may be a source of pain from internal derangement of the shoulder. Joint stability is provided by a thin capsule and by the glenohumeral ligaments, which are thickenings of the capsule anteriorly, posteriorly, and inferiorly. Anterior stability is predominantly conferred by the anterior band of the inferior glenohumeral ligament.

The rotator cuff provides dynamic stability of the joint. It is composed of four musculotendinous units: the supraspinatus, infraspinatus, and teres minor posteriorly, and the subscapularis anteriorly. The shoulder consists of three joints: the acromioclavicular (AC), sternoclavicular, and glenohumeral joints, and two gliding planes—the scapulothoracic and subacromial surfaces.

Figure 46-1 shows the musculoskeletal and topographic localization of pain associated with common shoulder disorders. Figure 46-2 shows the relationship of the three posterior rotator cuff muscles coursing anteriorly underneath the acromion to insert on the greater tuberosity. The subscapularis, the only anterior rotator cuff muscle, inserts on the lesser tuberosity. By understanding the relationship between the rotator cuff and the subacromial region, bounded inferiorly by the humeral head and superiorly by the undersurface of the acromion, the clinician can visualize the problems of impingement syndrome and can accurately
Clinical Evaluation of the Shoulder

Accurate diagnosis and successful treatment of a shoulder disorder begin with a thorough history and physical examination. Most of the information needed to make a correct diagnosis can be elicited with basic clinical skills, rather than by relying on expensive and highly technologic investigative aids. Diagnostic tests should be used only to confirm an established diagnosis or to assist in cases with a challenging presentation.

History

In establishing a diagnosis, it is important to consider the patient’s age and chief complaint. The differential diagnosis of shoulder pain in a 70-year-old sedentary individual is entirely different from that in a 20-year-old pitcher. Did the pain occur slowly over time or suddenly with a particular event? Gradual onset of pain over the anterolateral or deltoid region that is increased with forward elevation of the shoulder and nocturnal pain suggest impingement with rotator cuff tendinopathy. The presence of significant weakness with pain on overhead activities suggests impingement with rotator cuff tear. Pain and weakness may also be noted with reaching behind the back with the shoulder in extension and external rotation, as when reaching into the back seat of a car. Initiating factors relative to the onset of symptoms should be elicited, and any history of shoulder pain or trauma should be carefully documented.

Pain intensity, character, location, and periodicity and aggravating or alleviating factors should be assessed. Pain should be graded on a visual analog scale of 0 to 10, with 0 indicating no pain, and 10 indicating the worst pain the patient has ever experienced. Another indication of the severity of pain is disruption of sleep. The patient should be asked whether the pain prevents sleep or awakens the patient, and whether the patient can lie on the affected shoulder. Is the pain sharp or dull? Sharp, burning pain over the top of the shoulder indicates a neurogenic origin, whereas a dull, aching pain over the lateral deltoid suggests rotator cuff pathology with impingement. Location or distribution of the pain should be identified. Is it local around the shoulder girdle, or does the pain radiate down the arm? Is concomitant sensory loss or weakness present? Periodicity of the pain as constant or intermittent should be determined, as should factors that aggravate or alleviate the pain.

Pain caused by rotator cuff tendinopathy usually is exacerbated by repetitive activities that involve the elbow away from the side of the body.

Any history of neck pain should be considered, along with history of radicular pain. Radicular-type pain frequently extends below the elbow and is associated with sensory loss and weakness. Pain located in the paracervical region may indicate a cervical origin, or it can be localized to the trapezius. Trapezial pain often is associated with shoulder pain and results from the patient trying to favor the shoulder. Assuming a military brace position may produce fatigue, spasm, and trigger points of the trapezius.

Any pertinent medical history, such as a history of malignancy, should be considered. Neurologic, visceral, and vascular disease can produce referred pain to the shoulder and should always be considered, especially in a patient with a painless range of motion.

Physical Examination

Proper physical examination of the shoulder includes close inspection of the shoulder girdle from the front and back. The evaluation is started by standing behind the patient, who has both shoulders exposed. The normal shoulder is always inspected and compared with the injured shoulder. Examination can be performed with the patient in the sitting or standing position. Contour and symmetry are observed and compared between shoulders, and any atrophy or asymmetry in shoulder position or level is assessed. Spinalesmus muscle atrophy may result from disuse, chronic cuff tear, or suprascapular or brachial neuropathy. If scapular winging is evident, the patient should be asked to do a wall push-up, which accentuates winging.

### Table 46-1 Common Causes of Shoulder Pain

<table>
<thead>
<tr>
<th>Intrinsic Causes</th>
<th>Periarticular Disorders</th>
<th>Glenohumeral Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotator cuff tendinitis</td>
<td>Calcific tendinitis</td>
<td>Inflammatory arthritis</td>
</tr>
<tr>
<td>Rotator cuff tear</td>
<td>Bicipital tendinitis</td>
<td>Osteoarthritis</td>
</tr>
<tr>
<td>Acromioclavicular arthritis</td>
<td></td>
<td>Arthritis</td>
</tr>
<tr>
<td>Glenohumeral instability</td>
<td></td>
<td>Regional Disorders</td>
</tr>
<tr>
<td>Extrinsics Causes</td>
<td>Regional Disorders</td>
<td>Extrinsic Causes</td>
</tr>
<tr>
<td>Cervical radiculopathy</td>
<td>Brachial neuritis</td>
<td>Glenohumeral instability</td>
</tr>
<tr>
<td>Nerve entrapment syndromes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sternocleidomastoid arthritis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflex sympathetic dystrophy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibrositis</td>
<td>Neoplasms</td>
<td>Miscellaneous</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>Miscellaneous</td>
<td>Gallbladder disease</td>
</tr>
<tr>
<td>Gallbladder disease</td>
<td>Splenic trauma</td>
<td>Subphrenic abscess</td>
</tr>
<tr>
<td>Splectic trauma</td>
<td>Subphrenic abscess</td>
<td>Myocardial infarction</td>
</tr>
<tr>
<td>Subphrenic abscess</td>
<td>Myocardial infarction</td>
<td>Thyroid disease</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>Thyroid disease</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Thyroid disease</td>
<td>Diabetes mellitus</td>
<td>Renal osteodystrophy</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>Renal osteodystrophy</td>
<td></td>
</tr>
<tr>
<td>Renal osteodystrophy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 46  |  SHOULDER PAIN

Figure 46-1  Musculoskeletal (A and B) and topographic (C and D) areas localizing pain and tenderness associated with specific shoulder problems.

1. Subacromial space (rotator cuff tendinitis/impingement syndrome, calcific tendinitis, rotator cuff tear).
2. Bicipital groove (bicipital tendinitis, biceps tendon subluxation and tear).
3. Acromioclavicular joint.
4. Anterior glenohumeral joint (glenohumeral arthritis, osteonecrosis, glenoid labrum tears, adhesive capsulitis).
5. Sternoclavicular joint.
6. Posterior edge of acromion (rotator cuff tendinitis, calcific tendinitis, rotator cuff tear).
7. Suprascapular notch (suprascapular nerve entrapment).
8. Quadrilateral space (axillary nerve entrapment). These areas of pain and tenderness frequently overlap.

Figure 46-2  A, Superior view of the rotator cuff musculature as it courses anteriorly underneath the coracoacromial arch to insert on the greater tuberosity. B, Anterior view of the shoulder reveals the subscapularis, which is the only anterior rotator cuff muscle inserting on the lesser tuberosity. It internally rotates the humerus and provides dynamic anterior stability to the shoulder. (A and B, From the Ciba Collection of Medical Illustrations, Volume 8, Part I. Netter Illustration from www.netterimages.com ©Elsevier Inc. All rights reserved.)
Range of motion should be carefully recorded, along with notation of any absence of rhythmic shoulder motion or excessive scapulothoracic motion that may compensate for the lack of glenohumeral motion. Internal rotation of the shoulder is checked by having the patient reach behind the back with the thumb while the examiner notices the vertebral level. Loss of internal rotation is seen early with shoulder pain and usually indicates some tightness of the posterior shoulder capsule. The biceps tendon; the coracoid, lesser, and greater tuberosities; and the posterior cuff are palpated, and any tenderness is gauged (Figure 46-3A). Tenderness on palpation of the long head of the biceps frequently is associated with rotator cuff tendinopathy and tenderness of the greater tuberosity. Any spasm or tenderness of the trapezius or levator scapulae may be associated with rotator cuff disease or cervical spine disease. Cervical range of motion and palpation of the parascapular muscles are carried out. Parascapular tenderness and limited range of motion of the neck may indicate cervical spondylosis or neurogenic disease. A Spurling test is done by flexing the neck laterally while applying axial compression to the skull. Pain that radiates to the ipsilateral shoulder is considered a positive test result and indicates radiculopathy.

To elicit the impingement sign, the shoulder is elevated passively in forward flexion, while the scapula is depressed with the opposite hand, forcing the greater tuberosity against the anterior acromion and producing pain in cases of impingement (Figure 46-3B). This maneuver also may be painful in conditions such as adhesive capsulitis, glenohumeral and acromioclavicular (AC) arthritis, glenohumeral instability, and calcific tendinitis. A dynamic impingement test, the circumduction-adduction shoulder maneuver, also called the Clancy test, is 95% sensitive and 95% specific for diagnosing rotator cuff tendinopathy, including partial tears. The test is done with the patient in the standing position and with the head turned to the contralateral shoulder. The affected shoulder is circumducted and adducted across the body to shoulder level, while the elbow is kept in extension, the shoulder in internal rotation, and the thumb pointing toward the floor.

**Figure 46-3** A, Tenderness on palpation of trigger points may help localize the site of pathology. Tenderness on palpation of the long head of the biceps and greater tuberosity suggests impingement with possible cuff tendinopathy. B, To elicit the impingement sign, the shoulder is elevated in forward flexion while the scapula is depressed with the opposite hand, forcing the greater tuberosity and the rotator cuff against the anterior acromion and producing pain when impingement exists. Relief of pain after injection of local anesthetics (i.e., impingement test) provides additional evidence of subacromial pathology. C, The Clancy test is performed with the patient standing and with the head turned toward the contralateral shoulder. The affected shoulder is circumducted and adducted across the body to shoulder level, keeping the elbow in extension with the arm internally rotated with the thumb pointed toward the floor. In this position, the patient is asked to resist maximally as a uniform downward force is applied to the extended arm by the examiner. Production of pain or weakness localized to the anterior lateral portion of the shoulder is considered a positive test result. D, The test is performed by forward flexion of the arm at 90 degrees and subsequent cross-chest adduction of the arm. Pain localized to the acromioclavicular joint is considered a positive test result. E, The test is performed with the patient’s elbow flexed at 90 degrees and held at the patient’s side by the examiner. The patient is asked to attempt external rotation of the shoulder from a neutral position (0 degrees of abduction) as the examiner applies resistance to the forearm. Strength is compared with that of the contralateral arm. F, Abduction strength testing is performed with the patient’s shoulder in 30 degrees of forward flexion and 90 degrees of abduction and with the thumb pointed toward the floor. The patient is asked to resist as the examiner exerts a downward force on the abducted arm. Strength is compared with the contralateral shoulder. (From Martin TL, Martin SD: Rotator cuff tendinopathy, Hosp Med 12:23–31, 1998.)
(Figure 46-3C). In this position, the patient is instructed to resist maximally as a uniform downward force is applied to the extended arm by the examiner. The test result is considered positive if pain or weakness is elicited during the maneuver, with pain localized to the anterolateral aspect of the shoulder. A strong positive correlation of pain and weakness is noted with complete cuff tear.10

The sternoclavicular and AC joints should be observed for prominences and palpated for stability and tenderness. Many patients with impingement have tenderness on direct downward palpation of the AC joint owing to impingement on the cuff from undersurface osteophytes of the distal clavicle.2,8

AC joint tenderness may also result from primary AC joint arthrosis and should be differentiated by physical examination, including the cross-cheek adduction test and O’Brien’s test.11 Radiographic evidence of AC joint arthrosis is common in patients older than 40 years, but this condition is not usually painful.12

The cross-cheek adduction test or the horizontal adduction test is performed by forward flexing the shoulder 90 degrees with subsequent cross-cheek adduction of the arm (Figure 46-3D). Pain localized to the AC joint is considered a positive test result. If pain occurs posteriorly over the shoulder, a tight posterior capsule with impingement is suspected. O’Brien’s test is performed by forward flexing the arm 90 degrees and adducting the arm 10 degrees out of the sagittal plane of the body. The first part of the test is performed with the hand maximally pronated with the thumb pointed down. In this position, the patient is asked to resist as the examiner applies a downward force on the arm. If the test elicits pain, the patient is asked if the pain is on top of the shoulder or deep inside. Pain localized to the top of the shoulder indicates AC joint pain, and pain deep inside the shoulder indicates a superior labrum anterior posterior (SLAP) lesion. In the second part of the test, the patient is asked to supinate the hand maximally, while the examiner applies a downward force to the arm. If the patient notices significantly less pain, the test result is positive for a SLAP lesion. If the pain is unchanged and is located on top of the shoulder, the test result is positive for AC joint pathology.11

If the cause of AC joint tenderness is still in question, a lidocaine injection should be administered. The clinician should carefully avoid injecting the subacromial space by advancing the needle too far inferiorly through the AC joint; this can lead to false interpretation. Painful degenerative changes of the AC joint may exist concomitantly with subacromial impingement and should be evaluated thoroughly when surgical treatment (i.e., distal clavicle excision) is being considered.11

In patients with pain out of proportion to objective findings, other causes of shoulder pain should be sought, including calcific tendinitis, infection, reflex sympathetic dystrophy, and fracture. Patients with significant wasting of the supraspinatus and infraspinatus muscles and posterior shoulder pain, especially younger patients, may have suprascapular neuropathy or brachial neuropathy (Parsonage-Turner syndrome).8,14

Patients with chronic cuff disease frequently have variable disuse atrophy of the supraspinatus and infraspinatus fossae; in cases of chronic massive cuff tears, atrophy and weakness can be severe. Strength testing of external rotation should be done with the elbow at the side and supported by the examiner; the patient is asked to attempt external rotation of the shoulder from a neutral position (0 degrees of adduction), while the examiner applies resistance (Figure 46-3E).16,17 Weakness in this position may suggest a tear of the infraspinatus tendon. Abduction strength testing against resistance is done with the shoulder in 30 degrees of forward flexion and 90 degrees of abduction, and with the thumb pointed toward the floor (Figure 46-3F).16,17 Weakness in this position may suggest a tear of the supraspinatus tendon. A lift-off test should be performed with the shoulder in internal rotation; the patient is asked to try to hold the hand away from the back. Inability to do so indicates a subscapularis tear.

If after a thorough physical examination impingement is suspected, an impingement test should be performed with injection of 5 mL of local anesthetic into the subacromial space.18,19 Before the test is performed, the patient is asked to grade the pain during the impingement signs on a visual analog scale of 0 to 10, with 0 equal to no pain and 10 equal to the most severe pain the patient has ever experienced. The injection may be done anteriorly, laterally, or posteriorly, depending on the physician’s preference. Ten minutes after injection of local anesthetic into the subacromial space, the patient should be re-examined and asked to regrade the pain on the same visual analog scale. A 50% or greater reduction in pain is thought to be a positive test result for impingement; otherwise, an alternative cause of shoulder pain should be sought, or inadequate placement of the anesthetic should be suspected. If the AC joint is thought to be contributing to the shoulder pain, 1 to 2 mL of local anesthetic should be injected into the joint, and the shoulder should be re-examined. When subacromial impingement and the AC joint are thought to be contributing to shoulder pain, serial injections during separate office visits may be needed to evaluate the shoulder while minimizing discomfort to the patient.12

In cases of suspected bicipital tendinitis, Speed’s test is performed by having the patient flex the shoulder and extend the elbow while a downward force is applied to the arm. The production of pain over the long head of the biceps is a positive test result and suggests bicipital tendinitis.

Upper extremity strength testing should be performed and compared with the contralateral side so that any atrophy is detected. Grip strength is checked, and the hands are examined carefully for evidence of intrinsic atrophy. The biceps (C5), triceps (C7), and brachioradialis (C6) reflexes are checked for symmetry and briskness.

Light touch sensory testing should be conducted, and the dermatomal distribution of any deficits that may suggest cervical radiculopathy should be identified. The cervical, supraclavicular, axillary, and epitrochlear regions should be palpated for enlarged lymph nodes, which may suggest malignancy.

Imaging

Radiographic Assessment

For nontraumatic painful shoulder evaluation, standard radiographic profiles are used. An impingement series
should be obtained, which includes anteroposterior views with a 30-degree caudal tilt (Rockwood view), an outlet view (scapular Y with 10- to 15-degree caudal tilt), and an axillary view. Internal and external rotational views may be obtained if calcific tendinitis or instability is suspected. The Rockwood view can reveal any osteophytes off the anterior acromion and AC joint. In cases of traumatic injury, a trauma series is obtained that includes a true anteroposterior view, a scapular Y view, and an axillary view. The axillary view is useful in assessing posterior or anterior subluxation of the humeral head. Additional views, such as the West Point view, which evaluates the glenoid for evidence of a bony Bankart lesion, or the Styker notch view, which assesses the humeral head for a Hill-Sachs lesion, may be obtained to assist evaluation if the diagnosis of instability is in doubt. Secondary impingement-type rotator cuff tendinitis may be caused by increased anterior translation with subluxation of the humeral head. In such cases, an axillary view or fluoroscopy can help show the subluxation.

When AC joint pathology is suspected, a 10-degree, cephalic tilt view of the AC joint at 50% penetration, as described by Zanca, should be obtained (Figure 46-4). Stress views of the AC joint may be obtained by strapping 5 to 10 lb of weight to the patient’s forearms and determining AC separation. Comparing the coracoclavicular distance of both shoulders may be helpful. When clinically indicated, cervical spine radiographs should be obtained to exclude cervical spondylosis as a cause of shoulder pain.

Scintigraphy

Te 99m methyl diphosphonate (MDP) or gallium may be of diagnostic help in evaluating skeletal lesions around the shoulder joint. Bone scans generally are not helpful in the diagnosis of non-neoplastic or noninfectious shoulder disease.

Scintigraphy may have a role in identifying patients with complete rotator cuff tears that proceed to cuff-tear arthropathy. This is an important distinction because patients with complete rotator cuff tears may do well, whereas those who develop progressive changes of cuff-tear arthropathy have progressive arthritis, pain, and significant functional impairment. Synovitis or calcium pyrophosphate deposition disease may be an important factor in the pathogenesis of cuff-tear arthropathy. In such cases, scintigraphy may show the increased blood flow and blood pooling associated with chronic synovitis.

Arthrography

Double-contrast arthrotomography (DCAT) can be used to evaluate problems of the rotator cuff, glenoid labrum, biceps tendon, and shoulder capsule. Figure 46-5 shows normal DCAT of the shoulder. Rotator cuff tears can be shown by single-contrast or double-contrast studies. Proponents of double-contrast arthrography believe that the extent of the tear, the preferred surgical approach, and the quality of the rotator cuff tissue are best determined by double-contrast studies. Arthrography without MRI or CT can be misleading and may result in underestimation of the extent of a rotator cuff tear. Multidetector CT can enhance the accuracy of diagnosing labral and rotator cuff tears, especially in patients for whom MRI is not possible (Figure 46-6).

Tears of the glenoid labrum without shoulder dislocation are sources of anterior shoulder pain in athletes. Glenoid labrum tears (Figure 46-7), with or without associated glenohumeral subluxation, frequently can be identified by DCAT. Kneisl and colleagues described 55 patients who underwent DCAT followed by diagnostic shoulder...
arthroscopy. DCAT predicted the arthroscopic findings in 76% of anterior labrum studies and 96% of posterior labrum studies. This test was 100% sensitive and 94% specific in diagnosing complete rotator cuff tears. Partial rotator cuff tears identified at arthroscopy were missed in 83% of patients undergoing DCAT. Investigators believed that DCAT was better in diagnosing intra-articular and cuff pathology in cases of instability than when pain alone was the presenting diagnosis.

Shoulder arthrography can confirm a diagnosis of adhesive capsulitis by showing a contracted capsule with an obliterated axillary recess (Figure 46-8). The use of subacromial bursography has been beneficial in visualizing the outer surface of the rotator cuff and the subacromial space in cases of impingement. Fukuda and associates reported a small series of younger patients (average age, 41.8 years) who underwent subacromial bursography after a negative glenohumeral arthrographic result. These patients showed pooling of contrast medium on the bursal side of a tear, which was confirmed at the time of surgery. Subacromial bursography is not routinely used diagnostically, and, in our opinion, it is of little value in planning surgical procedures.

Computed Tomography

CT is helpful in evaluating the musculoskeletal system, and CT combined with contrast arthrography (CT-arthrography) has become a major diagnostic tool for the evaluation of glenoid labrum tears, loose bodies, and chondral lesions (Figure 46-9). Rafii and co-workers reported using CT-arthrography in an evaluation of shoulder derangement. This study found 95% accuracy of CT-arthrography for investigating lesions of the labrum and articular surface. More recently, multidetector CT-arthrography scans have been used to evaluate partial cuff tears (Figure 46-10A), cystic lesions (Figure 46-10B), and calcific tendinopathy (Figure 46-10C).

Ultrasonography

Technologic improvements in ultrasound equipment have led to improved ultrasound study of the rotor cuff. The technique is noninvasive, is rapid, and involves no radiation exposure. The cuff is examined in the horizontal and transverse planes with the arm in different positions to allow visualization of various areas of the cuff. These techniques generally provide visualization of the distal cuff, where most rotator cuff tears are located. Figure 46-11 shows normal and abnormal ultrasound images of the rotator cuff in longitudinal and transverse planes.

Several studies report high sensitivity and specificity for the diagnosis of a rotator cuff tear by ultrasound. The specificity and sensitivity of the procedure are reported to be greater than 90% as determined by arthrographic and surgical correlations. This technique also has been used for the postoperative evaluation of a rotator cuff repair and for evaluation of abnormalities of the biceps tendon.
PART 6 | DIFFERENTIAL DIAGNOSIS OF REGIONAL AND DIFFUSE MUSCULOSKELETAL PAIN

Gardelin and Perin\textsuperscript{41} reported ultrasound to be 96\% sensitive in determining rotator cuff and biceps tendon pathology. Mack and associates\textsuperscript{36} found ultrasound to be valuable in evaluating postoperative patients with recurrent shoulder symptoms. In a prospective study, Hodler and colleagues\textsuperscript{39} compared ultrasound with MRI and arthrography in evaluating rotator cuff lesions in 24 shoulders. Ultrasound identified 14 of 15 torn cuffs, MRI identified 10 of 15, and arthrography identified 15 of 15.\textsuperscript{39} Ultrasound identified 7 of 9 intact rotator cuffs, whereas MRI was accurate in 8 of 9 intact cuffs.\textsuperscript{39} Vestring and colleagues\textsuperscript{42} found ultrasound to be as accurate as MRI in the diagnosis of humeral head defects and joint effusions, but inferior to MRI in the diagnosis of labrum lesions, rotator cuff lesions, subacromial spurs, and synovial inflammatory disease. In the hands of an experienced sonographer, ultrasound may be the most cost-effective test for the initial evaluation of a rotator cuff injury, but most surgeons require CT-arthrography or MRI confirmation before beginning surgical exploration.\textsuperscript{36,39,41-43}

\textbf{Arthroscopy}

The use of arthroscopy for the diagnosis of shoulder pathology increased in the 1980s, in part because of its accuracy, which was far greater than that of clinical examination and better than the accuracy of other diagnostic modalities of the time. With technologic advances in fiberoptics, video output, and arthroscopic instrumentation, the use of arthroscopy to diagnose and treat shoulder problems exponentially increased to include procedures previously used only for open techniques.\textsuperscript{44}

Compared with DCAT, arthroscopy is more accurate in the diagnosis of intra-articular lesions associated with a painful shoulder.\textsuperscript{10} An additional benefit is that arthroscopy can be used to diagnose and treat shoulder problems of the glenohumeral joint and the subacromial region. With increased accuracy of MRI-arthrography in detecting partial cuff tears and labral lesions, diagnostic shoulder arthroscopy has become less common in the absence of clear indications and specific treatment plans. In combination with a detailed history and physical examination, and along with examination under anesthesia, shoulder arthroscopy has been helpful in the diagnosis of chronic instability patterns of the glenohumeral joint.\textsuperscript{44-47}

The indications and usefulness of shoulder arthroscopy in the treatment of common pathologic conditions have continued to increase as the technology improves, and as understanding of the pathophysiology of shoulder problems grows. Shoulder arthroscopy has been used routinely to confirm and treat SLAP lesions, labral tears, partial cuff

\textbf{Figure 46-9} CT-arthrography of the shoulder. \textbf{A}, Normal findings. \textbf{B}, Tear of the anterior glenoid labrum. \textbf{C}, Large defect of the articular surface of the posterior portion of the humeral head (Hill-Sachs lesion) (arrow). \textbf{D}, Loose body in the posterior recess (arrow).
Characteristic MRI findings in rotator cuff tears include a hypointense gap within the supraspinatus muscle tendon complex on T1-weighted films, absence of a demonstrable supraspinatus tendon with narrowing of the subacromial space, and an increased signal within the supraspinatus tendon on T2-weighted images.

Seeger and colleagues, reporting the results of 170 MRI studies, found that T1-weighted images were highly sensitive for identifying abnormalities within the supraspinatus tendon, but T2-weighted images were required to differentiate tendinitis from a small supraspinatus tendon tear. Large full-thickness tears could be identified, however, on T1-weighted and T2-weighted images.

Figure 46-12 depicts common shoulder pathology as seen by MRI. MRI is almost as sensitive as CT-arthrography in the diagnosis of abnormalities of the glenoid labrum.

Characteristic MRI findings in rotator cuff tears include a hypointense gap within the supraspinatus muscle tendon complex on T1-weighted films, absence of a demonstrable supraspinatus tendon with narrowing of the subacromial space, and an increased signal within the supraspinatus tendon on T2-weighted images.

Electromyography and Nerve Conduction Velocity Studies

EMG and nerve conduction velocity studies can help differentiate shoulder pain from pain of neurogenic origin. They also may be beneficial in determining the localization of neurogenic pain to a particular cervical root, the brachial plexus, or a peripheral nerve.
Intra-articular pathology (e.g., labrum tear, capsular tear, loose body, chondral defect) is suspected, MRI-arthrography is preferable to CT-arthrography. In diagnosing acute rotator cuff tears in a younger patient, ultrasound is the most cost-effective test to confirm a clinical suspicion. Although not as sensitive as the more sophisticated tests, plain radiographs can identify arthritic change, calcific tendinitis, established osteonecrosis, and most neoplasms.

If intra-articular pathology (e.g., labrum tear, capsular tear, loose body, chondral defect) is suspected, MRI-arthrography is preferable to CT-arthrography. In diagnosing acute rotator cuff tears in a younger patient, ultrasound is the most cost-effective test to confirm a clinical suspicion. In cases of impingement syndrome, MRI is sensitive, but it is difficult to differentiate tendinitis, partial tears, and small complete tears without MRI-arthrography. Orthopedic surgeons prefer MRI-arthrography for verification of labral tears or partial rotator cuff tears. In the case of a suspected full-thickness rotator cuff tear, MRI is preferred to determine the size of the tear, the amount of muscle atrophy and tendon retraction, and the quality of remaining tissue for repair.
CHAPTER 46 | SHOULDER PAIN

acromion, coracoacromial ligament, coracoid process, or AC joint on the rotator cuff as it passes beneath them during glenohumeral motion. The function of the posterior rotator cuff is to abduct and externally rotate the humerus. The cuff with the biceps tendon serves as a humeral head depressor to maintain the head centered within the glenoid fossa as the cuff and to use the deltoid to elevate the arm. Controversy continues, however, as to the exact cause of impingement, that is, whether it is a primary, intrinsic, degenerative event within the tendon with superior migration of the head on arm elevation and secondary

**INTRINSIC FACTORS CAUSING SHOULDER PAIN**

**Periarticular Disorders**

**Shoulder Impingement and Rotator Cuff Tendinopathy**

One of the most common nontraumatic causes of shoulder pain is impingement with rotator cuff tendinopathy. In 1972, Neer described his results of 100 anatomic shoulder dissections and coined the term *impingement syndrome*. Impingement may be defined as the encroachment of the

---

**Figure 46-12**  
A, Magnetic resonance imaging (MRI) proton density–weighted coronal view shows the supraspinatus tendon as a black band (A) that has an increased signal as it nears insertion on the greater tuberosity (B).  
B, Similar view in a T2-weighted image shows increased signal as gray (arrow), indicating a partial-thickness tear or tendinitis.  
C, MRI proton density–weighted coronal view shows abrupt end of supraspinatus tendon as it courses right to left (A). From A to B is an area of increased signal followed by a short portion of tendon (B) inserting at the greater tuberosity.  
D, Similar view in a T2-weighted image shows increased signal as white (fluid density), indicating fluid in the gap of a complete rotator cuff tear.  
E, MR arthrography shows a normal rotator cuff.  
F, MR arthrography shows a chronic cuff tear with retraction.
Impingement on the acromion, or purely mechanical attrition of the tendon with primary impingement against the acromion. The mechanical impingement of the rotator cuff may be influenced by variations in the shape and slope of the acromion. The supraspinatus outlet may become narrowed from proliferative spur formation of the acromion or degenerative changes in the AC joint. These changes, along with intrinsic degenerative changes of the rotator cuff, may lead to rotator cuff tear, but the exact pathogenesis remains controversial. Many studies have found a strong correlation between degenerative hypertrophic spur formation, with its resulting narrowing of the supraspinatus outlet, and the presence of full-thickness cuff tears, but clinical studies have failed to confirm whether hypertrophic changes in the coracoacromial arch are caused by the cuff lesions, or whether these changes themselves cause the lesions.

Neer developed a staging system for description of impingement lesions of the shoulder. A stage I lesion involves edema and hemorrhage of the rotator cuff and is typically found in individuals younger than 25 years who are active in overhead athletics. The condition usually responds to conservative treatment that includes rest, anti-inflammatory medication, and physical therapy. Stage II lesions usually occur in the 30s or 40s and represent the biologic response of fibrosis and thickening of the tendon after repeated episodes of mechanical impingement over time. Lesions are treated conservatively, as in stage I, but attacks may recur. If symptoms persist despite adequate conservative management for longer than 6 to 12 months, surgical intervention is warranted. Stage III lesions involve rotator cuff tears, biceps tendon rupture, and bone changes, and they rarely occur before age 40. Patients may present with pain, weakness, or supraspinatus atrophy, depending on the chronicity of the tear. Surgical treatment depends on the patient’s age, loss of function, weakness, and pain.

Patients usually present to the clinician with a complaint of pain that has failed to resolve after a variable period. Pain can be sudden and incapacitating in cases of traumatic cuff tears, or more commonly may manifest as a dull ache in cases of chronic impingement. Pain usually is located over the anterior and lateral aspects of the shoulder and may radiate into the lateral deltoid. It may worsen with sleeping on the affected extremity and is exacerbated by overhead activity. Tenderness on palpation may be elicited over the greater tuberosity and the long head of the biceps within the bicipital groove, indicating an associated biceps tendinitis. In cases with concomitant degenerative changes in the AC joint, tenderness may be noted on palpation over the AC joint, as an offending osteophyte impinges on the rotator cuff beneath.

The impingement sign as described by Neer (Figure 46-13) is useful in the diagnosis of rotator cuff tendinopathy. The patient often describes a catch as the arm is brought into the overhead position. The patient may be observed to raise the arm by abduction and external rotation to clear the greater tuberosity of the acromion, bypassing the painful area. A typical painful arc usually occurs between 70 degrees and 110 degrees of abduction. Neer also described an impingement test that involves injection of lidocaine into the subacromial bursa. Relief of pain is a positive impingement test result and usually indicates rotator cuff origin of the shoulder pain.

**Figure 46-13** The impingement sign is elicited by forced forward elevation of the arm. Pain results as the greater tuberosity impinges on the acromion. The examiner’s hand prevents scapular rotation. This maneuver may be positive in other periarticular disorders. (From Neer CS II: Impingement lesions, Clin Orthop Relat Res [173]:70, 1983.)
Radiographs in the early stages of cuff tendinopathy may be normal or may reveal a hooked acromion. As the disease progresses, sclerosis, cyst formation, and sclerosis of the anterior third of the acromion and the greater tuberosity may be observed. An anterior acromial traction spur may appear on the undersurface of the acromion lateral to the AC joint and represents contracture of the coracoacromial ligament. Late radiographic findings include narrowing of the acromiohumeral gap, superior subluxation of the humeral head in relation to the glenoid, and erosive changes in the anterior acromion. Arthrography, MRI, and ultrasound may be helpful in diagnosing a full-thickness tear of the rotator cuff in association with stage III disease. In some cases of chronic large rotator cuff tears, proximal migration of the humeral head leads to a pattern of degenerative arthritis termed cuff-tear arthropathy.

The choice of treatment and frequently its result are functions of the stage of the impingement and the response to pain. In stage I disease, in which little mechanical impingement occurs, most patients respond to rest. It is important to avoid immobilizing the shoulder for any period because con traction of the shoulder capsule and periarticular structures can produce an adhesive capsulitis. After a period of rest, a progressive program of stretching and strengthening exercises generally restores the shoulder to normal function. Use of aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) may shorten the symptomatic period. Modalities such as ultrasound, nerve probe, and transcutaneous electrical nerve stimulation generally are not helpful. Patients with stage I or II disease may have a dramatic response to local injection of glucocorticosteroids and local anesthetic agents. For stage II disease in which fibrosis and thickening occur anteriorly, it is frequently better to inject through a posterior approach. We prefer a combination of 3 mL of 1% lidocaine (Xylocaine), 3 mL of 0.5% bupivacaine, and 20 mg of triamcinolone. This injection combines a short-acting anesthetic to help confirm the diagnosis, a longer-acting anesthetic for analgesic purposes, and a steroid preparation in a depot form.

An integrated program of occupational and physical therapy often precludes the need for surgery in patients with stage II disease. Job modification for individuals with impingement syndrome caused by overuse may alleviate symptoms. Businesses are becoming increasingly aware of the cost savings associated with proper job ergonomics.

The initial rehabilitation in stage II impingement consists of cessation of repetitive overhand activity. Ice, NSAIDs, and local injections also may be beneficial. Initial physical therapy includes passive, active-assisted, and active range of motion combined with stretching and mobilization exercises to prevent contracture. As pain and inflammation subside, isometric or isotonic exercises are used to strengthen the rotator cuff musculature. Isokinetic training at variable speeds and in variable positions is instituted before the patient is returned to full activity. For patients with a job-related injury, it is crucial to review and modify job mechanics to prevent recurrent episodes that can cause further disability and may precipitate the need for surgery.

Neer suggested that a patient with refractory stage II disease may respond to division of the coracoacromial ligament and bursectomy of the subacromial bursa. Open anterior acromioplasty as described by Neer has become accepted as the procedure of choice for stage II and III impingement lesions, with many investigators reporting high success rates in treating impingement syndrome and rotator cuff tears.

Reported results show good and excellent relief of symptoms in 71% to 87% of patients treated by the open surgical procedure.

In 1985, Ellman described the technique of arthroscopic subacromial decompression. His initial results and the results of others are comparable with those of open surgical techniques. Arthroscopic subacromial decompression has become a widely accepted treatment for refractory stage II and III impingement lesions. The procedure can be done as an outpatient surgery, and because no deltoid is detached, as with the open technique, the procedure facilitates rehabilitation and increases overall recovery rates.

Calcific Tendinitis

Calcific tendinitis is a painful condition around the rotator cuff that is associated with deposition of calcium salts, primarily hydroxyapatite. The cause of calcific tendinitis is unknown. The commonly accepted cause is degeneration of the tendon, which leads to calcification through a dystrophic process. A common clinicopathologic correlation is seen in three distinct phases of the disease process: the precalcific or formative phase, which can be relatively painless; the calcific phase, which tends to be quiescent and may last months to years; and the resorptive or postcalcific phase, which tends to be painful, as calcium crystals are resorbed.

Although it is more common in the right shoulder, at least a 6% incidence of bilateralism has been reported. Patients with bilateral shoulder involvement often have the syndrome of calcific periarthritis, in which calcium hydroxyapatite crystals are found at multiple sites. Patients usually present with impingement-type pain in the affected shoulder during overhead activity. The pain may seem to be out of proportion to any objective physical findings. The patient may describe difficulty sleeping on the shoulder and trouble falling asleep. Symptoms may last a few weeks or a few months.

The incidence of calcific tendinitis varies in the literature among asymptomatic individuals from 2.7% to 20%. Most calcification occurs in the supraspinatus tendon, and 57% to 76.7% of patients are women. The average age of patients is 40 to 50 years.

Codman pointed out the localization of calcification within the tendon of the supraspinatus. He provided a detailed description of the symptoms and the natural history of this condition. In describing the phases of pain, spasm, limitation of motion, and atrophy, he noted the lack of correlation between symptoms and the size of the calcific deposit. According to Codman, the natural history includes degeneration of the supraspinatus tendon, calcification, and eventual rupture into the subacromial bursa. During the latter phase, pain and decreased motion can lead to adhesive capsulitis (see Figure 46-8).

Several factors may affect localization of calcium within the supraspinatus. Many patients have an early stage of impingement, which compresses the supraspinatus tendon on the anterior portion of the acromion. This longstanding impingement may lead to local degeneration of tendon fibers. In patients without impingement,
localization of calcium within the supraspinatus may be related to the blood supply of the rotator cuff, which normally is derived from an anastomotic network of vessels from the greater tuberosity or from the bellies of the short rotator muscles.84 The watershed of these sources is just medial to the tendinous attachment of the supraspinatus.85 Rathburn and Macnab86 referred to this watershed as the critical zone and pointed out that during abduction this area was rendered ischemic.

Treatment of calcific tendinitis depends on the clinical presentation and the presence of associated impingement. Patients can have an acute inflammatory reaction that may resemble gout. The acute inflammation can be treated with local glucocorticoid injection, NSAIDs, or both. Ultrasound may be beneficial. If impingement is associated, treatment depends on the stage at presentation. The radiographic appearance of the calcification can direct and perhaps predict the response to therapy. In the resorptive state, deposits appear floccular, suggesting that the process is in the phase of repair, and that a conservative program is indicated.

Patients with discrete calcification and perhaps associated adhesive capsulitis (see Figure 46-8) may be at a stable phase, in which calcium produces a mechanical block and is unlikely to be resorbed. For these patients, mechanical removal of calcific deposits and correction of associated pathologic lesions may be necessary.85,86 Percutaneous disruption of calcified areas may be performed using a needle directed by fluoroscopy. This technique allows lavage and injection, but does not treat associated impingement. Subacromial arthroscopy allows mechanical débridement of calcific deposits under direct visualization. This technique can be combined with arthroscopic removal of the inflamed bursa and decompression of associated impingement. Improved results have been noted with complete removal of calcific deposits.85 In many cases of refractory calcific tendinitis associated with impingement, open or arthroscopic acromioplasty, subacromial bursectomy, and decompression are indicated.

**Rotator Cuff Tear**

**Pathophysiology**

Spontaneous tear of the rotator cuff in an otherwise normal individual is rare.19 It can occur in patients with rheumatoid arthritis or systemic lupus erythematosus as part of the pathologic process with invasion from underlying pannus. Metabolic conditions such as renal osteodystrophy and agents such as glucocorticoids occasionally are associated with cuff tears. Most patients report a traumatic episode, such as falling on an outstretched arm or lifting a heavy object. The usual presenting symptoms are pain and weakness of abduction and external rotation. Crepitus and even a palpable defect may be associated. Long-standing tears generally are associated with atrophy of the supraspinatus and infraspinatus muscles. It may be difficult to differentiate a painful tendinitis from a partial-thickness or a small full-thickness cuff tear.

Controversy continues about the exact cause of cuff tendinopathy.86,92,94,98 Most likely, the pathophysiology involves a combination of factors, including decreased vascularity and cellularity of the tendon, along with changes in the collagen fibers of the tendon that occur with aging.

Loss of motion with subsequent capsular tightening, particularly in the posterior capsule, may lead to cephalad migration of the humeral head, with subsequent impingement of the cuff under the coracoacromial arch.96 Rehabilitation exercises stress regaining a normal range of motion. To achieve full, painless motion, the normal relationship of glenohumeral to scapulothoracic motion must be achieved.16,17,97

**Diagnosis**

**History.** Patients with nontraumatic tears of the rotator cuff report symptoms of chronic impingement. Loss of motion and a feeling of stiffness are often noted with extremes of motion, along with difficulty during activities of daily living, such as combing the hair, hooking a bra strap, putting on a shirt or coat, and reaching into the back pocket. In chronic cases of cuff tendinopathy, loss of motion usually occurs. Limitation of internal rotation occurs initially, is caused by posterior capsular contracture, and is often associated with posterior shoulder pain with adduction of the ipsilateral shoulder. Further shoulder impingement occurs with forward flexion because of superior migration of the humeral head against the anterior inferior acromion. This upward translation is analogous to the action of a yo-yo climbing on a string.96,98 Over time, loss of forward flexion, abduction, and external rotation occurs with passive and active motion of the shoulder.

**Imaging.** In acute cases, a history of trauma, such as a fall onto the affected shoulder, may be reported. In cases involving an anterior shoulder dislocation with subsequent profound weakness of the rotator cuff, a large cuff tear or a greater tuberosity avulsion should be suspected, in addition to axillary nerve palsy. In younger patients, traumatic failure of the cuff under tensile overload may result in cuff failure caused by forced adduction of the affected shoulder or active abduction against resistance, and this may occur with traumatic dislocation. Repetitive tensile overload also can result in partial rotator cuff tears in an overhead athlete.

Plain radiographs are used in initial evaluation of impingement-type shoulder pain with cuff tendinopathy. An impingement series should be ordered, including an anteroposterior radiograph with a 30-degree cephalic tilt (Rockwood view), which can reveal osteophytes of the anterior or acromion and AC joint; a scapular Y view with a 10-degree cephalic tilt (supraspinatus outlet view), which can evaluate the type of acromion and reveal anterior and AC osteophytes; and an axillary view, which can evaluate the acromion for possible os acromiale. Calcific deposits within the rotator cuff tendon can be viewed best with rotational anteroposterior radiographs. Cuff arthropathy should be suspected if the acromial-humeral distance is less than 7 mm, or with the presence of cyst formation within the greater tuberosity, humeral head osteopenia, sclerosis around the greater tuberosity, or humeral head collapse. In advanced stages of cuff arthropathy, complete loss of glenohumeral joint space may be seen with superior migration and abutment of the humeral head against the undersurface of the acromion.59
In the past, shoulder arthrography was considered the “gold standard” for diagnosing full-thickness and partial-thickness rotator cuff tears, with greater than 90% sensitivity and specificity.\textsuperscript{31,99} Currently, arthrography with CT or MRI is routinely used to diagnose rotator cuff pathology, including full-thickness and partial-thickness tears.

Ultrasoundography has been accurate in the diagnosis of full-thickness rotator cuff tears.\textsuperscript{39,100-103} Ultrasoundography offers the advantages of being inexpensive and noninvasive, but disadvantages include unproven effectiveness in determining subacromial impingement, capsular and labral abnormalities, and partial cuff tears. The procedure and its results are technician dependent. Ultrasoundography may have a useful role in determining the postoperative integrity of the cuff repair.\textsuperscript{38}

MRI has been invaluable in evaluating rotator cuff tears. Sensitivity and specificity of MRI for diagnosing full-thickness cuff tears are 100% and 95%,\textsuperscript{104} Through the use of gadolinium or saline, partial tears that are otherwise difficult to detect with conventional imaging can be detected.

Diagnosing cuff tears with MRI usually is based on discontinuity of the tendon on T1-weighted images and consistency with fluid signal on T2-weighted images. Ancillary findings include fluid in the subacromial space on T2-weighted images, loss of the subacromial fat plane on T1-weighted images, and proliferative spur formation of the acromion or AC joint. Large, chronic cuff tears also may be associated with cephalad migration of the humeral head and fatty atrophy of the spinatus muscle. Periarticular soft tissues, including the capsulolabral complex and the biceps tendon, as well as the rotator cuff can be thoroughly examined. The degree of tear and tendon retraction and evidence of muscle atrophy can be evaluated, all of which are crucial in preoperative planning for possible cuff repair.

\textbf{Treatment}

\textbf{Nonsurgical Treatment.} Codman and Akerson\textsuperscript{104} recommended early operative repair for acute full-thickness rotator cuff tears and reported the first documented repair in 1911. McLaughlin\textsuperscript{106} recommended early repair in cases of grossly displaced tuberosity fractures or massive tears. Several other clinical studies have supported the concept that a full-thickness tear does not preclude good shoulder function. DePalma\textsuperscript{105} reported that 90% of patients with rotator cuff tears responded to conservative measures, such as rest, analgesics, anti-inflammatory agents, and physiotherapy.

The reported percentage of patients responding to nonsurgical treatment in the literature varies from 33% to 90%.\textsuperscript{3,18,106-109} Conservative treatment includes pain control with NSAIDs, ultrasound, heat before shoulder stretching and exercise, and ice after overhead activity. Deep massage therapy is employed to reduce trigger point tenderness within the trapezius, levator scapulae, and periscapular muscles. Patients on long-term anti-inflammatory medications are monitored periodically for evidence of gastrointestinal bleeding and for hepatic or renal toxicity. Opiate-based drugs are used only in the acute setting, such as after a fall, or in the perioperative period.

Steroid and local anesthetic injections are used when the patient has significant pain that prohibits rehabilitation. Injections may be repeated once every 3 months if needed; injection into the cuff tendon is to be avoided. If the patient fails to improve after 3 months of conservative treatment, or does not continue to improve after three sequential injections, surgical options should be discussed.

The mainstay of conservative therapy is exercise. Rehabilitation stresses pain relief with exercises aimed at restoring shoulder motion and strengthening remaining cuff muscles, deltoid, and scapular stabilizers. Therapy can be divided into three phases. The goals of the initial phase of therapy are to relieve pain and restore shoulder motion. Motion therapy includes pendulum exercises, passive motion with use of a wand with assistance of the uninvolved shoulder, an overhead pulley system, and posterior capsular stretching. The arc of motion is gradually increased and is guided by the patient’s discomfort to avoid painful impingement arcs.

The second phase of therapy is entered after the patient has return of motion and little discomfort with overhead activity. Emphasis is placed on strengthening the remaining rotator cuff musculature and deltoid and periscapular muscles. Strengthening with elastic surgical tubing provides variable degrees of resistance, depending on the size of the tubing. Initial strengthening is performed out of the impingement arc (70 to 120 degrees of shoulder flexion). The goal of this phase is to strengthen the shoulder to prevent dynamic proximal humeral migration with impingement during active shoulder elevation.\textsuperscript{39,61} Normal shoulder kinematics relies on combined and synchronous glenohumeral flexion and scapular rotation.\textsuperscript{62,69} In addition to strengthening the cuff and deltoid, the scapular rotators, including the trapezius and the serratus anterior muscles, are emphasized.\textsuperscript{107}

After the patient has successfully completed phase two of the rehabilitation program with minimal symptoms and good shoulder function, the final phase is entered. Phase three is characterized by a gradual return to normal overhead activities, including work and sporting activities. This part of the rehabilitation program should be tailored to the individual patient’s needs and the demands placed on the shoulder.

\textbf{Surgical Treatment.} A Cochrane review of the effectiveness of surgery for rotator cuff disease failed to reach any firm conclusions about the effectiveness or safety of rotator cuff surgery.\textsuperscript{108} Severity and duration of pain are the primary indications for surgical intervention in a rotator cuff tear. Other factors important in surgical decision making include shoulder dominance, activity level, physiologic age, acuteness of the tear, degree of tear, loss of function, amount of tendon retraction, and fatty atrophy of the remaining cuff musculature.

A systematic review of indications for rotator cuff surgery found that earlier surgical intervention may be needed for patients with cuff tears with weakness and significant functional disability. In addition, older chronologic age did not portend a worse outcome; however, pending workman’s compensation claims, it did negatively affect treatment results.\textsuperscript{109}

\textbf{Acute Tears.} Acute tears of the rotator cuff can be treated with conservative measures of periscapular and cuff strengthening along with capsular stretching to restore
motion. Early surgical intervention should be considered in a young patient, especially an overhead athlete. Conservative shoulder rehabilitation should be maintained for 3 to 6 months before a decision is made regarding surgery for an older sedentary patient, in whom functional results without surgery may be acceptable. Many older patients may function well with chronic cuff tears, but they may become debilitated if an acute tear is superimposed on chronic changes. Surgical intervention may be required in these cases to return the patient to baseline function by repairing the acute tear and attempting to repair the chronic tear if possible.

Chronic Tears. For elderly patients whose pain and weakness do not create a functional problem, a conservative program is preferable for chronic tears. Pain unresponsive to conservative management is the main indication for surgery in an older patient with a chronic rotator cuff tear. In these cases, surgery should be considered on an individual basis after at least 3 months of conservative treatment, including subacromial steroid injection. If the cuff tear is massive and irreparable, débridement and subacromial decompression may provide good pain relief without extensive surgery and prolonged immobilization. In a younger patient with a chronic tear and weakness, surgery to repair the cuff may be indicated to improve strength and prevent further extension of the tear.

In cases of rotator cuff arthropathy with glenohumeral joint degeneration, a reverse total shoulder replacement may be indicated. This type of total shoulder replacement reverses the normal relationship between scapular and humeral components, moving the center of rotation medially and distally to increase the lever arm length of the deltoid muscle. The deltoid compensates for the deficient rotator cuff, allowing as near-normal function as possible (Figure 46-14). A recent publication also suggested reverse total shoulder arthroplasty for the treatment of irreparable rotator cuff tear with disability and no glenohumeral arthritis.

Bicipital Tendinitis and Rupture

The long head of the biceps passes through the bicipital groove, crosses over the head of the humeral, and inserts on the superior rim of the glenoid (see Figure 46-1A). The biceps tendon aids in flexion of the forearm, supination of the pronated forearm if the elbow is flexed, and forward elevation of the shoulder. Bicipital tendinitis, subluxation or dislocation of the biceps tendon within the bicipital groove, and rupture of the long head of the biceps generally are associated with anterior shoulder pain.

Bicipital tendinitis is sometimes an associated feature of a rotator cuff tear. The rotator cuff tear compromises centering of the humeral head on the glenoid. This compromise results in increased mechanical loading of the long head of the biceps, which initiates a hypertrophic tendinitis. Dislocation of the long head of the biceps usually is combined with a lesion of the subscapularis tendon. Isolated rupture of the long head of the biceps tendon is rare when the rotator cuff is intact. Rupture of the long head of the biceps is common, however, when a coexisting rotator cuff tear is present. The effects of rotator cuff tear and concomitant biceps tendon rupture on strength can be substantial.

Early phases of bicipital tendinitis are associated with hypervascularity, edema of the tendon, and tenosynovitis. Persistence of this process leads to adhesions between the tendon and its sheath, along with impairment of the normal gliding mechanism in the groove. Stretching of the adhesions may be associated with chronic bicipital tendinitis. The diagnosis of bicipital tendinitis is based on localization of tenderness. It is often confused with impingement symptoms and is frequently seen with an impingement syndrome. Isolated bicipital tendinitis can be differentiated by the fact that the tender area migrates with the bicipital groove as the arm is abducted and externally rotated. Many eponyms are associated with tests to identify bicipital tendinitis. Yergason’s supination sign refers to pain in the bicipital groove when the examiner resists supination of the pronated forearm with the elbow at 90 degrees. Ludington’s sign refers to pain in the bicipital groove when the patient interlocks the fingers on top of the head and actively abducts the arms.

Biceps tendon rupture can occur in some patients who report no history of shoulder pain. Patients often complain of an acute onset of pain and ecchymosis around the anterior shoulder and sagging of the biceps muscle belly. In these cases, a concomitant rotator cuff injury should be excluded by clinical examination. More often, the biceps tendon rupture is preceded by painful shoulder symptoms that often improve or disappear after the rupture.

Treatment generally is conservative and consists of rest, analgesics, NSAIDs, and local injection of glucocorticoids. The use of ultrasound and a neuroprobe is more beneficial in this condition than in isolated rotator cuff tendinitis. Patients with refractory bicipital tendinitis and recurrent symptoms of subluxation are treated by arthroscopic biceps tenodesis or open tenodesis, that is, opening the bicipital
groove and resecting the proximal portion of the tendon with tenodesis of the distal portion into the groove or beneath the pectoralis tendon.

**Acromioclavicular Disorders**

The AC joint is a common source of shoulder pain. Acute causes of AC joint pain are often related to direct trauma of the affected shoulder that may result in a distal clavicle injury with an intra-articular chondral fracture, or in AC joint instability from ligamentous disruption.  

Post-traumatic distal clavicle osteolysis, with resorption of the distal clavicle, may ensue 4 weeks after a shoulder injury, leading to AC joint pain. Osteolysis may be caused by microfracture of the subchondral bone and subsequent attempts at repair. Other authors believe the cause to be an autonomic nerve dysfunction affecting the blood supply to the clavicle. The increased blood supply leads to resorption of bone from the distal clavicle. More commonly, chronic osteolysis results from repetitive microtrauma to the AC joint from activities such as weight lifting, gymnastics, and swimming.

The underlying pathophysiology is believed to be an inflammatory process caused by stress fracture of the subchondral bone with hyperemic resorption of the distal clavicle. Other causes of osteolysis include rheumatoid arthritis, hyperparathyroidism, and sarcoidosis, which should be considered in the differential diagnosis, especially in bilateral cases. Patients with atraumatic osteolysis of the distal clavicle should be forewarned that bilateral involvement may occur; an incidence of 70% was reported for one long-term follow-up. Other chronic causes of AC pain include idiopathic, intra-articular disk pathology, post-traumatic degenerative arthrosis from joint incongruity, primary degenerative arthrosis, and rheumatoid arthrosis.

Evaluation should always include a detailed history, physical examination, and radiographic evaluation. A history of trauma to the AC joint from a direct fall or blow to the ipsilateral shoulder may be reported. Less commonly, the AC joint may have been injured indirectly, as during a fall on the outstretched arm with forces transmitted through the arm to the AC joint. Patients with osteolysis of the distal clavicle sometimes give a history of acute trauma, although the more common cause is repetitive microtrauma to the AC joint caused by activities such as weight lifting or gymnastics. Patients frequently complain of pain over the AC joint when adducting the ipsilateral shoulder, such as during a golf swing or when buckling a seat belt. Often, pain occurs when sleeping on the affected shoulder. Athletes may experience AC joint pain on bench pressing, push-ups, and dips. Pain and weakness of the affected shoulder also may be experienced with forward flexion and adduction of the arm.

On physical examination, a visible step-off may be observed between the medial acromion and the distal clavicle, indicating a probable AC separation. Pain usually can be elicited on direct palpation of the AC joint and is made worse by a cross-arm adduction maneuver. This test is performed by internally rotating the arm, which is maximally adducted across the chest, and is considered positive if pain is produced in the AC joint (see Figure 46-3D). Pain also may be elicited by moving the arm from a horizontally abducted position to the extended position and on maximal internal rotation of the shoulder. These tests cause rotation and compression of the AC joint and are sensitive but less specific. They also may be positive with other disorders of the shoulder, such as posterior capsular stiffness.

Frequently, AC joint pain coexists with subacromial impingement and rotator cuff pathology. In these cases, impingement signs are positive, and rotator cuff weakness may be present. Otherwise, no muscle weakness should be detectable on manual resistance testing, and no evidence of muscle atrophy should be found. The AC joint and the subacromial space may have to be injected on separate occasions to determine the true source of the symptoms. Some physicians have noticed an association of AC joint symptoms with shoulder instability. Glenohumeral motion can vary, depending on chronicity and isolation of the problem to the AC joint. In isolated cases, some loss of internal rotation of the affected shoulder may be caused by pain.

Radiographs should include anteroposterior views of the shoulder in the scapular plane in neutral, internal, and external rotation; a transscapular Y view; an axillary view; and a 15-degree cephalic tilt view of the AC joint at 50% penetrance, as described by Zanca (see Figure 46-4). Stress views may be obtained by strapping 5 to 10 lb of weight to the forearms and determining AC separation. Comparing the coracoclavicular distance of both shoulders also may be helpful. When clinically indicated, cervical spine radiographs should be obtained to exclude cervical spondylosis.

Radiographic evaluation may reveal AC joint arthrosis with microcystic changes in the subchondral bone, sclerosis, osteophytic lipping, and joint space narrowing. In cases of osteolysis, radiographs may reveal loss of subchondral bone detail with microcystic appearances in the subchondral region of the distal clavicle and osteopenia of the lateral one-third of the clavicle. In late stages of osteolysis, resorption of the distal end of the clavicle results in marked widening of the AC joint and sometimes complete resorption of the distal clavicle. AC separation may be evident with widening of the coracoclavicular distance and post-traumatic ossification of the coracoclavicular ligaments.

AC symptoms do not always correlate with the radiographic appearance of the joint. DePalma found AC joint degeneration to be an age-related process, with symptoms not always correlating with radiographic findings of AC joint arthrosis. AC joint pain may occur despite normal radiographs.

A technetium 99m phosphate bone scan may assist in the diagnosis, revealing increased uptake in the distal clavicle and the medial acromion. In cases of atraumatic osteolysis of the distal clavicle, increased uptake may be isolated to the distal clavicle, but in approximately 50% of cases, scintigraphic activity of the adjacent medial acromion is increased. The bone scan may reveal pathologic changes in the AC joint when plain radiographs appear normal.

In selected cases, MRI can be valuable in determining a diagnosis and evaluating the glenohumeral and subacromial regions for coexisting pathology (Figure 46-15). AC joint involvement may reveal increased fluid with synovitis, soft
tissue enlargement, and periarticular ossifications with encroachment on underlying bursal and cuff tissue.

Patients with AC joint pain usually respond well to nonoperative treatment; however, complete relief of symptoms may require an extended period. Conservative therapy includes heat, NSAIDs, steroid injections, shoulder rehabilitation, and avoidance of painful positions and activities. Steroid injections are repeated at 3-month intervals if painful conditions persist.

Open resection of the distal clavicle for chronic AC joint pain was initially reported by Gurd\textsuperscript{140} and by Mumford,\textsuperscript{141} both with good results. Since that time, other surgeons have reported similar good results with open resection, but significant morbidity, such as disruption of the deltotrapezial fascia and anterior deltoid rupture, can occur.\textsuperscript{123,124,137,139,142} Arthroscopic resection of the distal clavicle has been described with results similar to open resection.\textsuperscript{13,129,134,135,143-146}

**Glenohumeral Disorders**

The various arthritides that affect the shoulder joint are discussed in detail in other chapters. They are presented here to address aspects that are unique to the glenohumeral joint. The usual presentation of intra-articular disorders consists of pain with motion and symptoms of internal derangement, such as locking and clicking. Pain is generalized throughout the shoulder girdle and sometimes is referred to the neck, back, and upper arm. The usual response to pain includes decreased glenohumeral motion and substitution with increased scapulothoracic mobility. Patients with adequate elbow and scapulothoracic motion require little glenohumeral motion for activities of daily living; patients with glenohumeral arthrodesis can achieve adequate function.\textsuperscript{147,148} The response to pain consists of diminution of motion and secondary soft tissue contractures with muscle atrophy. With increasing weakness and involvement of adjacent joints, pain, limitation of motion, and weakness can cause a substantial functional deficit.

**Inflammatory Arthritis**

Although the most common inflammatory arthritis involving the shoulder joint is rheumatoid arthritis (RA), other systemic disorders, such as systemic lupus erythematosus, psoriatic arthritis, ankylosing spondylitis, reactive arthritis, and scleroderma, may cause glenohumeral degeneration. Motion is limited by splinting of the joint with secondary soft tissue contractures, or by primary soft tissue involvement with scarring or rupture. Plain radiographs confirm glenohumeral involvement (Figure 46-16A). Narrowing of the glenohumeral joint space may occur, with erosion and cyst formation and without significant sclerosis or osteophytes. As the disease progresses, superior and posterior erosion of the glenoid with proximal subluxation of the humeral head may occur. Eventually, secondary degenerative changes and even osteonecrosis of the humeral head may occur.
Treatment is initially conservative and is directed toward controlling pain, inducing a systemic remission, and maintaining joint motion through physical therapy. The use of intra-articular glucocorticoids may be beneficial in controlling local synovitis. In rheumatoid arthritis, the involvement of periarticular structures with subacromial bursitis and rupture of the rotator cuff magnifies the functional deficit. When synovial cartilage interactions produce significant symptoms and radiographic changes that cannot be controlled by conventional therapy, glenohumeral resurfacing should be considered.

When following an RA patient with shoulder involvement, the rheumatologist should assess range of motion carefully and should obtain periodic radiographs. Patients with progressive loss of motion or radiographic destruction should be referred for evaluation for possible surgical treatment. The treatment of choice is an unconstrained total shoulder arthroplasty. Total shoulder arthroplasty is best performed in patients with rheumatoid arthritis before end-stage bony erosion and soft tissue contractions have occurred. Acute inflammatory arthritis of the glenohumeral joint may be associated with gout, pseudogout, hydroxyapatite deposition of renal osteodystrophy, and recurrent hemophilic hemarthrosis.

**Osteoarthritis**

Osteoarthritis of the glenohumeral joint is less common than that in the hip, its counterpart in the lower extremity; this condition is caused by non–weight-bearing characteristics of the shoulder joint and the distribution of forces throughout the shoulder girdle. Osteoarthritis is divided into conditions associated with high unit loading of articular cartilage and conditions in which an intrinsic abnormality within the cartilage causes abnormal wear at normal loads. Because the shoulder is normally a non–weight-bearing joint and is not usually susceptible to repeated high loading, the presence of osteoarthritis of the glenohumeral joint should alert the physician to consider other factors. Has the patient engaged in unusual activities, such as boxing, heavy construction, or long-term use of a pneumatic hammer? Has some disorder, such as epiphyseal dysplasia, created joint incongruity with high unit loading of the articular cartilage? Is this a neuropathic process caused by diabetes, syringomyelia, or leprosy? Have associated hemochromatosis, hemophilia, or gout altered the ability of articular cartilage to withstand normal loading? Is unrecognized chronic dislocation responsible?

Pain is the usual presentation, but generally it is not as acute or it may be associated with the spasm seen in inflammatory conditions. Plain radiographs show narrowing of the glenohumeral joint, osteophyte formation, sclerosis, and some cyst formation (Figure 46-16B). Because the rotator cuff usually is intact, less bone erosion of the glenoid and proximal subluxation of the humerus is noted. Patients with osteoarthritis of the glenohumeral joint frequently do well with functional adjustments and conservative therapy. Analgesics and NSAIDs may provide symptomatic relief. The use of glucocorticoid injections is less beneficial, unless evidence of synovitis is observed. Patients with severe involvement who fail to respond are best treated by shoulder arthroplasty.

**Osteonecrosis**

Osteonecrosis of the shoulder refers to necrosis of the humeral head seen in association with a variety of conditions. Symptoms are due to synovitis and joint incongruity resulting from resorption, repair, and remodeling. Pathogenesis and various causes are discussed in Chapter 103.

The most common cause of osteonecrosis of the shoulder is avascularity resulting from a fracture through the anatomic neck of the humerus. Fracture through this area disrupts intramedullary and capsular blood supplies to the humeral head. Another common cause of osteonecrosis of the shoulder is steroid therapy provided in conjunction with organ transplantation, systemic lupus erythematosus, or asthma. Other conditions associated with osteonecrosis of the humeral head include hemoglobinopathies, pancreatitis, and hyperbarism.

Early diagnosis is difficult because the presence of symptoms is often delayed. Bone scans may be helpful in early cases, before radiographic changes are evident. MRI is highly sensitive and is more specific than scintigraphy. Plain radiographs show progressive phases of necrosis and repair (as discussed in Chapter 103). In early stages, the films may be normal or may show osteopenia or bone sclerosis. A crescent sign representing subchondral fracture or demarcation of the necrotic segment appears during the reparative process. Patients who fail to remodel show collapse of the humeral head with secondary degenerative changes. A considerable discrepancy is often noted between symptoms and radiographic involvement. Patients with extensive bone changes may be asymptomatic. Treatment should be directed by the patient’s symptoms rather than by the radiographs and is similar to that provided for osteoarthritis. Arthroscopy occasionally is helpful by removing loose chondral fragments and debriding chondral incongruities. Patients with severe symptoms that cannot be controlled by conservative means are best treated with unconstrained shoulder arthroplasty, hemiarthroplasty, or resurfacing arthroplasty.

**Cuff-Tear Arthropathy**

In 1873, Adams described the pathologic changes that characterize rheumatoid arthritis of the shoulder and a condition that has since that time been referred to as Milwaukee shoulder or cuff-tear arthropathy. McCarty called the condition Milwaukee shoulder and reported that factors predisposing to this syndrome included deposition of calcium pyrophosphate dihydrate crystals, direct trauma, chronic joint overuse, chronic renal failure, and denervation. Patients with Milwaukee shoulder have elevated levels of synovial fluid 5-nucleotidase activity and elevated levels of synovial fluid inorganic pyrophosphate and nucleotide pyrophosphohydrolase activity.

Neer and colleagues reported a similar condition in which untreated massive tears of the rotator cuff with proximal migration of the humeral head are associated with erosion of the humeral head. Erosion of the humeral head differs from that seen in other arthritides and is presumed to be caused by a combination of mechanical and nutritional factors acting on the superior glenohumeral cartilage.
Patients with cuff-tear arthropathy present a difficult therapeutic problem because bone erosion and disruption of the cuff jeopardize the functional result from an unconstrained prosthesis.\textsuperscript{151} Hemiarthroplasty or a reverse total shoulder arthroplasty may be indicated.\textsuperscript{160,161} The major challenge in treating cuff-tear arthropathy is to determine which patients with massive rotator cuff tears will proceed to the syndrome of cuff-tear arthropathy. Patients with massive rotator cuff tears who develop localized calcium pyrophosphate disease may be predisposed to further proximal migration and further joint destruction. This situation poses a dilemma for the treating physician. Many patients with massive rotator cuff tears remain stable and require little or no treatment. Occasionally, symptomatic patients can be treated by arthroscopic débridement of the cuff tear. In a recent study, patients with massive rotator cuff tears without arthritis did well when treated with reverse total shoulder arthroplasty.\textsuperscript{115} It is crucial to define the patient who will proceed to the syndrome of cuff-tear arthropathy. If crystal deposition disease predisposes patients to proximal migration and joint destruction, joint aspiration with crystal analysis and scintigraphy to determine synovial reaction may be helpful diagnostic tools.

Hamada and co-workers\textsuperscript{162} followed 22 patients with massive rotator cuff tears treated conservatively. Radiographic findings included narrowing of the acromiohumeral interval and degenerative changes in the humeral head, tuberosities, acromion, AC joint, and glenohumeral joint. Five of seven patients followed for longer than 8 years progressed to cuff-tear arthropathy. Investigators concluded that progressive radiographic changes were associated with repetitive use of the arm in elevation, rupture of the long head of the biceps, impingement of the humeral head against the acromion, and weakness of external rotation.\textsuperscript{162}

**Septic Arthritis**

Septic arthritis can masquerade as any of the conditions classified as periarthritic or glenohumeral disorders (see Chapters 99 and 110). Sepsis must be included in any differential diagnosis of shoulder pain because early recognition and prompt treatment are necessary to achieve a good functional result. The diagnosis is confirmed by joint aspiration with synovial fluid analysis and culture. Cultures should include aerobic, anaerobic, mycobacterial, and fungal studies.

**Labral Tears**

The glenoid labrum increases the depth of the glenoid and serves as an anchor for the attachment of the glenohumeral ligaments. Historically, labral tears have been difficult to diagnose. Findings on physical examination can be confused with impingement and rotator cuff tendinopathy and bicipital tendinitis. Diagnosis can be confirmed with MRI-arthrography, CT-arthrography, and DCAT.\textsuperscript{117} Arthroscopy has greatly increased our knowledge of the glenoid labrum in normal and pathologic situations and has aided clinicians in the diagnosis and treatment of labral lesions.

Labral tears can be divided into tears associated with symptoms of internal derangement and tears associated with anterior or posterior instability. A soft tissue Bankart lesion is associated with a tear of the anterior band of the inferior glenohumeral ligament and with anterior instability. Isolated labral tears that do not involve detachment of the ligaments can cause internal derangement and may have an arthroscopic appearance similar to that of a meniscal tear of the knee.

Andrews and associates\textsuperscript{7} first described lesions of the anterior superior labrum in throwing athletes; these lesions were often associated with biceps tendon tears (10%), which may result from traction of the biceps tendon. Snyder and co-workers\textsuperscript{163} introduced the term SLAP lesion in 1990 to describe an injury involving the long head of the biceps tendon and the superior portion of the glenoid labrum.

The long head of the biceps tendon originates at the supraglenoid tubercle and the glenoid labrum in the superior-most portion of the glenoid. The major portion of the tendon blends with the posterior superior aspect of the labrum. The most common mechanism of a SLAP injury is a fall onto an outstretched arm with the shoulder in abduction and slight forward flexion.\textsuperscript{163} The lesion also can result from acute traction on the arm and from an abduction and external rotation mechanism.\textsuperscript{164,165}

Patients usually complain of pain with overhead activities and a frequent catching or popping sensation in the shoulder. The most reliable diagnostic test is O’Brien’s test. The test is performed against resistance with the arm in forward flexion and with the elbow extended and the forearm pronated. In the second part of the test, the arm is supinated. Less pain during the latter part of the test suggests a SLAP lesion.\textsuperscript{163} The most accurate diagnostic test is MRI-arthrography with gadolinium.\textsuperscript{166} Treatment for symptomatic SLAP lesions is surgical.

**Adhesive Capsulitis**

Adhesive capsulitis, or frozen shoulder syndrome (FSS), is a condition characterized by limited motion of the shoulder joint with pain at the extremes of motion. It was first described by Putman\textsuperscript{167} in 1882 and later by Codman.\textsuperscript{1} The initial presentation is pain, which is generalized and is referred to the upper arm, back, and neck. As pain increases, loss of joint motion ensues. The process generally is self-limiting and in most cases resolves spontaneously within 10 months, unless an underlying problem is present.

The exact cause of FSS is unknown.\textsuperscript{9,168} It is frequently associated with conditions such as diabetes mellitus, parkinsonism, thyroid disorders, and cardiovascular disease. When one of these conditions exists, a history of some mild trauma that initiated the frozen shoulder is often reported. Major skeletal trauma and soft tissue injury may coexist with FSS. It also may be seen with a variety of other conditions, including apical lung tumor, pulmonary tuberculosis, cervical radiculopathy, and post myocardial infarction.\textsuperscript{169,171} In one review of FSS, 3 of 140 patients with this syndrome had local primary invasive neoplasms.\textsuperscript{117} Another study described 3 patients with adhesive capsulitis who subsequently were found to have a neoplastic lesion of the midshaft of the humerus.\textsuperscript{173} In a high-risk patient with an underlying disorder, even minor surgery or trauma in a remote location, such as the hand, can precipitate FSS.\textsuperscript{9,174,175}
Glenohumeral Instability

Glenohumeral instability is a pathologic condition that manifests as pain associated with excessive translation of the humeral head on the glenoid during shoulder motion. Instability can range from excessive laxity with episodes of subluxation to frank dislocation of the joint. Traumatic dislocation of the glenohumeral joint reveals characteristic clinical and radiographic findings that are beyond the scope of this chapter and have been reviewed in detail elsewhere. The most common type of instability is anterior, although posterior and multidirectional laxity of the shoulder is increasingly recognized as a cause of shoulder pain. Anterior dislocation usually occurs with the arm in an abducted and externally rotated position, and the diagnosis is usually obvious. Posterior dislocation is frequently associated with convulsive disorders or unusual trauma with the arm in a forward flexed and internally rotated position. The diagnosis is often missed and should always be suspected in the patient who is unable to rotate the arm externally after trauma.

Recurrent subluxation without dislocation may be difficult to diagnose and may be mistakenly identified as impingement with chronic cuff tendinitis. An overhead athlete may experience repetitive stresses to the shoulder, causing microtrauma to the static stabilizers. Jobe and colleagues described a syndrome of shoulder pain in overhead or throwing athletes that manifests as impingement but is caused by anterior subluxation of the joint, with the humeral head impinging on the anterior aspect of the coracohumeral arch. Fu and co-workers underscored this distinction by dividing the causes of rotator cuff tendinitis into primary impingement of the tendon on the coracohumeral arch and anterior subluxation with secondary impingement in young athletes performing overhead movements. Walch and colleagues described intra-articular impingement between the undersurface of the rotator cuff (supraspinatus and infraspinatus) and the posterior superior glenoid rim and labrum. This “internal impingement” usually is observed in overhead athletes with subtle anterior glenohumeral instability and results in tendinitis or partial tears of the rotator cuff (see Figure 46-16).

The pathophysiology involves a diffuse inflammatory synovitis with subsequent adherence of the capsule and loss of the normal axillary pouch and joint volume, which leads to significant loss of motion. Capsular contracture is thought to result from adhesion of the capsular surfaces or fibroblastic proliferation in response to cytokine production. The condition is common in women in their 40s and 50s. Typically, the patient relays a history of diffuse, dull aching around the shoulder, with weakness and loss of motion occurring over a few months.

Usually, three distinct clinical stages of the syndrome can be identified. Stage one is the painful inflammatory or freezing phase. During this stage, pain is severe, is exacerbated by any attempts at movement, and usually lasts a few weeks or months. The patient usually feels most comfortable with the arm at the side in an adducted and internally rotated position. Phase two, the adhesive or stiffening phase, generally lasts 4 to 12 months. Pain is usually minimal during this phase, although periscapular symptoms may develop from compensatory motion to achieve elevation of the arm. The third phase of the syndrome is the resolution or thawing phase, which may last 5 to 26 months. During this time, pain eases and motion slowly improves, although some patients may improve dramatically over a short period.

In the early stages, any attempts at motion may produce severe pain and associated weakness. The syndrome usually is associated with a prolonged period of immobilization. Night pain is common, along with an inability to sleep on the associated shoulder; this is similar to findings of impingement syndrome.

In patients with a history of minimal or no trauma and FSS, a metabolic cause should be excluded. Complete blood cell count, erythrocyte sedimentation rate, serum chemistry, and thyroid function tests are done as a screening panel. Further testing is done if results suggest that the patient may have a systemic illness. Plain radiographs should include true anteroposterior, axillary, and scapular Y views of the shoulder. In patients with no underlying detectable illness and a negative workup, a Tc 99m pertechnetate scan may show increased uptake in FSS, but more important, it is used to exclude occult lesions or metastasis.

Literature review reveals a multitude of treatment options along with significant deficits on accurate reporting of disease staging with response to treatment.

Treatment of FSS is mainly conservative and consists of intra-articular injections, heat, gentle stretching, NSAIDs, and modalities such as transcutaneous electrical nerve stimulation. The disease usually is self-limited and after the painful phase is not severely disabling. Communication between the physician and the patient, together with a thorough explanation of the condition, is essential because resolution of the syndrome occurs slowly over time. Closed manipulation and surgery (open and arthroscopic) are reserved for patients whose condition is recalcitrant to conservative measures, or for whom the diagnosis is in question. Paramount in the prevention of FSS is avoiding overimmobilization in a minor shoulder injury, in addition to careful identification of patients at risk for FSS.

Fareed and Gallivan reported good results with hydraulic distention of the glenohumeral joint using local anesthetic agents. Rizk and associates conducted a prospective, randomized study to assess the effects of steroid or local anesthetic injection in 48 patients with FSS. No significant difference in outcome was noted between individuals who received intrabursal or intra-articular injection. Steroid with lidocaine offered no advantage over lidocaine alone in restoring shoulder motion. However, transient pain relief occurred in two-thirds of steroid-treated patients.

General anesthesia occasionally is indicated for closed manipulation. Hill and Bogumill reported the results of manipulation of 17 frozen shoulders in 15 patients who did not respond to physical therapy. On average, 78% of individuals who were working before their shoulder problems returned to work 2.6 months after manipulation. Investigators concluded that manipulation allowed patients to return to a normal lifestyle and to work sooner than the reported natural history of the condition. Surgical intervention for adhesive capsulitis should be limited to treatment of an underlying problem, such as calcific tendinitis or an impingement syndrome.
The diagnosis of glenohumeral instability with subluxation in one or multiple directions is made with the combination of a detailed history and physical examination and the use of adjuncts, such as arthrography, CT, MRI, and arthroscopy with examination under anesthesia. The syndrome of multidirectional instability has been recognized in patients with symptomatic inferior instability, in addition to anterior or posterior instability. Approximately 50% of affected patients have evidence of generalized laxity. Frequently, the syndrome occurs in young athletic patients who are loose jointed, in particular in the dominant arm of pitchers, racket sports players, and swimmers. In this type of athlete, repetitive microtrauma may cause stretching of the shoulder, resulting in a large capsular pouch without labral detachment. A traumatic event may damage the shoulder, resulting in the syndrome of multidirectional instability and a Bankart lesion.

The most common manifestation in these patients is pain, which is often mistakenly considered to be rotator cuff tendinitis. The patient may relate a history of minor trauma causing acute pain and a “dead arm” syndrome lasting minutes or hours. Other associated symptoms include a sense of instability, weakness, and radial symptoms suggestive of neuropathy. Few or no positive physical findings may be associated with chronic subluxation or multidirectional instability. The patient may have signs of generalized ligamentous laxity, and pain may be reproduced by subluxating the glenohumeral joint in multiple directions. One particularly helpful sign of inferior laxity is the sulcus sign, which refers to the subacromial indentation that occurs when longitudinal traction is applied to the humerus with the arm at the side. This sign occurs with inferior translation of the humeral head. Because this syndrome frequently occurs in athletes with highly developed musculature around the shoulder girdle, physical findings of subluxation may be difficult to reproduce in the office setting.

Plain radiographs are generally normal, although some inferior subluxation may be shown on stress radiographs obtained with the use of weights. Special radiographs, as discussed previously, may show a Bankart lesion (i.e., avulsion of the anterior inferior glenoid rim) or a Hill-Sachs lesion (i.e., osteochondral defect of the posterior humeral head) with subluxation of the humeral head in front of the anterior glenoid rim. CT-arthrography or MRI-arthrography may show increased capsular volume, a labral detachment, or a Hill-Sachs lesion (see Figure 46-9). When surgery is indicated, examination under anesthesia and shoulder arthroscopy may assist in diagnosing the primary direction of instability in the syndrome of multidirectional instability. In selected patients with traumatic anterior dislocation who have no history of multidirectional instability, arthroscopic stabilization may be done with stabilization of the capsulolabral complex.

Treatment of patients with chronic subluxation or the syndrome of multidirectional instability is first directed toward prolonged rehabilitation. Activities that stress the shoulder and produce symptoms are avoided. Strengthening exercises of the shoulder girdle may control symptoms, dynamically stabilizing the glenohumeral joint, and may obviate the need for surgical intervention. If a conservative treatment program fails, surgery is performed on the side associated with the greatest clinical instability. Stabilization is directed toward tightening of the capsular structures to stabilize the glenohumeral joint.

**EXTRINSIC OR REGIONAL FACTORS CAUSING SHOULDER PAIN**

Because the shoulder girdle connects the thorax with the upper extremity, and because major neurovascular structures pass in proximity to the joint, shoulder pain is a hallmark of many nonarticular conditions.

**Cervical Radiculopathy**

Cervical pathology may manifest with associated shoulder pain. The area of referred pain has a dermatomal pattern, consistent with the distribution of dermatomal nerve roots. Isolation of the pain usually defines the exact location of associated cervical pathology. Pain can be differentiated from shoulder pain on the basis of history, physical examination, EMG, cervical radiographs, and myelography or MRI when indicated. Because conditions causing cervical neck pain and conditions causing shoulder pain, such as calcific tendinitis and cervical radiculopathy, may coexist, it is often difficult to distinguish which lesion is responsible for the symptoms. These conditions often can be differentiated by injection of local anesthetics to block certain components of the pain.

The thoracic outlet is an interval created by the anterior and middle scalene muscles and the first rib through which the brachial plexus and vessels pass to the arm. In thoracic outlet syndrome, compression of these nerves and vessels often manifests as vague shoulder pain with numbness of the ipsilateral fourth and fifth digits. Cervical rib or hyperplasia of the scalene muscles can be related to the onset of pain. The occurrence of pain also has been related to scapular ptosis, poor posture, and clavicular fracture with malunion or copious formation of callus.

**Brachial Neuritis**

In the 1940s, Spillane and Parsonage and Turner described a painful condition of the shoulder associated with limitation of motion. As pain subsided and motion improved, muscle weakness and atrophy became apparent. The deltoid, supraspinatus, infraspinatus, biceps, and triceps are the most frequently involved muscles, although diaphragmatic paralysis also has been reported. The cause is unclear, but the clustering of cases suggests a viral or postviral syndrome. Occasionally, an associated influenza-like syndrome or previous vaccination has been reported.

Hershman and colleagues described acute brachial neuropathy in athletes. Findings that suggest an acute brachial neuropathy include acute onset of pain without trauma; persistent, severe pain that continues despite rest; and patchy neurologic signs. The diagnosis is confirmed by EMG and nerve conduction studies. The prognosis for recovery is excellent, although full recovery may take 2 to 3 years. Tsairis and associates reported 80% recovery within 2 years and more than 90% recovery by the end of 3 years.
Nerve Entrapment Syndromes

Peripheral compression neuropathies of the upper extremities may produce referral pain to the shoulder. Distant compression neuropathies, such as carpal tunnel (median nerve) and cubital tunnel (ulnar nerve) syndromes, may manifest with concomitant and separate shoulder impingement with rotator cuff disease. Associated numbness and paresthesias with mapping of the dermatomal distribution and with peripheral neuropathy often direct the examiner to the appropriate diagnosis. Patients often give a history of dropping objects and a feeling of clumsiness with the affected hand. A Tinel sign may be elicited over the region of entrapment at the elbow or wrist. Provocative maneuvers such as Phalen's test may be positive and usually indicate median nerve compression at the wrist. Diminished vibratory sensation is an early finding in the disease and is easily reproducible, whereas decreased two-point discrimination and intrinsic atrophy are late findings of peripheral compression neuropathy.

The diagnosis usually can be made by clinical examination with exclusion of other possible causes. EMG and nerve conduction velocity tests may reveal slowed conduction and latency at appropriate compression points to aid in diagnosis. Spinal accessory nerve injury with subsequent denervation of the trapezius may cause weakness and pain in the shoulder consistent with impingement. The injury can occur from traction injury to the neck or a direct blow or pressure to the base of the neck. Iatrogenic nerve injury may result from surgical procedures on the neck such as lymph node biopsy. The injury produces weakness in shoulder abduction with associated pain that radiates from the neck into the trapezius and shoulder. Subsequent atrophy of the trapezius may lead to dysymmetry and ptosis of the involved shoulder, with narrowing of the suprascapular notch. The injury can be well localized and often can be elicited by palpation over the region of the suprascapular notch. Any activity that brings the scapula forward, such as reaching across the chest, may aggravate the pain. The location of pain and other symptoms can mimic more common entities, such as impingement, rotator cuff disease, cervical disk disease, brachial neuropathy, biceps tendinitis, thoracic outlet syndrome, AC disease, and instability of the shoulder.

The resulting suprascapular neuropathy produces pain in the posterolateral aspect of the shoulder that may radiate into the ipsilateral extremity, shoulder, or side of the neck. Although this condition is uncommon, it can have a prolonged and disabling course when undiagnosed. Because the suprascapular nerve has no cutaneous innervation, no numbness, tingling, or paresthesias are associated. Weakness is usually noted in abduction and external rotation, and significant atrophy is often observed at diagnosis. The pain frequently is described as a deep burning or aching that can be well localized and often can be elicited by palpation over the region of the suprascapular notch. Any activity that brings the scapula forward, such as reaching across the chest, may aggravate the pain. The location of pain and other symptoms can mimic more common entities, such as impingement, rotator cuff disease, cervical disk disease, brachial neuropathy, biceps tendinitis, thoracic outlet syndrome, AC disease, and instability of the shoulder.

Fritz and colleagues reported the efficacy of MRI in the diagnosis of suprascapular nerve entrapment secondary to space-occupying lesions. Definitive diagnosis is made with EMG and nerve conduction studies. EMG changes usually reveal spontaneous activity in the muscle at rest and fibrillations indicating motor atrophy and denervation. Nerve conduction studies may reveal slowing across the site of entrapment. As with axillary nerve entrapment, the syndrome is often associated with young, athletic individuals with excessive overhead activity. It also has been associated with trauma.

Lack of consensus continues regarding the optimal treatment of suprascapular neuropathy. Post and Grinblat reported good to excellent results with surgical treatment in 25 of 26 cases. No difference in residual atrophy and in strength deficits has been shown, however, for operative and nonoperative treatments. Ferretti and co-workers evaluated 96 top-level volleyball players from the 1985 European Championships and found that 12 had isolated suprascapular neuropathy with atrophy of the infraspinalus of the dominant shoulder. All players were unaware of any impairment, however, and played without limitations. After a space-occupying lesion has been excluded, a 6-month trial of conservative treatment may be indicated for some individuals. If the entrapment does not improve, or if symptoms worsen with conservative treatment, surgical...
decompression for pain relief is warranted; however, resolution of atrophy and strength gains can vary.213

**Sternoclavicular Arthritis**

Occasionally, traumatic, nontraumatic, or infectious conditions can cause pain around the sternoclavicular joint (see Figure 46–1). The most common problem involves ligamentous injury and painful subluxation or dislocation. This can be diagnosed by palpable instability and crepitus over the sternoclavicular joint. Sternoclavicular views may radiographically show dislocation.221

Inflammatory arthritis of the sternoclavicular joint has been associated with rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, and septic arthritis. The association of plantar-palmar pustulosis with sternoclavicular arthritis has been reported.222 Seven of 15 patients who underwent biopsy for this condition had cultures positive for *Propionibacterium acnes*, suggesting an infectious origin of the condition.222

Two other conditions involving the sternoclavicular joint are Tietze’s syndrome, a painful, nonsuppurative swelling of the joint and adjacent sternochondral junctions, and Friedrich’s syndrome, a painful osteonecrosis of the sternal end of the clavicle. Condensing osteitis of the clavicle is a rare benign idiopathic lesion of the medial one-third of the clavicle. This condition, which is better described as aseptic enlarging osteosclerosis of the clavicle, is most commonly seen in middle-aged women and manifests as a tender swelling over the medial one-third of the clavicle.223

**Reflex Sympathetic Dystrophy**

Since its original description by Mitchell224 in 1864, reflex sympathetic dystrophy (RSD) has remained a poorly understood and frequently overlooked condition. Its cause is unknown but may be related to sympathetic overflow or short-circuiting of impulses through the sympathetic system. Any clinician who deals with painful disorders must be familiar with the diagnosis and treatment of this condition. Bonica’s225 excellent review covers the clinical presentation, various stages of the disease, and the importance of early intervention to ensure a successful outcome.

RSD has been called *causalgia, shoulder-hand syndrome,* and *Studeck’s atrophy,* which has caused some confusion. It is generally associated with minor trauma and is to be differentiated from causalgia, which involves trauma to major nerve roots.224 RSD is divided into three phases, which are important in determining appropriate treatment.225 Phase one is characterized by sympathetic overflow with diffuse swelling, pain, increased vascularity, and radiographic evidence of demineralization. If left untreated for 3 to 6 months, the condition may progress to phase two, which is characterized by atrophy. The extremity may now be cold and shiny, with atrophy of the skin and muscles. Phase three refers to progression of trophic changes, with irreversible flexion contracture and a pale, cold, painful extremity. It has been speculated that phase one is related to peripheral short-circuiting of nerve impulses, phase two represents short-circuiting through the intermuncial pool in the spinal cord, and phase three is controlled by higher thalamic centers.225,226

Steinbrocker227 reported that recovery is possible as long as vasomotor activity with swelling and hyperemia is evident. After the trophic phase two or three is established, the prognosis for recovery is poor. Prompt recognition of the syndrome is important because early intervention to control pain is mandatory. Careful supervision and reassurance are crucial because many of these patients are emotionally labile as a result of the pain or an underlying problem. The syndrome may be remarkably reversed by a sympathetic block. Patients who receive transient relief from sympathetic blockade may be helped by surgical sympathectomy.

**Neoplasms**

Primary and metastatic neoplasms may cause shoulder pain by direct invasion of the musculoskeletal system or by compression with referred pain.228 Primary tumors are more likely to occur in younger individuals. More common lesions have a typical distribution, such as the predilection of a chondroblastoma for the proximal humeral epiphysis or an osteogenic sarcoma for the metaphysis.229 The differential diagnosis of spontaneous onset of shoulder pain in older individuals should include metastatic lesions and myeloma. Neoplasms are best identified by plain radiographs, MRI, Tc 99m MDP scintigraphy, and CT.

Neoplasms also may involve the shoulder region through metastases to the region. An associated history of carcinomas should alert the examiner to the possibility of a bone tumor, especially in patients who have had malignancies with a predilection for metastasis to bone (e.g., thyroid, renal, lung, prostate, breast). Pain often is present at rest and is exacerbated at night. Atypical pain distribution that is not relieved by injection without specific dermatomal distribution should alert the examiner to other underlying possibilities. Plain radiographs should be evaluated thoroughly for any cortical destruction and for lytic lesions.

Pancoast syndrome or apical lung tumor may manifest as shoulder pain or cervical radiculitis caused by invasion of the brachial plexus or invasion of C8 or T1 roots.230,232 With invasion of the cervical sympathetic chain, the patient also may develop Homer’s syndrome.

**Miscellaneous Conditions**

With increasing numbers of patients undergoing long-term maintenance hemodialysis, a shoulder pain syndrome known as *dialysis shoulder arthropathy* has been described. It consists of shoulder pain, weakness, loss of motion, and functional limitation. The cause and pathogenesis of this syndrome are unclear, although rotator cuff disease, pathologic fracture, bursitis, and local amyloid deposition have been implicated as causative factors.231 Surgical or necropsy data are insufficient to confirm a specific diagnosis. Patients generally respond poorly to local measures of injection, heat, and NSAIDs, but their condition may improve with correction of underlying metabolic disorders, such as osteomalacia and secondary hyperparathyroidism.

In patients older than 50 years of age with bilateral shoulder pain and stiffness, polymyalgia rheumatica should be considered as a diagnosis. It is twice as common among females and is almost exclusively found in whites.234 Most
patients have sedimentation rates greater than 40 mm/hr and respond to low doses of corticosteroids.  

Selected References

1. Codman EA: The shoulder: rupture of the supraspinatus tendon and other lesions in or about the subacromial bursa, Boston, 1934, Thomas Todd.
References

1. Codman EA: The shoulder: rupture of the supraspinatus tendon and other lesions in or about the subacromial bursa, Boston, 1934, Thomas Todd.


64.e2 PART 6  DIFFERENTIAL DIAGNOSIS OF REGIONAL AND DIFFUSE MUSCULOSKELETAL PAIN


Low Back Pain

RAJIV DIXIT

KEY POINTS

Up to 80% of the population will experience low back pain (LBP), and degenerative changes of the lumbar spine is the most common cause.

More than 90% of these patients are largely pain free within 8 weeks.

Initial evaluation should identify the few patients with neurologic involvement or suspicion of systemic disease (infection, malignancy, or spondyloarthritis) because they may need urgent or specific intervention.

Psychosocial and other factors that predict risk of chronic disabling LBP should be assessed.

Imaging is rarely indicated in the absence of significant neurologic involvement or suspicion of systemic disease.

Imaging abnormalities should be carefully interpreted because they are frequently present in asymptomatic individuals.

A precise pathoanatomic diagnosis with identification of the pain generator cannot be made in up to 85% of patients.

Persistent LBP should be treated with an individually tailored program that includes analgesia, core strengthening, stretching, aerobic conditioning, loss of excess weight, and patient education.

Intensive interdisciplinary rehabilitation with an emphasis on cognitive-behavioral therapy should be strongly considered if conservative measures fail.

There is no evidence for the effectiveness of epidural corticosteroid injections in patients without radiculopathy.

A large number of injection techniques, physical therapy modalities, and nonsurgical interventional therapies lack evidence of efficacy.

The major indication for back surgery is presence of a serious or progressive neurologic deficit.

Back surgery in the absence of neurologic deficits, especially spinal fusion for degenerative changes, is not clearly effective.

EPIDEMIOLOGY

Low back pain (LBP) is one of the most common conditions encountered in clinical medicine. It affects the area between the lower rib cage and gluteal folds.

An estimated 65% to 80% of the population will experience LBP during their lifetime. LBP is the most prevalent chronic pain syndrome and the leading cause of limitation of activity in patients younger than the age of 45. It is also the second most frequent reason for a visit to the physician’s office and the third most common surgical indication. The incidence of LBP increases with age, and LBP more commonly affects women.

The natural history of back pain, especially the duration and chronicity, remains somewhat controversial. Regardless, studies show that pain and function improve substantially in most patients within 1 month, and more than 90% are better at 8 weeks. These patients, however, remain susceptible to future relapses that also tend to be brief. The remaining 7% to 10% develop chronic LBP and it is these individuals who are largely responsible for the high costs associated with LBP. The most recent reliable data, from 1998, estimated the direct cost of managing LBP in the United States to be $90 billion with substantial additional indirect costs due to lost time from work and decreased productivity.

A number of risk factors have been associated with LBP including heredity, psychosocial factors, heavy lifting, obesity, pregnancy, weaker trunk strength, and cigarette smoking. Persistence of disabling LBP has been associated with the presence of maladaptive pain coping behavior, nonorganic signs, functional impairment, poor general health status, and psychiatric comorbidities.

ANATOMY

The lumbar spine is composed of five vertebrae. Each vertebra consists of a body anteriorly and a neural arch that encloses the spinal canal posteriorly (Figure 47-1). Their cartilaginous end plates cover the superior and inferior surfaces of the vertebral body.

Adjacent vertebrae are united by an intervertebral disk. The outer circumference of the disk is made up of concentric layers of dense, tough fibrous tissue, the annulus fibrosus. The annulus encloses a shock-absorbing gelatinous nucleus pulposus. In addition to this discovertebral joint anteriorly, at each level of the lumbar spine, there are two posterolaterally placed synovial facet (apophyseal) joints. These are formed by articulation of the superior and inferior articular processes of adjacent vertebrae.

The vertebral column is further stabilized by ligaments and paraspinal muscles (erector spinae, trunk, and abdominal muscles). The anterior and posterior longitudinal ligaments run the length of the spinal column. They anchor the anterior and posterior vertebral body surfaces and intervertebral disks. The ligamentum flavum interconnects the laminae while the interspinous and supraspinous ligaments interconnect the spinous processes. The intertransverse ligaments interconnect the transverse processes.
The sacroiliac joints join the spinal column to the pelvis. The anterior and inferior part of the joint is lined with synovium, whereas the posterior and superior part is fibrous. There is little or no movement at the sacroiliac joint.

The spinal canal in the lumbar region contains the cauda equina (the bundle of lumbar and sacral nerve roots that occupy the vertebral canal below the cord), blood vessels, and fat. Because the spinal cord ends at the L1 level, cord compression is generally not a feature of lumbar pathology. At each level a pair of nerve roots leaves the spinal canal and exits through the intervertebral foramina.

**CLINICAL EVALUATION**

LBP has a large number of causes, and the spectrum of clinical presentation is broad. Many individuals will have self-limited episodes of acute LBP that resolve without specific treatment, whereas others may present with recurrent or chronic LBP. A thorough history is the most important part of the clinical evaluation of these patients. Imaging is often unnecessary.

**HISTORY**

The major focus in the initial evaluation of a patient with LBP is to identify the small fraction (<5%) of patients who may have neural compression or underlying systemic disease (infection, malignancy, or spondyloarthritis) as the cause of back pain. These patients require early diagnostic testing and may require specific treatment (e.g., antibiotics for vertebral osteomyelitis) or urgent treatment (e.g., surgical decompression in a patient with major or progressive neural compression). As such, clues to the presence of underlying systemic disease (Table 47-1), often referred to as “red flags,” should be carefully sought. It is also important to look for any social or psychologic distress such as job dissatisfaction, pursuit of disability compensation, and depression that may amplify or prolong the pain.

Mechanical LBP is due to an anatomic or functional abnormality in the spine that is not associated with inflammatory or neoplastic disease. It typically increases with physical activity and upright posture and tends to be relieved by rest and recumbency. More than 95% of LBP is mechanical, and degenerative change in the lumbar spine is the most common cause of mechanical LBP. Severe and acute mechanical LBP in a postmenopausal woman would be suspicious for a vertebral compression fracture secondary to osteoporosis. Nocturnal pain suggests the possibility of underlying infection or neoplasm as the cause of LBP.

Inflammatory LBP, as seen in the spondyloarthritides, is more common in men younger than 40 years of age. It is associated with marked morning stiffness that usually lasts for more than 30 minutes. The pain frequently improves with exercise but not with rest. Pain is often worse during the second half of the night, and some patients complain of alternating buttock pain.

It is important to ask the patient if the back pain radiates into the lower extremities, suggesting pseudoclaudication.

**Table 47-1 Red Flags for Potentially Serious Underlying Causes of Low Back Pain**

<table>
<thead>
<tr>
<th>Spinal Fracture</th>
<th>Significant trauma</th>
<th>Prolonged glucocorticoid use</th>
<th>Age &gt; 50 yr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infection or Cancer</strong></td>
<td>History of cancer</td>
<td>Unexplained weight loss</td>
<td>Immunosuppression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Injection drug use</td>
<td>Nocturnal pain</td>
</tr>
<tr>
<td><strong>Cauda Equina Syndrome</strong></td>
<td>Urinary retention</td>
<td>Overflow incontinence</td>
<td>Fecal incontinence</td>
</tr>
<tr>
<td></td>
<td>Bilateral or progressive motor deficit</td>
<td>Saddle anesthesia</td>
<td></td>
</tr>
<tr>
<td><strong>Spondyloarthritis</strong></td>
<td>Severe morning stiffness</td>
<td>Pain improves with exercise, not rest</td>
<td>Pain during second half of night</td>
</tr>
<tr>
<td></td>
<td>Alternating buttock pain</td>
<td>Age &lt; 40 yr</td>
<td></td>
</tr>
</tbody>
</table>

---

*Figure 47-1 Anatomy of the lumbar spine. A, Cross-sectional view through a lumbar vertebra. B, Lateral view of the lumbar spine.*
(neurogenic claudication) secondary to spinal stenosis or sciatica (usually secondary to a herniated disk). Young adults are more likely to experience disk herniations, and elderly patients are more likely to have spinal stenosis. Sciatica results from nerve root compression and produces pain in a dermatomal (radicular) distribution, usually to the level of the foot or ankle. The pain is lancinating, shooting, and sharp in quality. It is frequently accompanied by numbness and tingling and may be accompanied by sensory and motor deficits. Sciatica due to disk herniation typically increases with cough, sneezing, or the Valsalva maneuver. Sciatica should be differentiated from non-neurogenic sclerotomal pain. This pain can arise from pathology within the disk, facet joint, or lumbar paraspinal muscles and ligaments. Like sciatica, sclerotomal pain is often referred into the lower extremities, but unlike sciatica, sclerotomal pain is nondermatomal in distribution, it is dull in quality, and the pain usually does not radiate below the knee or have associated paresthesias. Most radiant pain is sclerotomal. Bowel or bladder dysfunction should suggest the possibility of the cauda equina syndrome.

**PHYSICAL EXAMINATION**

A physical examination usually does not lead to a specific diagnosis. Nevertheless, a general physical examination including a careful neurologic examination may help identify those few but critically important cases of LBP that are secondary to a systemic disease or have clinically significant neurologic involvement (see Table 47-1).

Inspection may reveal the presence of scoliosis. This can be either structural or functional. A structural scoliosis is associated with structural changes of the vertebral column and sometimes the rib cage as well. In adults structural scoliosis is usually secondary to degenerative changes, although some adults may have a history of adolescent idiopathic scoliosis. With forward flexion, structural scoliosis persists. In contrast, functional scoliosis, which usually results from paravertebral muscle spasm or leg length discrepancy, usually disappears. A tuft of hair in the lumbar spine region may indicate a congenital structural abnormality such as spina bifida occulta.

Palpation can detect paravertebral muscle spasm. This often leads to loss of the normal lumbar lordosis. Point tenderness on percussion over the spine has sensitivity but not specificity for vertebral osteomyelitis. A palpable step-off between adjacent spinous processes suggests spondylolisthesis.

Limited spinal motion (flexion, extension, lateral bending, and rotation) is not associated with any specific diagnosis because LBP due to any cause may limit motion. Range of motion measurements, however, can help in monitoring treatment. Chest expansion of less than 2.5 cm has specificity but not sensitivity for ankylosing spondylitis. The hip joints should be examined for any decrease in range of motion because hip arthritis, which normally causes groin pain, may occasionally present as LBP. Trochanteric bursitis with tenderness over the greater trochanter of the femur can be confused with LBP. The presence of more widespread tender points, especially in a female patient, suggests the possibility that LBP may be secondary to fibromyalgia.

In patients with a history of LBP that radiates into the lower extremities (sciatica, pseudoclaudication, or referred sclerotomal pain) a straight leg–raising test should be performed. With the patient lying on his or her back, the heel is placed in the palm of the examiner’s hand. With the knee fully extended the leg is raised progressively. This places tension on the sciatic nerve (that takes origin from L4, L5, S1, S2, and S3) and thereby stretches the nerve roots (especially L5, S1, and S2). If any of these nerve roots is already irritated, such as by impingement from a herniated disk, further tension on the nerve root by straight leg–raising will result in radicular pain that extends below the knee. The test is positive if radicular pain is produced when the leg is raised less than 70 degrees. Dorsiflexion of the ankle further stretches the sciatic nerve and increases the sensitivity of the test. Pain experienced in the posterior thigh or knee during straight leg–raising generally results from hamstring tightness and does not represent a positive test. The straight leg–raising test is sensitive but not specific for clinically significant disk herniation at the L4-5 or L5-S1 level (the sites of 95% of clinically meaningful disk herniations). False-negative tests are more frequently seen with herniation above the L4-5 level. The straight leg–raising test is usually negative in patients with spinal stenosis. The crossed straight leg–raising test (with sciatica reproduced when the opposite leg is raised) is highly specific but insensitive for a clinically significant disk herniation. The neurologic evaluation (Figure 47-2) of the lower extremities in a patient with sciatica can identify the specific nerve root involved. The evaluation should include motor testing with focus on dorsiflexion of the foot (L4), great toe dorsiflexion (L5), and foot plantar flexion (S1); determination of knee (L4) and ankle (S1) deep tendon reflexes; and tests for dermatomal sensory loss. The inability to toe walk (mostly S1) and heel walk (mostly L5) indicate muscle weakness. Muscle atrophy can be detected by circumferential measurements of the calf and thigh at the same level bilaterally.

Patients involved with litigation or with psychologic distress occasionally exaggerate their symptoms. They may display nonorganic signs where the objective findings do not match the subjective complaints such as with nonanatomic motor or sensory loss. A number of tests to detect this have been described by Waddell and co-workers. The most reproducible tests are the presence of superficial tenderness, overreaction during the examination, and observation of a discrepancy in the straight leg–raising test done in the seated and supine positions.

**DIAGNOSTIC TESTS**

**Imaging**

The major function of diagnostic testing, especially imaging, is the early identification of pathology in those few patients who have evidence of a major or progressive neurologic deficit and those in whom an underlying systemic disease is suspected (see Table 47-1). Otherwise, imaging is not required unless significant symptoms persist beyond 6 to 8 weeks. This approach avoids unnecessary early testing because more than 90% of the patients will have recovered spontaneously by 8 weeks. Furthermore, neither magnetic
resonance imaging (MRI) or plain radiographs obtained early in the course of LBP evaluation improves clinical outcome, predicts recovery course, or reduces the overall cost of care.\textsuperscript{2,16}

A significant problem with all imaging studies is that many of the anatomic abnormalities identified in patients with LBP are also commonly present in asymptomatic individuals and are frequently unrelated to the back pain.\textsuperscript{9} Often these abnormalities result from age-related degenerative changes, which begin to appear even in early adulthood and are among the earliest degenerative changes in the body.\textsuperscript{17} Although clinically challenging and sometimes impossible, one should refrain from making causal inferences based solely on imaging abnormalities in the absence of corresponding clinical findings because this may lead to unnecessary, invasive, and costly interventions.

Given the weak association between imaging abnormalities and symptoms, it is not surprising that in up to 85% of patients a precise pathoanatomic diagnosis with identification of the pain generator cannot be made.\textsuperscript{11} Patients should understand that the reason for imaging is to rule out serious conditions and that common degenerative findings are expected. Ill-considered attempts to make a diagnosis on the basis of imaging studies may reinforce the suspicion of serious disease, magnify the importance of nonspecific findings, and label patients with spurious diagnoses.

Plain radiographs and MRI are the major modalities used in the evaluation of patients with LBP. In patients with persistent LBP of greater than 6 to 8 weeks' duration despite standard therapies, radiography may be a reasonable first option if there are no symptoms suggesting radiculopathy or spinal stenosis.\textsuperscript{15} Anteroposterior and lateral views are usually adequate. Oblique views substantially increase radiation exposure and add little new diagnostic information. Gonadal radiation in a woman from a two-view radiograph of the lumbar spine is equivalent to radiation exposure from a chest radiograph taken daily for more than 1 year.\textsuperscript{16}

Abnormalities on radiography such as single-disk degeneration, facet joint degeneration, Schmorl's nodes (protrusion of the nucleus pulposus into the spongiosa of a vertebra), spondylolysis, mild spondylolisthesis, transitional vertebrae (the “lumbarization” of S1 or “sacralization” of L5), spina bifida occulta, and mild scoliosis are equally prevalent in individuals with and without LBP.\textsuperscript{8,9,19}

MRI without contrast is generally the best initial test for patients with LBP who require advanced imaging. It is the preferred modality for the detection of spinal infection and cancers, herniated disks, and spinal stenosis.\textsuperscript{8} MRI testing for LBP should largely be limited to patients in whom there is a suspicion of systemic disease (such as infection or malignancy), for the preoperative evaluation of patients who are surgical candidates on clinical grounds\textsuperscript{21,28} (e.g., the presence of a significant or progressive neurologic deficit), or for those patients with radiculopathy or spinal stenosis who are candidates for epidural corticosteroids.\textsuperscript{18} Disk abnormalities are commonly noted on MRI studies but often have little or no relationship with the patient's symptoms. A disk bulge is a symmetric, circumferential extension of disk material beyond the interspace. A disk herniation is a focal or asymmetric extension. Herniations are subdivided into protrusions and extrusions. Protrusions are broad-based, whereas extrusions have a “neck” so that the base is narrower than the extruded material. Bulges (52%) and protrusions (27%) are common in asymptomatic adults, but extrusions are rare.\textsuperscript{8} MRI with the intravenous contrast agent, gadolinium, may be useful for the evaluation of patients with prior back surgery (with no hardware present) to help in the differentiation of scar tissue from recurrent disk herniation.

MRI is generally preferred over computed tomography (CT) scanning in the evaluation of patients with LBP. However, when bone anatomy is critical, CT is superior. Unlike MRI, CT can safely be done in patients with a ferromagnetic implant, although imaging artifacts may make interpretation difficult. CT myelography is therefore sometimes preferred in patients with surgically placed spinal hardware.

Bone scanning is used primarily to detect infection, bony metastases, or occult fractures and to differentiate them from degenerative changes. Bone scans have limited specificity due to poor spatial resolution, and thus abnormal findings often require further confirmatory imaging such as MRI.

\textbf{Figure 47-2} Neurologic features of lumbosacral radiculopathy.
Electrodiagnostic Studies

Electrodiagnostic studies can be helpful in the evaluation of some patients with lumbosacral radiculopathy. The main procedures are electromyography and nerve conduction studies. These studies can confirm nerve root compression and define the distribution and severity of involvement. Whereas studies such as MRI can only provide anatomic information, electrodiagnostic studies provide physiologic information that may support or refute the findings on imaging. Electrodiagnostic testing is therefore mostly considered in patients with persistent disabling symptoms of radiculopathy where there is discordance between the clinical presentation and findings on imaging. Electromyography and nerve conduction studies can also be helpful in differentiating the limb pain of peroneal nerve palsy or lumbosacral plexopathy from that of L5 radiculopathy. These studies are also useful in evaluating possible factitious weakness. Electrodagnosis is unnecessary in a patient with an obvious radiculopathy. It should be noted that electromyographic abnormalities may persist for over a year following decompressive surgery.\textsuperscript{19}

Laboratory Studies

Laboratory studies are used mostly in identifying patients with systemic causes of LBP. A patient with normal blood cell counts, erythrocyte sedimentation rate, and radiographs of the lumbar spine is unlikely to have underlying infection or malignancy as the cause of LBP.\textsuperscript{11}

DIFFERENTIAL DIAGNOSIS

LBP usually originates from pathology within the lumbar spine or associated muscles and ligaments (Table 47-2). Rarely pain is referred to the back from visceral disease. In the vast majority of patients with LBP, the pain is mechanical.\textsuperscript{11} Degenerative change in the lumbar spine is the largest contributor to the mechanical causes of LBP\textsuperscript{8} (see Table 47-2) and indeed the most commonly identified cause of back pain.

Lumbar Spondylosis

The current common usage of the term lumbar spondylosis incorporates degenerative changes in both the anteriorly placed discovertebral joints and the posterolaterally placed facet joints.\textsuperscript{6} These degenerative or osteoarthritic changes are seen radiographically as disk or joint space narrowing, subchondral sclerosis, and osteophytosis (Figure 47-3). Imaging evidence of lumbar spondylosis is common in the general population, increases with age, and may be unrelated to back symptoms. Radiographic abnormalities such as single disk degeneration, facet joint degeneration, Schmorl’s nodes, mild spondylolisthesis, and mild scoliosis are equally prevalent in persons with and without back pain.\textsuperscript{22} The situation is further complicated by the observation that patients with severe mechanical LBP may have minimal radiographic changes, and conversely patients with advanced changes may be asymptomatic.

The clinical spectrum of mechanical LBP is wide. Patients may present with acute LBP (with recurrent attacks in some), whereas chronic LBP (often with periods of acute exacerbation) may develop in others. Somatic referral may lead to sclerotomal pain that radiates into the buttocks and lower extremities. Lumbar spondylosis predisposes patients to intervertebral disk herniation, spondylolisthesis, and spinal stenosis.

In some patients with facet joint osteoarthritis the pain may radiate into the buttock and posterior thigh, be alleviated with forward flexion, and be exacerbated by bending ipsilateral to the involved joint (facet syndrome).

The terms internal disk disruption and diskogenic low back pain are used interchangeably and remain controversial diagnoses.\textsuperscript{13} The disorder is diagnosed by provocative diskography. Following contrast injection into several disks in sequence, the radiographic appearance and induced pain at each level are assessed. If injection into a disk reproduces a patient’s usual LBP, the test is considered positive. Advocates of this technique interpret a positive diskogram as defining the particular disk as the primary pain generator, and spinal fusion or disk arthroplasty is frequently recommended.\textsuperscript{2} However, injection into a disk can simulate the quality and location of pain known not to originate from that disk.\textsuperscript{24} Furthermore, diskographic abnormalities and induced pain are frequently seen in asymptomatic persons and, more importantly, the diskogenic pain attributed to disk disruption frequently improves spontaneously.\textsuperscript{8,11} Therefore the clinical importance and appropriate management of this condition remains unclear.

Focal high signal in the posterior annulus fibrosus as seen on T2-weighted MRI images, sometimes referred to as a

<table>
<thead>
<tr>
<th>Table 47-2 Causes of Low Back Pain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mechanical</strong></td>
</tr>
<tr>
<td>Lumbar spondylosis*</td>
</tr>
<tr>
<td>Disk herniation*</td>
</tr>
<tr>
<td>Spondylolisthesis*</td>
</tr>
<tr>
<td>Spinal stenosis*</td>
</tr>
<tr>
<td>Fractures (mostly osteoporotic)</td>
</tr>
<tr>
<td>Nonspecific (idiopathic)</td>
</tr>
<tr>
<td><strong>Neoplastic</strong></td>
</tr>
<tr>
<td>Primary</td>
</tr>
<tr>
<td>Metastatic</td>
</tr>
<tr>
<td><strong>Inflammatory</strong></td>
</tr>
<tr>
<td>Spondyloarthritides</td>
</tr>
<tr>
<td><strong>Infectious</strong></td>
</tr>
<tr>
<td>Vertebral osteomyelitis</td>
</tr>
<tr>
<td>Epidural abscess</td>
</tr>
<tr>
<td>Septic diskitis</td>
</tr>
<tr>
<td>Herpes zoster</td>
</tr>
<tr>
<td><strong>Metabolic</strong></td>
</tr>
<tr>
<td>Osteoporotic compression fractures</td>
</tr>
<tr>
<td>Paget’s disease</td>
</tr>
<tr>
<td><strong>Referred Pain to Spine</strong></td>
</tr>
<tr>
<td>From major viscera, retroperitoneal structures, urogenital system, aorta, or hip</td>
</tr>
</tbody>
</table>

*Related to degenerative changes.
occurs at L4-5\(^2\) (with more torsion at the L4-5 level). Probably related to this, 90% to 95% of clinically significant compressive radiculopathies occur at these two levels.\(^1\)

Disk herniation is rare in young individuals with the frequency increasing with age. The peak frequency of herniation at the L5-S1 and L4-L5 levels is between the ages of 44 and 50 with a progressive decline in frequency thereafter.\(^2\)

The genesis of sciatica is felt to have both a mechanical (disk material impinging on a nerve root) and biologic component. Inflammation, vascular invasion, immune responses, and an array of cytokines have been implicated.

**Figure 47-3**  Lumbar spondylosis. Anteroposterior (A) and lateral (B) radiographs of the lumbar spine show the cardinal features of disk-space narrowing, marginal osteophytes, and end plate sclerosis. (Courtesy Dr. John Crues, University of California, San Diego.)

**Figure 47-4**  Schematic drawing showing posterolateral disk herniation resulting in nerve root impingement.

**Disk Herniation**

Intervertebral disk herniation occurs when the nucleus pulposus in a degenerated disk prolapses and pushes out the weakened annulus, usually posterolaterally. Imaging evidence of disk herniation has a high prevalence in the general population with one study finding MRI evidence of disk herniation in 27% of asymptomatic individuals.\(^2\) Occasionally, however, the herniated disk can cause nerve root impingement leading to lumbosacral radiculopathy (Figures 47-4 and 47-5). A herniated intervertebral disk is the most common cause of sciatica.\(^3\)

The lumbosacral spine is susceptible to disk herniation because of its mobility. Seventy-five percent of flexion and extension occurs at the lumbosacral joint (L5-S1), and 20% occurs at L4-5\(^2\) (with more torsion at the L4-5 level). Probably related to this, 90% to 95% of clinically significant compressive radiculopathies occur at these two levels.\(^1\)

Disk herniation is rare in young individuals with the frequency increasing with age. The peak frequency of herniation at the L5-S1 and L4-L5 levels is between the ages of 44 and 50 with a progressive decline in frequency thereafter.\(^2\)

The genesis of sciatica is felt to have both a mechanical (disk material impinging on a nerve root) and biologic component. Inflammation, vascular invasion, immune responses, and an array of cytokines have been implicated.
Spondylolisthesis

Spondylolisthesis is the anterior displacement of a vertebra on the one beneath it. There are two major types: isthmic and degenerative.

Isthmic spondylolisthesis (Figure 47-6) is caused by bilateral spondylolysis. Spondylolysis is a defect in the pars interarticularis that is most commonly seen at L5. It is typically a fatigue fracture acquired early in life that is more commonly seen in boys. Spondylolysis progresses to spondylolisthesis in approximately 15% of patients.

Degenerative spondylolisthesis develops in some patients with severe degenerative changes with subluxation at the facet joints allowing anterior or posterior movement of one vertebra over another. It is usually seen in an older age group (typically older than age 60), is more common in women, and most frequently involves the L4-5 level.

Most patients, especially those with a minor degree of spondylolisthesis, are asymptomatic. Some may complain of an aching mechanical LBP. Neurologic complications may occur in some with greater degrees of spondylolisthesis. Nerve root impingement is more likely to be seen in patients with isthmic spondylolisthesis (especially L5 nerve root), whereas in degenerative spondylolisthesis the more likely clinical presentation is of spinal stenosis. Rarely extreme slippage results in cauda equina syndrome. In view of its potential dynamic nature, spondylolisthesis may be missed if standing radiographs are not obtained.

Spinal Stenosis

Lumbar spinal stenosis is defined as a narrowing of the central spinal canal, its lateral recesses, and neural foramina that may result in a compression of lumbosacral nerve roots. Spinal stenosis can occur at one or multiple levels, and the
narrowing may be asymmetric. It is important to recognize that 20% to 30% of asymptomatic adults older than age 60 have imaging evidence of spinal stenosis. The prevalence of symptomatic lumbar spinal stenosis is not established. It is, however, the most frequent indication for spinal surgery in patients older than age 65.

Congenital idiopathic spinal stenosis (Table 47-3) is not uncommon and results from congenitally short pedicles. These patients tend to become symptomatic early (third to fifth decade of life) when superimposed mild degenerative changes that would normally be tolerated result in sufficient further narrowing of the spinal canal to cause symptoms.32

Degenerative changes are the cause of spinal stenosis in the vast majority of cases. The intervertebral disk loses height as it degenerates. This results in a bulging or buckling of the now redundant and often hypertrophied ligamentum flavum into the posterior part of the canal. Any herniation of the degenerated disk narrows the anterior part of the canal while hypertrophied facets and osteophytes may compress nerve roots in the lateral recess or intervertebral foramen (Figures 47-7 and 47-8). Any degree of spondylolisthesis will further exacerbate spinal canal narrowing.

The hallmark of spinal stenosis is neurogenic claudication (pseudoclaudication). The symptoms of neurogenic claudication are usually bilateral but often asymmetric. The primary complaint is of pain in the buttocks, thighs, and legs. The pain may be accompanied by paresthesias. Neurogenic claudication is induced by standing erect or walking and relieved by sitting or flexing forward. This forward flexion increases the spinal canal dimensions and may lead to the patient adopting a simian stance. It is therefore not

<table>
<thead>
<tr>
<th>Table 47-3 Causes of Lumbar Spinal Stenosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Congenital</strong></td>
</tr>
<tr>
<td>Idiopathic</td>
</tr>
<tr>
<td>Achondroplastic</td>
</tr>
<tr>
<td><strong>Acquired</strong></td>
</tr>
<tr>
<td>Degenerative</td>
</tr>
<tr>
<td>Hypertrophy of facet joints</td>
</tr>
<tr>
<td>Hypertrophy of ligamentum flavum</td>
</tr>
<tr>
<td>Disk herniation</td>
</tr>
<tr>
<td>Spondylolisthesis</td>
</tr>
<tr>
<td>Scoliosis</td>
</tr>
<tr>
<td><strong>Iatrogenic</strong></td>
</tr>
<tr>
<td>Postlaminectomy</td>
</tr>
<tr>
<td>Postsurgical fusion</td>
</tr>
<tr>
<td><strong>Miscellaneous</strong></td>
</tr>
<tr>
<td>Paget’s disease</td>
</tr>
<tr>
<td>Fluorosis</td>
</tr>
<tr>
<td>Diffuse idiopathic skeletal hyperostosis</td>
</tr>
<tr>
<td>Ankylosing spondylitis</td>
</tr>
</tbody>
</table>

Figure 47-6  A, Spondylolysis with bilateral defects in the pars interarticularis (arrows). B, Spondylolysis of the L5 vertebra (arrow) resulting in isthmic spondylolisthesis at L5-S1.

Figure 47-7  Spinal stenosis secondary to a combination of disk herniation (A), facet joint hypertrophy (B), and hypertrophy of the ligamentum flavum (C).
symptoms remained unchanged in 70%, improved in 15%, and worsened in 15%. As such, prophylactic surgical intervention is not warranted.

Diffuse Idiopathic Skeletal Hyperostosis

Diffuse idiopathic skeletal hyperostosis (DISH) is characterized by calcification and ossification of paraspinous ligaments and the entheses. It is a noninflammatory condition of unknown etiology that is not associated with HLA-B27 positivity. DISH has been associated with obesity, diabetes mellitus, and acromegaly. It is rarely diagnosed before the age of 30, is more commonly seen in men, and the prevalence rises with age.

The thoracic spine is most commonly involved, although the cervical and lumbar regions may also be affected. Ossification of the anterior longitudinal ligament is best seen on a lateral radiograph of the thoracic spine. This together with bridging enthesophytes in the spine give the appearance of flowing wax on the anterior and right lateral aspects of the spine. Involvement of the left lateral aspect in patients with situs inversus has led to speculation that the descending aorta plays a role in the location of the calcification. Intervertebral disk spaces and facet joints are preserved (unless there is coexisting lumbar spondylosis) and the sacroiliac joints appear normal. This helps differentiate DISH from spondylosis and the spondyloarthritides. Almost any extraspinal osseous or articular site may be affected. Irregular new bone formation (“whiskering”) is often best seen at the iliac crests, ischial tuberosities, and femoral trochanters. Ossification of tendons and ligaments at sites of attachment (such as the patella, olecranon process, and calcaneus) and periarticular osteophytes (such as the lateral acetabulum

Figure 47-8 Degenerative spinal stenosis. A, The sagittal T2-weighted magnetic resonance image shows decreased anteroposterior diameter of the neural canal at the L4-5 level due to redundancy of the ligamentum flavum. B, The axial image through the L4-5 disk shows decreased cross-sectional area of the thecal sac from hypertrophic changes of the facet joints posterolateral to the thecal sac. (Courtesy Dr. John Crues, University of California, San Diego.)
and inferior portion of the sacroiliac joint on pelvic radiographs) may also be seen. Severe ligamentous calcification may be seen in the sacrotuberous and iliolumbar ligaments and heterotopic bone formation following hip replacement in patients with DISH has been described.40

DISH may be entirely asymptomatic. The most common complaint encountered is of pain and stiffness involving the spine, often the thoracic region. Usually there is only a moderate limitation of spinal motion. Extensive ossification of the anterior longitudinal ligament together with large anterior enthesophytes may occasionally compress the esophagus and cause dysphagia.46 Ossification of the posterior longitudinal ligament is almost exclusively seen in the cervical spine and may occur either as a discrete disorder or as part of DISH. This can rarely lead to cervical myelopathy. Pain and tenderness may be present at the entheses, and these patients may have findings of lateral or medial humeral epicondylitis, Achilles tendinitis, or plantar fasciitis.

If treatment of DISH is necessary at all, it is symptomatic. Most patients respond to acetaminophen, nonsteroidal anti-inflammatory drugs (NSAIDs), and judicious use of glucocorticoid injections for painful enthesopathy.

Nonspecific Low Back Pain

This is also referred to as idiopathic LBP. As mentioned earlier, a precise pathoanatomic diagnosis, with identification of the pain generator, cannot be made in up to 85% of the patients. This is largely because of the nonspecific nature of the symptoms in patients with LBP and the weak association of these symptoms with findings on imaging. Thus terms such as lumbago, strain, and sprain have come into use. Strain and sprain have never been histologically characterized. Therefore nonspecific LBP is a more accurate label for these patients who have a mostly self-limited syndrome of acute mechanical LBP. The severity of pain can vary from mild to severe, and whereas sometimes the back pain develops immediately after a traumatic event such as lifting a heavy object or a twisting injury, other patients may just wake up with LBP. Most patients are better within 1 to 4 weeks1 but remain susceptible to similar future episodes. Less than 10% of patients develop chronic nonspecific LBP.

Neoplasm

Neoplasms are an uncommon, but nevertheless important, cause of LBP. In a primary care setting, neoplasia accounts for less than 1% of the patients with LBP.11 In a large prospective study of patients in a walk-in clinic, a history of cancer, unexplained weight loss, failure to improve after 1 month of conservative therapy, and age older than 50 years were each associated with a higher likelihood for cancer.43 By far the most important predictor for the likelihood of underlying cancer as the cause of LBP was a prior history of cancer.

The typical patient with LBP secondary to spinal malignancy presents with a persistent and progressive pain that is not alleviated by rest and indeed is often worse at night. In some patients a spinal mass can result in a lumbar-sacral radiculopathy or cauda equina syndrome. Acute LBP may be the presentation in a patient with a pathologic compression fracture. Rarely, leptomeningeal carcinomatosis (in patients with breast cancer, lung cancer, lymphoma, or leukemia) may present with a lumbosacral polyradiculopathy.42 Most cases result from involvement of the spine by metastatic carcinoma4 (especially prostate, lung, breast, thyroid, or kidney) or multiple myeloma. Metastatic vertebral lesions, more commonly seen in the thoracic spine, account for 39% of bony metastases in patients with primary neoplasms.44 Spinal cord tumors, primary vertebral tumors, and retroperitoneal tumors may rarely be the cause of LBP.11

Osteoid osteoma, a benign tumor of bone, typically presents with LBP in the second or third decade of life. The pain is often accompanied by a functional scoliosis secondary to paravertebral spasm. Patients may present with pain even before the osteoid osteoma is visible radiographically. Osteoid osteomas predominantly involve the posterior elements of the spine, usually the neural arch. A sclerotic lesion measuring less than 1.5 cm with a lucent nidus is pathognomonic.44 A bone scan, CT scan, or MRI should be ordered if an osteoid osteoma is suspected but not detected on radiography.

Plain radiographs are less sensitive than other imaging tests in detecting neoplastic lesions because approximately 50% of trabecular bone must be lost before a lytic lesion is visible.4 Further, Metastatic lesions may be lytic (radiolucent), blastic (radiodense), or mixed. The majority of metastases are osteolytic. Vertebral bodies are primarily involved because of their rich blood supply associated with red marrow, and unlike infections the disk space is usually spared. It should be noted that a purely lytic lesion such as multiple myeloma will not be detected by a bone scan. MRI offers the greatest sensitivity and specificity in the evaluation of spinal tumors and is generally the modality of choice.

Infection

Vertebral osteomyelitis (spinal osteomyelitis, spondylodisitis) may be acute (usually pyogenic) or chronic (pyogenic, fungal, or granulomatous). Acute vertebral osteomyelitis evolves over a period of a few days or weeks and is the major focus of this discussion.

Vertebral osteomyelitis usually results from hematogenous seeding, direct inoculation at the time of spinal surgery, or contiguous spread from an infection in the adjacent soft tissue. The lumbar spine is the most common site of vertebral osteomyelitis followed by the thoracic and cervical spine.45 Staphylococcus aureus is the most common microorganism followed by Escherichia coli. Coagulase-negative staphylococci and Propionibacterium acnes are almost always the cause of exogenous osteomyelitis after spinal surgery, particularly if internal fixation devices are used.46 A source of infection is detected in about half the cases with endocarditis diagnosed in up to a third of cases of vertebral osteomyelitis.45 Other common sites for the primary focus of infection are the urinary tract, skin, soft tissue, a site of vascular access, bursitis, or septic arthritis.46 Most patients with hematogenous pyogenic vertebral osteomyelitis have underlying medical disorders such as diabetes, coronary artery disease, immunosuppressive disorders,
malignancy, and renal failure. Intravenous drug abuse is also a risk factor for vertebral osteomyelitis.

Vertebral osteomyelitis may be complicated by an epidural or paravertebral abscess. This may result in neurologic complication.

Back pain is the initial symptom in most patients. The pain tends to be persistent, present at rest, exacerbated by activity, and at times well localized. Point tenderness on percussion over the spine has sensitivity but not specificity for vertebral osteomyelitis. Fever is present in only about half of the patients, partly because most patients are using analgesic medications. Because most cases of vertebral osteomyelitis result from hematogenous seeding, the dominant manifestations initially may be of the primary infection. An epidural abscess may result in a radiculopathy or cauda equina syndrome.

Leukocytosis is seen in only about two-thirds of the patients. However, almost all the patients have increases in the erythrocyte sedimentation rate and C-reactive protein, with the latter best correlating with clinical response to therapy. If blood cultures are negative in a patient suspected of having vertebral osteomyelitis, a bone biopsy (CT-guided or open) with appropriate culture studies and histopathologic analysis is indicated.

Plain radiography is usually the initial imaging study. Radiographic changes, however, occur relatively late and are nonspecific. Typically there is loss of disk height and loss of cortical definition followed by bony lysis of adjacent vertebral bodies. MRI is the most sensitive and specific imaging technique to detect spinal infections. The classic finding of pyogenic osteomyelitis is involvement of two vertebral bodies with their intervening disk. In a patient with neurologic impairment, MRI should be done early to rule out an epidural abscess. Whenever possible, antimicrobial therapy should be directed against an identified susceptible pathogen. There are no data from randomized, controlled trials to guide decisions about specific antimicrobial regimens or the duration of therapy. Intravenous therapy of at least 4 to 6 weeks, and possibly additional oral antibiotic therapy, is usually recommended. Surgery may be necessary to drain an abscess, although CT-guided catheter drainage may be sufficient in some cases. Surgical débridement is always required when infection is associated with a spinal implant with removal of the implant whenever possible.

Tuberculosis and nontubercular granulomatous infections (blastomycosis, cryptococcosis, actinomycosis, coccidioidomycosis, and brucellosis) of the spine should be considered in the appropriate clinical and geographic setting.

Lumbar nerve roots are commonly involved in patients with herpes zoster. In most cases a single unilateral dermatome is involved. Pain is often severe and may precede the appearance of a maculopapular rash that evolves into vesicles and pustules.

Inflammation

The spondyloarthritides cause inflammatory LBP (see Table 47-1) and are discussed in detail elsewhere (see Chapters 74 to 78).

Metabolic Disease

The major consideration in this category is the occurrence of acute mechanical LBP secondary to a vertebral compression fracture in a patient with osteoporosis (Chapter 101). Most patients are postmenopausal women.

Paget’s disease of bone (Chapter 101) is most often detected in an asymptomatic patient by the incidental finding of either an elevated alkaline phosphatase or characteristic radiographic abnormality. The spine is the second most commonly affected site after the pelvis. Within the spine the L4 and L5 vertebrae are most commonly involved. Paget’s disease of the spine may involve single or multiple levels. The vertebral body is almost always involved together with a variable portion of the neural arch. Radiographically Paget’s disease is seen as areas of enlargement of the bone with thickened, coarsened trabeculae. Usually a mixed picture of sclerotic and lytic Paget’s disease is encountered. The vertebrae may enlarge, weaken, and fracture. LBP may occur due to the pagetic process itself (with peristomial stretching and vascular engorgement), microfractures, overt fractures, secondary osteoarthritic of the facet joints, spondylolysis with or without spondylolisthesis, or sarcomatous transformation (rare). Neurologic complications secondary to Paget’s disease of the lumbar spine include sciatica secondary to nerve root impingement, spinal stenosis, and rarely a cauda equina syndrome.

Visceral Pathology

Disease in organs that share segmental innervation with the spine can cause pain to be referred to the spine. In general, pelvic diseases refer pain to the sacral area, lower abdominal diseases to the lumbar area, and upper abdominal diseases to the lower thoracic spine area. Local signs of disease such as tenderness to palpation, paravertebral muscle spasm, and increased pain on spinal motion are absent.

Vascular, gastrointestinal, urogenital, or retroperitoneal pathology may on occasion cause LBP. A partial list of causes includes an expanding aortic aneurysm, pyelonephritis, ureteral obstruction due to renal stones, chronic prostatitis, endometriosis, ovarian cysts, inflammatory bowel disorders, colonic neoplasms, and retroperitoneal hemorrhage (usually in a patient taking anticoagulants). Most abdominal aortic aneurysms are asymptomatic but may become painful as they expand. Aneurysmal pain is usually a harbinger of rupture. Rarely the aneurysm may develop leakage. This produces severe pain with abdominal tenderness. Most patients with aortic dissection present with a sudden onset of severe “tearing” pain in the chest or upper back. Pain originating from a hollow viscus such as the ureter or colon is often colicky.

Miscellaneous

LBP may be part of the clinical spectrum in innumerable conditions. It would not be practical or useful to discuss these entities here. Considered next are some of the more important or controversial causes of LBP.

The piriformis syndrome is felt to be an entrapment neuropathy of the sciatic nerve related to anatomic variations in the muscle-nerve relationship or to overuse. The
piriformis is a narrow muscle that originates from the anterior part of the sacrum and inserts into the greater trochanter. It is an external rotator of the hip. Unlike sciatica from lumbosacral nerve root compression, the pain is not restricted to a specific dermatome. The straight leg-raising test is usually negative. There may be tenderness over the sciatic notch. Physical examination maneuvers for the diagnosis of piriformis syndrome are based on the notion that stretching the irritated piriformis muscle may provoke sciatic nerve compression. This can be done by internally rotating the hip (Freiburg's sign) or by flexion, adduction, and internal rotation (FAIR maneuver) of the hip. Physical therapy that focuses on stretching the piriformis muscle and NSAIDs are generally the treatments offered.

Sacroiliac joint dysfunction is a controversial diagnosis. It is a term used to describe pain in the sacroiliac region related to abnormal sacroiliac joint movement or alignment. However, tests of pelvic symmetry or sacroiliac joint movement have low intertester reliability and fluoroscopically guided sacroiliac joint injections have been unreliable in diagnosis and treatment. Radiographic degenerative changes of the sacroiliac joint are often noted in the evaluation of patients with LBP. It remains unresolved as to whether these changes are the primary cause of the back pain.

Lumbosacral transitional vertebrae include sacralization of the lowest lumbar vertebral body and lumbarization of the uppermost sacral segment. The association of these variants with LBP remains controversial.

A “back mouse” is a mobile subcutaneous fibro-fatty nodule in the lumbosacral area. The nodule may be tender. Although there are case reports, the association with LBP remains unproven.

Epidural lipomatosis may be seen in obese patients, but it is more commonly seen as a rare side effect of long-term use of corticosteroids. There is an increase in epidural adipose tissue that causes a narrowing of the spinal canal. This is usually an incidental finding, although it may lead to compression of neural structures.

LBP during pregnancy is common. The pain usually starts between the fifth and seventh months of pregnancy. The etiology of LBP in pregnancy is unclear. Biomechanical, hormonal, and vascular factors have been implicated. Most women have resolution of their pain postpartum.

Fibromyalgia (see Chapter 52) and polymyalgia rheumatica (see Chapter 88) are two frequently encountered rheumatologic conditions in which LBP may be a prominent part of the clinical syndrome.

**TREATMENT**

Specific treatment is available only for the small fraction of patients with LBP who have either evidence of clinically significant neural compression or an underlying systemic disease (cancer, infection, visceral disease, and spondyloarthritis). In the vast majority of patients with LBP, either the precise pathoanatomic cause (i.e., the pain generator) cannot be determined or, when the cause is determined, no specific treatment is available. These patients are managed with a conservative program centered on analgesia, education, and physical therapy. The goal of treatment is relief of pain and restoration of function. Surgery is rarely necessary (Figure 47-9).

One should be wary of the proliferation of unproven medical, surgical, and alternative therapies. Most have not been rigorously tested in well-designed randomized controlled trials. Uncontrolled studies can produce a misleading impression of efficacy due to fluctuating symptoms and the largely favorable natural history of LBP in most patients.

For management purposes, patients with LBP are considered to have either acute LBP (duration <3 months), chronic LBP (duration >3 months), or a nerve root compression syndrome.

**Acute Low Back Pain**

The typical patient seeks medical attention for sudden onset of severe mechanical LBP. Examination usually reveals paravertebral muscle spasm, often resulting in loss of the normally present lumbar lordosis and severe decrease in range of motion secondary to pain. The prognosis for acute LBP is excellent. Indeed, only about a third of these patients seek medical care and more than 90% recover within 8 weeks or earlier.

Patients with acute LBP are advised to stay active and continue ordinary daily activities within the limits permitted by pain. This leads to more rapid recovery than bed rest. Bed rest of more than 1 or 2 days is discouraged.

Pharmacologic therapy is used for symptomatic relief. Unfortunately, no medication has consistently been shown to result in large average benefits on pain and evidence of beneficial effects on function is even more limited. Acetaminophen and NSAIDs are first-line options for analgesia. Short-term use of opioids is reasonable in patients with severe disabling LBP or in those at high risk of complications due to NSAIDs. For patients with acute LBP, short-acting opioids are generally recommended. Muscle relaxants are moderately effective for short-term symptomatic relief but have a high prevalence of adverse effects including drowsiness and dizziness. It is unclear whether these medications truly relax muscles or their effects are related more to sedation or other nonspecific effects. Benzodiazepines have similar efficacy to muscle relaxants for short-term pain relief but are associated with risks for abuse, addiction, and tolerance.

Back exercises are not helpful in the acute phase, and a physical therapy referral is usually unnecessary in the first month. Later an individually tailored program that focuses on core strengthening, stretching exercises, aerobic conditioning, functional restoration, and loss of excess weight is recommended to prevent recurrences. The purpose of back exercises is to stabilize the spine by strengthening trunk muscles. Flexion exercises strengthen the abdominal muscles, and extension exercises strengthen the paraspinal muscles. Numerous exercise programs have been developed and appear to be equally effective.

Patient education including use of education booklets is strongly recommended. The information provided should include causes of LBP, basic anatomy, favorable natural
Injection therapy is used mostly in subacute (>6 weeks) and chronic LBP. Epidural corticosteroid injections have gained remarkable, but unjustified, popularity. The rationale for their use is that the genesis of radicular pain, when a herniated disk impinges on a nerve root, is at least partly related to locally induced inflammation. Indeed, there is evidence of moderate benefit compared with placebo injection for short-term relief of leg pain in patients with radiculopathy due to a herniated nucleus pulposus. However, epidural corticosteroid injections offer no significant functional benefit, nor do they reduce the need for surgery. It is important to note that there is no evidence for the effectiveness of epidural corticosteroid injections in LBP patients without radiculopathy.

A variety of other injection therapies using glucocorticoids or anesthetic agents, often in combination, are used in individuals with LBP with or without radicular pain and other symptoms in the leg. These include injection of trigger points, ligaments, sacroiliac joints, facet joints, and intradiscal steroid injections. There is no convincing evidence of

history, minimal value of diagnostic testing, importance of remaining active, effective self-care options, and coping techniques.

Spinal manipulation is provided mainly by chiropractors and osteopaths. It may involve low-velocity mobilization or manipulation with a high-velocity thrust that stretches spinal structures beyond the normal range and is frequently accompanied by a cracking or popping sound. For acute LBP, current evidence suggests that manipulative therapy is no more effective than conventional medical therapy. There is no evidence that ongoing manipulation reduces the risk of recurrence of LBP. There is insufficient evidence regarding the efficacy of massage and acupuncture in the treatment of acute LBP.

Application of heat by heating pads or blankets is a reasonable self-care option for short-term relief of acute LBP. There is, however, insufficient evidence to recommend application of cold packs or the use of corsets and braces. Traction provides no significant benefit for LBP patients with or without sciatica.
the efficacy of these interventions.\textsuperscript{58,59} Medial branch block for presumed facet joint pain and nerve root blocks for therapeutic or diagnostic purposes are also not recommended.\textsuperscript{50} Unfortunately these invasive and expensive procedures are commonly used in interventional pain clinics.

A number of physical therapy modalities are currently used in the treatment of patients with subacute and chronic LBP. These include transcutaneous electrical nerve stimulation (TENS), percutaneous electrical nerve stimulation, interferential therapy, low-level laser therapy, shortwave diathermy, and ultrasound. There is insufficient evidence of efficacy to recommend their use.

Vertebral compression fractures secondary to osteoporosis are common. There is resolution of pain with fracture healing within a few weeks in most patients. Vertebroplasty and balloon kyphoplasty are two increasingly popular, invasive, and expensive procedures that are being used for the treatment of persistent pain associated with these fractures. Both procedures involve the percutaneous placement of needles into the vertebral body through or lateral to the pedicles, as well as the injection of bone cement to stabilize the fracture. Kyphoplasty differs from vertebroplasty in that the cement is injected into a void in the vertebral body created by inflation of a balloon. Several early studies had suggested a positive treatment effect for vertebroplasty.\textsuperscript{50} However, two blinded, randomized, placebo-controlled trials of vertebroplasty for painful osteoporotic spinal fractures found no beneficial effect of vertebroplasty as compared with a sham procedure.\textsuperscript{51,62} Therefore on the basis of current evidence, the routine use of vertebroplasty or indeed kyphoplasty for relief of pain from osteoporotic compression fractures cannot be justified.

**Chronic Low Back Pain**

The clinical spectrum in patients with chronic LBP is wide. Some complain of severe, unrelenting pain, but most have a nagging mechanical LBP that may radiate into the buttocks and upper thighs. Patients with chronic LBP may experience periods of acute exacerbation. These exacerbations are managed according to the principles discussed earlier. A significant number of patients with chronic LBP remain functional and continue working, but overall the results of treatment are unsatisfactory and complete relief of pain is unrealistic for most. Patients with chronic LBP are largely responsible for the high costs associated with LBP. It is therefore incumbent on physicians who treat these patients to judiciously use proven therapies.

For most patients, first-line medication options are acetaminophen or NSAIDs. They may provide some degree of analgesia, but the evidence for their long-term efficacy is not compelling. Opioid analgesics or tramadol are an option when used judiciously in patients with severe disabling pain. Because of substantial risks including aberrant drug-related behaviors with long-term use in patients vulnerable to abuse or addiction, potential benefits and harms of opioid analgesics should be carefully weighed before starting therapy.\textsuperscript{18,63} There is no evidence that long-acting, around-the-clock dosing is more effective than short-acting or as-needed dosing, and continuous exposure to opioids could induce tolerance and lead to dose escalations.\textsuperscript{5} Muscle relaxants are not recommended for long-term use in patients with chronic stable LBP. Antidepressants that inhibit norepinephrine uptake are thought to have pain-modulating properties independent of their effects on depression. As such, low-dose tricyclic antidepressants are an option for chronic LBP, although the treatment effect is small and adverse side effects are common.\textsuperscript{5} There is no evidence of efficacy of selective serotonin reuptake inhibitors for LBP. Depression is, however, common in patients with chronic LBP and should be treated appropriately. Duloxetine, a serotonin-norepinephrine reuptake inhibitor, may have marginal efficacy in patients with chronic LBP.\textsuperscript{64,65} There is insufficient evidence to recommend antiepileptic medications such as gabapentin and topiramate for pain relief in patients with LBP with or without radiculopathy.\textsuperscript{5}

An individually tailored physical therapy program and patient education, as discussed in the section earlier on the treatment of acute LBP, are particularly important aspects in the management of a patient with chronic LBP. The use of physical therapy modalities and injection techniques (as discussed earlier) is not recommended for patients with chronic LBP. Lumbar supports and traction are ineffective. For most patients with LBP a medium-firm mattress or a back-conforming mattress (waterbed or foam) may be superior to a firm mattress.\textsuperscript{66,67}

A number of physical treatments have been used in treating chronic LBP. Spinal manipulation has been shown to be superior to sham manipulation but is no more effective than conservative medical therapy.\textsuperscript{68} There is less evidence for the efficacy of massage and acupuncture.\textsuperscript{68}

There has been a proliferation of nonsurgical interventional therapies for back pain. Chemonucleolysis is used for the treatment of herniated disks with intradiskal injections of chymopapain (extracted from papaya). Chymopapain enzymatically digests the nucleus pulposus while leaving the annulus fibrosus intact. Potentially life-threatening anaphylactic reactions have occurred rarely. Chemonucleolysis has lost favor in the United States but remains popular in Europe. Radiofrequency denervation has most commonly been used for the treatment of presumed facet joint pain by targeting the medial branch of the primary dorsal ramus. It involves fluoroscopic placement of an electrode near the nerve and application of heat by using a radiofrequency current to coagulate the nerve. There is a lack of convincing evidence about the effectiveness of this invasive procedure.\textsuperscript{58} Intradiskal electrothermal therapy (IDET) and percutaneous intradiskal radiofrequency thermocoagulation (PIRFT) involve placement of an electrode into the intervertebral disk of patients with presumed diskogenic pain and using electric or radiofrequency current to provide heat to thermocoagulate and shrink intradiskal tissue and destroy nerves. Current evidence does not support the use of IDET or PIRFT.\textsuperscript{58,69} Prolotherapy (also referred to as sclerotherapy) involves repeated injections of an irritant sclerosing agent into ligaments and tendinous attachments. It is based on the hypothesis that back pain in some patients stems from weakened ligaments and repeated injections of a sclerosing agent will strengthen the ligaments and reduce pain. On the basis of trial data, a guideline from the American Pain Society recommends against prolotherapy for chronic LBP.\textsuperscript{58}

Spinal cord stimulation is a procedure involving the placement of electrodes, percutaneously or by laminectomy, in the epidural space adjacent to the area of the
spine presumed to be the source of pain and applying an electric current in order to achieve sympatholytic and other neuromodulatory effects.\textsuperscript{58} Power for the spinal cord stimulator is supplied by an implanted battery. Spinal cord stimulation is associated with a greater likelihood for pain relief compared with reoperation or conventional medical management in patients with failed back surgery syndrome with persistent radiculopathy.\textsuperscript{58} At present there is no good evidence for the use of spinal cord stimulation for chronic LBP not related to the failed back surgery or failed back surgery syndrome without radiculopathy. Approximately a third of the patients involved in studies have experienced a complication following spinal cord stimulation including electrode migration, infection, wound breakdown, and lead- and generator pocket–related complications.\textsuperscript{58}

Intraspinal drug infusion systems, using a subcutaneously implanted pump with attached catheter, have been used in some patients with chronic intractable LBP for the intrathecal delivery of analgesics, usually morphine. Adequate evidence to support this intervention is not available.

Chronic LBP is a complex condition that involves biologic, psychologic, and environmental factors. For patients with persistent and disabling nonradicular LBP despite recommended noninterdisciplinary therapies, the clinician should strongly consider intensive interdisciplinary rehabilitation with an emphasis on cognitive-behavioral therapy.\textsuperscript{59} Interdisciplinary rehabilitation (also called multidisciplinary therapy) is an intervention that combines and coordinates physical, vocational, and behavioral components and is provided by multiple health professionals with different clinical backgrounds. Cognitive-behavioral therapy is a psychotherapeutic intervention that involves working with cognitions to change emotions, thoughts, and behaviors. There is strong evidence of improved function and moderate evidence of pain improvement with intensive interdisciplinary rehabilitation programs.\textsuperscript{21} Functional restoration (also called work hardening) is an intervention that involves simulated or actual work in a supervised environment in order to enhance job performance skills and improve strength, endurance, flexibility, and cardiovascular fitness in injured workers. When combined with a cognitive-behavioral component, functional restoration is more effective than standard care alone for reducing time lost from work.\textsuperscript{68}

As previously discussed, the precise identification of the pain generator in an LBP patient with degenerative changes involving the lumbar spine and no radicular pain is usually not possible in contradistinction to the patient with radicular symptoms. It is therefore not surprising that as a general rule the results of back surgery are disappointing when the goal is relief of back pain rather than relief of radicular symptoms resulting from neurologic compression. As such, the role of surgical treatment for chronic disabling LBP without neurologic involvement in patients with degenerative disease remains controversial. The most common surgery performed is spinal fusion. Interbody fusion is achieved from either a posterior or an anterior approach or both combined for a circumferential fusion. All fusion techniques involve placement of a bone graft between the vertebrae. Instrumentation refers to the use of hardware such as screws, plates, or cages that serve as an internal splint while the bone graft heals. Bone morphogenetic proteins are sometimes used to speed fusion. The rationale for fusion is based on its successful use at painful peripheral joints.

The current evidence is that for nonradicular back pain with degenerative changes, fusion is no more effective than intensive interdisciplinary rehabilitation but is associated with small to moderate benefits compared with standard nonsurgical care.\textsuperscript{70} Furthermore, the majority of patients who undergo surgery do not experience an optimal outcome defined as no pain, discontinuation or occasional pain medication use, and return of high-level function.\textsuperscript{59}

Lumbar disk replacement using a prosthetic disk is a newer alternative to fusion. Disk replacement is approved in the United States for patients with disease limited to one disk between L3-S1 and no spondylolisthesis or neurologic deficit. No data support the hypothetical advantage that, unlike spinal fusion, prosthetic disks will protect adjacent levels from further degeneration by preserving motion. At present there is insufficient evidence regarding long-term benefits and harms of disk replacement to support recommendations.

### Nerve Root Compression Syndromes

#### Disk Herniation

Patients with a herniated disk with radicular pain secondary to nerve root compression should be treated nonsurgically, as described in the section on acute LBP unless they have a serious or progressive neurologic deficit. Only about 10% of patients have sufficient pain after 6 weeks of conservative care that surgery is considered.\textsuperscript{11} A decision to continue with nonsurgical therapy beyond 6 weeks in these patients does not increase risk for paralysis or cauda equina syndrome.\textsuperscript{59} Surgery in these patients is associated with moderate short-term (through 6 to 12 weeks) benefits compared with nonsurgical therapy, though differences in outcome diminish with time and are generally no longer present after 1 to 2 years.\textsuperscript{56,59}

Open discectomy or microdiscectomy is the usual surgery performed on patients with serious or progressive neurologic deficit or electively on patients with persistent disabling pain secondary to radiculopathy (Table 47-4). Open discectomy generally involves a laminectomy, whereas microdiscectomy, using a smaller incision and an operating microscope, involves a hemilaminectomy to remove the disk fragment compressing the nerve root. There are no

<table>
<thead>
<tr>
<th>Table 47-4</th>
<th>Indications for Surgical Referral</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Disk Herniation</strong></td>
<td><strong>Cauda equina syndrome (emergency)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Serious neurologic deficit</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Progressive neurologic deficit</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Greater than 6 weeks of disabling radiculopathy (elective)</strong></td>
</tr>
<tr>
<td><strong>Spinal Stenosis</strong></td>
<td><strong>Serious neurologic deficit</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Progressive neurologic deficit</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Persistent and disabling pseudoclaudication (elective)</strong></td>
</tr>
<tr>
<td><strong>Spondylolisthesis</strong></td>
<td><strong>Serious or progressive neurologic deficit</strong></td>
</tr>
</tbody>
</table>
clear differences in the outcome between open diskectomy and microdiscectomy. There is insufficient evidence to evaluate the efficacy of sequestrectomy, or various laser-assisted, endoscopic, percutaneous, and other minimally invasive methods.20,71

Epidural corticosteroid injections may offer moderate benefit for short-term relief of radicular pain but do not offer significant functional benefit and do not reduce the need for surgery.27

Anti-tumor necrosis factor therapy is under investigation in patients with lumbar radiculopathy. A small randomized controlled trial with addition of a short course of adalimumab to the treatment regimen of patients with severe and acute sciatica resulted in a small decrease in leg pain and fewer surgical procedures.72 However, another randomized controlled trial found no difference between infliximab and a saline infusion.73

Spinal Stenosis

It is critical to understand the natural history of degenerative lumbar spinal stenosis before making treatment decisions. The symptoms of spinal stenosis remain stable for years in most patients and may improve in some. Dramatic improvement is uncommon. Even when symptoms progress, there is little likelihood of rapid deterioration of neurologic function. Therefore conservative nonoperative treatment is a rational choice for most patients.

There is a paucity of good data to guide the conservative management of lumbar spinal stenosis. Physical therapy is the mainstay of management, but evidence for the efficacy of specific standardized regimens is not available. Most regimens include core strengthening, stretching, aerobic conditioning, loss of excess weight, and patient education. Exercises that involve lumbar flexion such as bicycling are better tolerated. Strengthening of abdominal muscles may be helpful by promoting lumbar flexion and reducing lumbar lordosis. Lumbar corsets that maintain slight flexion may provide symptomatic relief. They should only be used for a limited number of hours a day to avoid atrophy of paraspinal muscles.

Acetaminophen, NSAIDs, and mild narcotic analgesics are used for symptomatic relief of pain.

Lumbar epidural corticosteroid injections are used on the assumption that symptoms may result from inflammation at the interface between the nerve root and compressing tissues.32 A small randomized controlled trial showed a reduction in pain and improvement in function at 6 months following use of epidural steroid injections.74 However, observational data suggest that epidural injections do not influence functional status or the need for surgery at 1 year.75

Surgery is indicated for the few patients with lumbar spinal stenosis who have a serious or progressive neurologic deficit. However, most surgery for lumbar spinal stenosis is elective. The indication for elective surgery is to relieve persistent and disabling symptoms of neurogenic claudication that have not responded to conservative care. In patients without fixed neurologic deficits, delayed surgery produces similar benefits to surgery selected as the initial treatment.52,76 The surgical goal is to decompress the central spinal canal and the neural foramina to eliminate pressure on the nerve roots. This is accomplished by laminectomy, partial facetectomy of hypertrophied facet joints, and excision of the hypertrophied ligamentum flavum and any protruding disk material. Laminectomy with lumbar fusion should generally be reserved for patients who have spinal stenosis with spondylolisthesis. Unfortunately there is an alarming increase in spinal fusion surgery with routine use of complex fusion techniques in the absence of evidence of greater efficacy. The techniques include instrumentation, bone graft augmentation with bone cement and human bone morphogenetic proteins, and combined anterior and posterior fusion (often at multiple levels). These techniques are associated with increased perioperative mortality, major complications, rehospitalization, and cost.77,79

Overall, for patients with spinal stenosis, with or without spondylolisthesis, who have disabling symptoms of neurogenic claudication despite conservative care, there is some evidence supporting the effectiveness of decompressive laminectomy in reducing pain and improving function through 1 to 2 years.32,59,70 Beyond this time frame the benefits appear to diminish.

A less invasive alternative to decompressive laminectomy is the implantation of a titanium interspinous spacer at one or two vertebral levels. This spacer distracts adjacent spinous processes and thereby imposes lumbar flexion, which in turn potentially increases the spinal canal dimensions. There is preliminary evidence of efficacy in patients with one- or two-level spinal stenosis, without spondylolisthesis, and with a history of relief of neurogenic claudication with flexion.70 There are no trials comparing the interspinous spacer with decompressive surgery.

Spondylolisthesis

The vast majority of patients with spondylolisthesis and chronic LBP are treated conservatively. Rarely a patient may need decompression surgery with fusion if a serious or progressive neurologic deficit develops from nerve root impingement or the patient develops disabling pseudoclaudication secondary to spinal stenosis. A randomized trial involving patients with isthmic spondylolisthesis and disabling isolated LBP or sciatica for at least a year suggested better results from fusion surgery than from nonsurgical care,70 although the differences in outcome narrowed over a 5-year follow-up period.81

OUTCOME

The natural history of acute LBP is favorable. There is substantial improvement in pain and function within a month in the majority of patients,1 and more than 90% are better at 8 weeks.4 Only about a third of patients with acute LBP seek medical care. Presumably the rest improve on their own. Relapses that also tend to be brief are common and may affect up to 40% of patients within 6 months.

Improvement is also the norm for patients with sciatica secondary to a herniated disk.32 A third of these patients are significantly better in 2 weeks, and 75% improve after 3 months.5 Only about 10% of these patients ultimately undergo surgery.

The symptoms of spinal stenosis tend to remain stable in 70%, improved in 15%, and worsened in 15%.35
The 7% to 10% of patients who develop chronic pain are largely responsible for the high costs associated with LBP and remain a major challenge. Factors that predict persistence of chronic disabling LBP include maladaptive pain coping behaviors, presence of nonorganic signs, functional impairment, poor general health status, psychiatric comorbidities, job dissatisfaction, disputed compensation claims, and a high level of “fear avoidance” (an exaggerated fear of pain leading to avoidance of beneficial activities). 23,83

SUMMARY

The outcome for most patients with LBP is good. The management of patients with chronic LBP, however, remains a challenge. The results of conservative and surgical management in these patients are unsatisfactory. There has been a proliferation and increasing utilization of a large number of expensive but unproven nonsurgical interventional techniques and physical therapy modalities.

Surgical intervention is indicated in the presence of a serious or progressive neurologic deficit. However, surgery in the absence of neurologic deficits, especially spinal fusion for degenerative changes, is controversial and not clearly effective. Rates of back surgery (including spinal fusion) in the United States are the highest in the world and continue to rise rapidly. 20 A particularly worrisome trend is the routine use of complex fusion techniques (with associated increased perioperative mortality, major complications, and cost) in the absence of evidence of greater efficacy. 7–10 The use of sham surgery in controlled trials is controversial for ethical reasons. However, randomized trials incorporating a sham operation may be justifiable to test the efficacy of spinal fusion because the surgery is not performed for a life-threatening condition, the primary clinical outcomes are subjective, and the rate of complications is high. 79

An Australian study indicated that a television campaign advising people with back pain to stay active and keep working reduced work-injury claims and medical expenses and had a sustained effect of altering physicians’ and patients’ perceptions regarding back pain. 84 Perhaps public health initiatives may help prevent episodes of LBP from becoming chronic and disabling. 23

References

Hip and Knee Pain

JAMES I. HUDDELESTON • STUART GOODMAN

It is estimated that musculoskeletal pain affects one-third to one-half of the general population. Disease is occurring as the baby boomers have reached middle age and beyond. This is exemplified by the increasing prevalence of hip and knee replacement operations, which rose by 16.2% to 884,400 procedures annually in the United States between 2002 and 2004. Furthermore, the prevalence of total knee and total hip arthroplasty is expected to double by 2016 and 2026, respectively. The hip and knee joints are two of the most commonly affected sites of musculoskeletal pain, with the prevalence of hip pain ranging from 8% to 30% in persons 60 years of age and older and the prevalence of knee pain ranging from 20% to 52% in persons 55 years of age or older. In general, women experience more musculoskeletal pain than men. There are also geographic and ethnic variations in the rates of both hip and knee pain. For example, there tends to be significantly less hip and knee pain with decreasing latitude, as well as significantly less hip pain and osteoarthritis in China than in the United States.

When evaluating complaints of knee or hip pain, knowledge of the anatomy of these joints is necessary for formulating a differential diagnosis. Given the thin soft tissue envelope around the knee and the fact that knee pain is rarely referred, the pain generators around the knee can often be elucidated with a complete history and thorough physical examination. Diagnosis of hip pain may be more challenging because the joint is deeper and the region is not infrequently the site of referred pain from the spine. An understanding of the basic biomechanics of these joints is also important in formulating a differential diagnosis because certain activities are likely to cause specific injuries.

This chapter focuses on the important aspects of the history, physical examination, and imaging modalities involved in evaluating patients with complaints of knee and hip pain. An appropriate, thorough workup of these patients will allow the clinician to formulate an accurate differential diagnosis in an efficient manner.

KNEE PAIN

History

A detailed history is perhaps the most important step in accurately diagnosing the cause of knee pain. Knee complaints generally fall into two broad categories, pain or instability. Pain may arise from injury to the articular surfaces (e.g., osteoarthritis, inflammatory arthritis, osteochondral defects, osteochondritis dissecans), torn menisci, quadriceps and patellar tendons, bursitis, nerve damage, fractures, neoplasia, or infection. Referred pain from the hip or spine is less common. Instability is usually episodic and stems from injuries to the quadriceps-patellar extensor mechanism, collateral ligaments, or cruciate ligaments. It is important to distinguish true instability from the common complaint of “giving way” because the latter is usually due to a robust pain response rather than specific structural pathology.

Patients in certain age groups tend to experience similar injuries. In patients younger than 40 years, ligament injuries, acute meniscus tears, and patellofemoral problems are frequently encountered. In contrast, degenerative conditions such as osteoarthritis and degenerative meniscal lesions tend to occur more frequently in older patients.

The location and character of the pain are particularly important when evaluating knee pain because many of the structures vital to proper knee function are subcutaneous and can be palpated easily. We prefer to conceptualize the knee as three separate compartments—medial, lateral, and patellofemoral. Each compartment should be examined separately. The patient should be able to point to the exact area where the pain is most severe. The onset of the pain...
should be determined. Osteoarthritis and inflammatory arthritis tend to have an insidious onset, whereas injuries to menisci and ligaments are usually associated with a traumatic event. Knowing the details of a traumatic event will be helpful. For example, a twisting injury, especially one sustained with a flexed knee, suggests a meniscus tear, whereas a noncontact knee injury associated with change of direction is more likely to produce a tear of the anterior cruciate ligament (ACL). Pain from degenerative arthritis tends to be associated with stiffness, is generally worse with ongoing activity during the day, and is exacerbated by exercise, stair climbing, getting up from a chair, getting in and out of a car, and so on.

The presence or absence of knee swelling is an important part of the history because knee effusions (fluid in the knee joint) usually accompany internal derangement. An effusion may also be present with synovitis, osteoarthritis, inflammatory arthritis, fractures, infection, and neoplasm. Distinguishing among soft tissue swelling around the knee, synovial thickening, and a true knee effusion is critical (see later). The timing or onset of the swelling is also important for determining the diagnosis. An acute cruciate or collateral ligament injury or osteochondral fracture will usually present with an acute hemarthrosis (occurring within an hour), whereas an effusion associated with arthritis tends to be more insidious in nature.

Complaints of “locking” are common. In a younger patient, locking may be due to a displaced meniscal tear. In older patients with degenerative arthritis, complaints of locking are often due to loose bodies. It is important to distinguish between true locking and diminished range of motion due to pain (so-called pseudolocking) because this distinction will determine which imaging studies are most appropriate.

Timing of the pain with activity is also important for making the correct diagnosis. Meniscus tears and ligament injuries leading to instability will be particularly troublesome with activities such as walking on uneven surfaces, stairs, and movements requiring knee flexion and pivoting. Osteoarthritis tends to be exacerbated by all load-bearing activities and relieved by rest.

The clinician should also explore the patient’s exercise tolerance and ability to perform activities of daily living. These details may give insight into the severity of the injury and will also guide treatment. Important details include the use of ambulatory assist devices (cane, crutches, walker, brace, and wheelchair), walking tolerance, and capability for other exercises (physical therapy).

A history of any previous treatments rendered should also be recorded. One’s response to physical therapy, analgesics, nonsteroidal anti-inflammatories, nutritional supplements (such as glucosamine and chondroitin), intra-articular injections of corticosteroids or hyaluronic acid derivatives, and any operative treatments will lend further insight into the accurate diagnosis and have implications for treatment once the diagnosis has been confirmed.

At the end of taking a detailed history, the clinician should be able to formulate a differential diagnosis with a short list of potential conditions. This information should then allow the physician to concentrate on specific aspects of a focused physical examination that will lead to confirmation of the diagnosis.

**Physical Examination**

**General**

After a brief overall assessment of the patient, the physical examination should begin with observation of the patient’s lower extremity coronal alignment and leg lengths. We prefer to have the patient stand with legs slightly apart while he or she faces the examiner (Figure 48-1). A goniometer is then used to measure the varus/valgus alignment of the knees. Evaluation of leg lengths should be performed with step blocks of known sizes. The total height of the blocks needed to make the iliac crests level with the floor is equivalent to the leg-length discrepancy (Figure 48-2).

Gait is examined next. Although a comprehensive discussion of gait analysis is beyond the scope of this chapter, all clinicians should routinely make a few basic observations when evaluating the patient with a knee problem. Antalgic gaits (shortened stance phase) and thrusts are commonly seen. Any disorder that causes lower extremity pain may cause an antalgic gait. Seen in the stance phase of gait, thrusts may be due to a progressive angular deformity secondary to degenerative changes or chronic ligamentous instability. Medial thrusts result from medial collateral ligament and/or posteromedial capsular laxity. Lateral thrusts arise from lateral collateral ligament or posterolateral corner laxity (Figure 48-3). Patients may also thrust into recurvatum (so called back-knee deformity) due to posterior capsular laxity or quadriceps weakness.

The patient should then transfer to the examination table for evaluation in a comfortable supine position. The examination should proceed with inspection and palpation before performing any provocative maneuvers. A pillow should be placed under the knee if full extension is not possible due to pain (e.g., fractures, displaced meniscus tears, large effusion). If there is no known pre-existing pathology, the contralateral knee can serve as an adequate control. The lower extremity should be inspected for any
passive range of motion of both knees should be recorded with a goniometer. The examiner should then proceed with palpation of all structures of the knee. It is important to do this in a systematic manner to ensure completeness. Palpation should be gentle but firm enough to detect subtle pathology. Structures to be palpated include the quadriceps tendon, the patella (superior and inferior poles), the pes anserinus bursa, the medial (Figure 48-6A) and lateral (Figure 48-6B) joint lines, the origins and insertions of the collateral ligaments, the tibial tubercle, and the popliteal fossa. Fullness in the posterior knee may be indicative of a Baker's cyst.

**Ligaments**

Injuries to the collateral or cruciate ligaments may lead to knee instability. It is important to mention that for each skin lesions, areas of ecchymosis, or surgical scars. Quadriceps atrophy should be noted, and a tape measure should be used to record thigh circumference. It is good practice to measure the thigh circumference at the same distance from the patella or joint line in each knee. The presence of an effusion should be noted. This will be seen as fullness or swelling in the suprapatellar pouch. The effusion should be confirmed by ballottement of the patella (Figure 48-4). Small effusions will require “milking” of the fluid upward into the suprapatellar pouch. This will allow for quantification of the amount of fluid (Figure 48-5). The active and
translational and rotational motion of the knee, there are both primary and secondary restraints. When a primary restraint is disrupted, motion will be limited by the secondary restraint. If a secondary restraint is injured and the primary restraint remains intact, then motion will not be abnormal. For example, the ACL is the primary restraint to anterior translation of the tibia, while the medial meniscus is the secondary restraint. ACL disruption will lead to a significant increase in anterior tibial translation. This translation will be increased if the patient had a prior medial meniscectomy.

The collateral ligaments can be examined with stress applied in the coronal plane. They should be examined in full extension and in 30 degrees of flexion to remove the influence of the cruciate ligaments and the capsular restraints. With the patient in a supine position, a varus force is applied across the knee to test the lateral collateral ligament and a valgus force is applied across the knee to evaluate the medial collateral ligament.

The ACL is one of the most frequently injured structures in the knee. ACL insufficiency is also common in advanced osteoarthritis. Common mechanisms of injury include a direct blow to the lateral side of the knee (the “clipping” injury in football causing the triad of medial collateral ligament, ACL, and medial meniscus injuries), as well as noncontact injuries that occur during cutting, pivoting, and jumping. Patients often report an audible “pop” accompanied by the acute onset of knee swelling. Multiple tests have been described to evaluate the ACL. The most sensitive tests for diagnosis of an ACL injury include the anterior drawer, Lachman, and pivot-shift tests. All three tests are performed with the patient in the supine position. The anterior drawer test is performed with the knee flexed to 90 degrees. The examiner places his or her hands on the posterior surface of the proximal tibia and subluxates the tibia anteriorly. Any movement of the tibia that is different from the contralateral side is considered abnormal. The Lachman test is performed with the knee in 30 degrees of flexion to remove the contribution of secondary restraints. The examiner applies an anterior force on the tibia while stabilizing the femur with his or her contralateral hand. Any increase in anterior tibial translation relative to the contralateral side is considered abnormal. The pivot-shift test is performed with the knee in extension. The examiner holds the tibia in slight internal rotation and flexes the knee to 90 degrees. The tibia is then subluxated anteriorly, and the examiner looks for a second step or “shift” of the tibia on the femur.
rotation and applies a valgus stress while the knee is slowly flexed. This combination of forces should cause the tibia to subluxate anteriorly if the ACL is injured. The test is positive if the tibia reduces with a “clunk” or a “glide” at 20 to 40 degrees of flexion (Figure 48-9).

The posterior cruciate ligament (PCL) is the strongest ligament in the knee, and thus injuries to the PCL are usually a result of significant knee trauma. The “dashboard” injury is a common mechanism for PCL injury and occurs during a motor vehicle accident when the flexed knee strikes the dashboard (Figure 48-10). The PCL can be evaluated with the posterior drawer, posterior sag, and quadriceps active tests. All tests are performed with the patient in the supine position. The posterior drawer test is performed with the knee in 90 degrees of flexion. The examiner applies a posteriorly directed force to the tibia. Placement of one’s thumb tips at the anterior joint line will allow for quantification of any abnormal translation (Figure 48-11). The posterior sag test is positive when the tibia subluxates posteriorly with the knee at 90 degrees of flexion. Loss of the medial tibial step-off at the joint line should alert the examiner to a PCL injury (Figure 48-12). This test is usually positive in the chronic setting or under anesthesia in the acute setting. The quadriceps active test is performed with the knee in 60 degrees of flexion. The patient is asked to extend the knee while keeping his or her foot on the examination table. One will see reduction of the tibia in a positive test.

Figure 48-10  An injury to the posterior cruciate ligament can occur when the tibia strikes the dashboard, causing the tibia to subluxate posteriorly on the femur.

Figure 48-9  A and B, The pivot-shift test is positive if the tibia reduces with a “clunk” or a “glide” at 20 to 40 degrees of flexion.

Figure 48-11  The posterior drawer test is performed by subluxating the tibia posteriorly with the knee in 90 degrees of flexion. The amount of posterior translation (mm) is noted. The end point is characterized as "soft" or "hard."
Menisci

Meniscal tears usually occur with rotation of the flexed knee as it moves into extension. Tears of the medial meniscus are more common than tears of the lateral meniscus, likely due to the relative lack of mobility of the medial meniscus. Patients will frequently complain of “locking” and “clicking” or of something “wrong” with the knee, and this usually results from displacement of the torn meniscus during motion. Common physical findings include pain with hyperflexion and with hyperextension, joint line tenderness, and an effusion. Many provocative tests have been described to diagnose meniscal tears. The McMurray test and Apley compression tests are frequently performed, though they do lack sensitivity and specificity. The flexion McMurray test is performed with the patient supine and the hip and knee flexed to 90 degrees. A compressive and rotational force is applied to the knee as it is moved from a flexed to an extended position. The test is positive if the patient complains of pain (Figure 48-14). The Apley compression test is performed with the patient prone and the knee flexed to 90 degrees. In a positive test, the patient will complain of pain with rotation of the tibia. An arthroscopic
Prevalence of quadriceps tendon rupture after total knee arthroplasty is a rare (0.1%) but devastating complication. Patients usually present with intense anterior knee pain after experiencing an eccentric quadriceps contraction during a fall or twisting injury. Physical examination reveals a palpable defect in the tendon, an effusion due to hemarthrosis, and hypermobility of the patella. Patients will usually not be able to fully extend their knee (Figure 48-16).

Patella Tendon
Problems with the infrapatellar tendon include tendinitis and rupture. Tendinitis is usually an overuse injury and is often associated with jumping, changes in activity level, and eccentric contractions during falls. Patients will exhibit tenderness at their tibial tubercle or at the inferior pole of their patella. Rupture of the patella tendon usually occurs in patients younger than 40 years of age and is associated with chronic patella tendinitis. Patients usually present with anterior knee pain and the inability to extend their knee.

Patellofemoral Pain
Anterior knee pain is a common complaint seen by many orthopedic surgeons. It is more common in women, and it accounts for up to 25% of all sports-related knee injuries. A variety of factors contribute to the biomechanics of the patellofemoral joint and include overuse, the depth of the trochlea, the shape of the patella, quadriceps strength, the line of pull of the quadriceps relative to the patella tendon (the Q angle), the length of the patella tendon, the shape of the femoral condyles, and the articular cartilage. Abnormalities of any of these factors may contribute to this pain syndrome, and successful treatment is possible only with correct identification of any contributing factors.

Physical examination of the patellofemoral joint begins with an analysis of coronal alignment of the knee because any valgus deformity may contribute to lateral subluxation. The height of the patella relative to the tibial tubercle should be noted (patella alta or baja). The J sign is present when the patella slides laterally at terminal extension,
have been described should the interpretation phase begin. It is the natural tendency to bypass the description and proceed directly to interpretation. If this is done, it is likely that certain findings will be missed or dismissed prematurely.

The basic radiographic evaluation of the knee consists of standing anteroposterior (AP) weight-bearing, lateral, and Merchant’s views. The AP view allows for evaluation of coronal alignment and height of the tibiofemoral joint spaces. The normal coronal alignment of the knee should be 5 to 7 degrees of anatomic (tibiofemoral) valgus. The lateral tibiofemoral joint space should be wider than the medial tibiofemoral joint space in a normal knee. Any abnormality in mobility may stem from changes in the tightness of the retinaculum. The apprehension test is performed by attempting to subluxate the patella with the knee in extension. The test is positive when it elicits pain and an unwillingness to allow the examiner to move the patella laterally (Figure 48-17).

At the conclusion of the history and physical examination, the astute clinician should have formulated a short list of possible diagnoses. With this list in mind, the appropriate imaging studies can now be obtained. The goal of the initial imaging studies should be to confirm the diagnosis with the most appropriate and least expensive study. Advanced imaging studies should not replace a thorough history and physical examination.

**Imaging**

**Conventional Radiographs**

Conventional roentgenograms are usually the first study obtained after knee injury and should be read in a systematic fashion. Soft tissues should be evaluated before examining the bony structures. Findings should be described in terms of radiolucent and radiopaque lines. Only after the findings have been described should the interpretation phase begin. It is the natural tendency to bypass the description and proceed directly to interpretation. If this is done, it is likely that certain findings will be missed or dismissed prematurely.

The basic radiographic evaluation of the knee consists of standing anteroposterior (AP) weight-bearing, lateral, and Merchant’s views. The AP view allows for evaluation of coronal alignment and height of the tibiofemoral joint spaces. The normal coronal alignment of the knee should be 5 to 7 degrees of anatomic (tibiofemoral) valgus. The lateral tibiofemoral joint space should be wider than the medial tibiofemoral joint space in a normal knee. The presence of marginal osteophytes, joint space narrowing, subchondral sclerosis, and cystic change will be seen in the presence of osteoarthritis (Figure 48-18). Periarticular

![Figure 48-17](image) The apprehension test is positive when subluxation of the patella causes pain.

![Figure 48-18](image) Standing anteroposterior (A), lateral (B), and Merchant’s (C) views of an osteoarthritic knee.
Computed Tomography

Computed tomography (CT) has largely been replaced by magnetic resonance imaging (MRI) in evaluation of routine knee problems. CT is now used primarily for detection of bony tumors and in the trauma setting for detection of subtle fractures that are not easily visualized with conventional radiographs, as well as for a more thorough evaluation of intra-articular fractures. In cases of distal femoral or proximal tibia fractures, CT is used to help the surgeon plan operative treatment. CT is also used to assess axial alignment of the femoral and tibial components in cases of the painful total knee arthroplasty.\(^\text{33,34}\)

Ultrasound

The use of ultrasound has become more common in the diagnosis of knee disorders due to recent improvements in transducer technology. Ultrasound is an attractive imaging modality because of its low cost, real-time capabilities, and portability. The ability to perform provocative maneuvers during sonography is particularly appealing. Ultrasound can easily and reliably detect joint effusions, as well as quadriceps and patella tendon disruptions. It has been reported that ultrasound can detect a 1-mm increase in joint fluid.\(^\text{35}\)

Nuclear Scintigraphy

Nuclear scintigraphy is sensitive but not specific, and it is used to detect areas of increased osseous remodeling. It requires clinical correlation and should be used in conjunction with other imaging modalities. Technetium phosphate compounds are injected intravenously. Approximately 50% of the tracer is excreted by the kidneys, and the remainder is taken up in areas of increased osseous turnover. Imaging of the skeleton is typically performed 2 to 3 hours after injection because this allows for maximum contrast between the soft tissues and the skeletal structures while still providing for an adequate photon count.\(^\text{36}\) Three-phase bone scanning can yield additional information. The three phases include an angiographic pool, followed by blood pool and bone imaging. The angiographic phases allow for detection of regional hyperemia. This technique has been reported to have greater specificity and can be used in cases of suspected osteomyelitis, osteonecrosis, stress fracture, and implant loosening.\(^\text{36}\) It has been reported that increased radionuclide uptake can be seen for up to 12 to 18 months after total knee arthroplasty. Asymmetric uptake in one area around the prosthesis should raise the question of loosening or periprosthetic fracture (Figure 48-20).\(^\text{37,38}\) Addition of labeled leukocytes to the technetium 99m sulfur colloid yields an 80% sensitivity and 100% specificity for diagnosing infection.\(^\text{39}\)

Magnetic Resonance Imaging

MRI has supplanted many imaging modalities due to its direct multiplanar capabilities and superior soft tissue contrast. Although conventional radiographs remain the gold standard for defining osseous structures, MRI provides excellent visualization of articular cartilage, the cruciate ligaments, the collateral ligaments, the patella tendon, the osteopenia, concentric joint space narrowing, and a paucity of osteophytes are commonly seen in inflammatory arthritis (Figure 48-19). The lateral radiograph allows for evaluation of an effusion, patella tendon length, and the quadriceps tendon. The Merchant’s view is taken tangential to the patellofemoral joint.\(^\text{31}\) It allows for detection of patellofemoral arthritis and malalignment.

Additional views include a posteroanterior (PA) standing view with the knees flexed approximately 45 degrees, the tunnel or intercondylar notch view, and the 36-inch AP standing view of bilateral lower extremities. The flexed PA standing view is taken with the radiographic beam directed 10 degrees caudad from anterior to posterior. This allows for evaluation of the posterior femoral condyles for joint space narrowing.\(^\text{32}\) The tunnel view is obtained with the knee flexed and the radiographic beam directed inferiorly at an angle perpendicular to the tibial plateau. It is useful in detecting posterior tibiofemoral joint space narrowing, tibial spine fractures, loose bodies, and osteochondral lesions on the medial aspect of the femoral condyles. The 36-inch standing view is used for determining the mechanical axis of the lower extremity and evaluating any deformity that may be present. The normal mechanical axis is a straight line joining the center of the hip, knee, and ankle joints. Surgeons use it for preoperative planning and postoperative evaluation in total knee arthroplasty, as well for the planning of distal femoral and proximal tibia osteotomies in arthritis surgery.

Figure 48-19 Standing AP (A), lateral (B), and Merchant’s (C) views of the knee in a patient with rheumatoid arthritis.
Common Disorders in the Differential Diagnosis of Knee Pain

**General**

Though many diseases may involve the knee, a limited number are encountered frequently. In evaluating the complaint of knee pain, the clinician should be familiar with osteoarthritis; rheumatoid arthritis; inflammatory arthritis associated with the seronegative spondyloarthropathies; tears of the menisci, ligaments, and tendons; osteochondritis dissecans; osteochondral fractures; fractures; referred pain from the hip (such as with slipped capital femoral epiphysis in adolescents); vascular claudication; neurogenic claudication; complex regional pain syndrome; sarcoma; metastases; and infection.

**Bursitis**

The prepatellar bursa lies between the retinaculum and the subcutaneous fat and runs from the patella to the tibial tubercle. The bursa may become inflamed and fill with fluid when exposed to a direct blow or repetitive microtrauma (kneeling). Patients with prepatellar bursitis present with anterior knee pain on flexion and a fluctuant mass over the anterior knee. If the area becomes warm, tender to palpation, and erythematous, septic bursitis should be ruled out with aspiration. The pes anserinus bursa, located over the insertions of the sartorius, gracilis, and semitendinosus muscles on the proximal medial tibia, can also be a source of knee pain if inflamed.

**Neoplasia**

Tumors around the knee are often diagnosed after trauma prompts medical evaluation. Pain at night, pain at rest, and constitutional symptoms should alert the clinician...
to consider the appropriate workup. Some of the benign tumors seen around the knee include enchondroma, pigmented villonodular synovitis, osteochondromatosis, and giant cell tumor. Malignant tumors seen around the knee include, but are not limited to, metastases, osteosarcoma, Ewing's sarcoma, chondrosarcoma, and malignant fibrous histiocytoma.

**Popliteal Cysts**

A popliteal cyst, originally called Baker's cyst, is a synovial fluid-filled mass located in the popliteal fossa. The most common synovial popliteal cyst is considered to be a distention of the bursa located beneath the medial head of the gastrocnemius muscle. Usually, in an adult patient, an underlying intra-articular disorder (osteoarthritis) is present. In children, the cyst can be isolated and the knee joint normal. Patients usually present with episodic posterior knee pain. The diagnosis is made by ultrasonography or MRI. Treatment options include benign neglect, aspiration, surgical excision, or removal of the underlying pathology (arthritis) with knee arthroplasty.

**HIP PAIN**

**History**

Taking an accurate history is an important initial step in formulating a differential diagnosis for patients who present with a complaint of hip pain. In general, more conditions should be considered in the differential diagnosis for hip pain than for knee pain because the hip is a common site for referred pain from lumbosacral and intrapelvic pathology. A detailed, comprehensive history will direct the clinician to a focused physical examination.

Most patients who present with hip pathology will complain of pain. It is important to define the exact location of the pain because “hip” pain may refer to discomfort in the groin, lateral thigh, or buttock. Pain in the groin or medial thigh region is most often due to hip disease and is believed to arise from irritation of the capsule and/or synovial lining. Pain generated in the lumbosacral spine may be referred to the buttocks and/or lateral thigh. Lateral thigh pain may stem from so-called trochanteric bursitis (usually abductor tendinitis) as well. Activities or positions that aggravate and relieve the pain should be explored. The severity, frequency, and patterns of radiation of the pain should also be evaluated. It is not uncommon for knee pain to be generated from the hip joint. Metastatic and primary tumors that occur in the pelvic and proximal thigh regions should always be included in the differential diagnosis. Intrapelvic pathology from the prostate, seminal vesicles, hernias, ovaries, gastrointestinal (GI) system, and vasculature should also be considered.

Knowledge of the patient’s general level of functioning is important because this will lend insight into the severity of disease and may influence treatment. Patients with hip pathology may have difficulty trimming their toenails, donning shoes and socks, and using stairs. Walking tolerance and use of assist devices should also be recorded. The Harris Hip Score and WOMAC Osteoarthritis Index are two rating scales that are widely used to assess function in this patient population.

The patient should be asked about any hip problems that he or she encountered in childhood. Diseases such as developmental dysplasia, slipped capital femoral epiphysis, Legg-Calvé-Perthes disease, polio, and trauma may lead to osteoarthritis later in life. Any treatment rendered for these diseases should be asked about as well.

Osteoarthritis and inflammatory arthritis are two common causes of hip pain. In general, pain from osteoarthritis will be exacerbated by activity and relieved by rest. Mild arthritis of the hip may not become symptomatic until a certain activity level is reached. Stiffness (usually from synovitis) is also a common complaint with both degenerative and inflammatory arthritis. When the hip pain continues despite a trial of rest, an underlying inflammatory or infectious process should be considered. The American Rheumatism Association revised their classification of rheumatoid arthritis in 1988. The current criteria include 1 hour of morning stiffness for 6 weeks, symmetric joint swelling in at least three joints, subcutaneous nodules, typical radiographic changes, and a positive rheumatoid factor.

Any previous treatments for hip pain should be discussed. The patient's response to nonsteroidal anti-inflammatory medications, nutritional supplements (e.g., chondroitin and glucosamine), physical therapy, corticosteroid injections, local anesthetic injections, hyaluronic acid injections, ultrasound, and operative interventions should be recorded. Lastly, a more general medical history should be explored. The physician should be aware of alcoholism, neuromuscular disorders, smoking history, and general support systems.

**Physical Examination**

The physical examination of the patient with hip pain begins as the clinician watches the patient for the first time. Ease of chair rise, postures, and walking speed all provide insight into the extent of a patient's disability. A general evaluation of the patient's spine, lower extremity alignment, and leg lengths comes next. With the examiner behind the patient, the spine is examined for coronal and sagittal balance. The patient is asked to touch his or her toes. A rib hump indicates the presence of scoliosis. Any gross deformity of the spine will alert the examiner to the potential of a pelvic obliquity and resultant leg-length discrepancy. The overall coronal alignment of the lower extremities is evaluated next. If a leg-length discrepancy is detected, blocks can be used, as discussed previously, to determine the amount of apparent inequality. If the leg-length discrepancy is due to a fixed pelvic obliquity from lumbosacral disease, blocks may not be able to level the pelvis. Previous surgical scars about the hip are noted. Palpation of the bony landmarks (iliac crest, anterior superior iliac spine, posterior superior iliac spine, ischial tuberosity, coccyx, spinous processes, and greater trochanter) should be performed (Figure 48-22). The femoral neck is located approximately three fingerbreadths below the anterior superior iliac spine.

A basic evaluation of gait should be performed. Though gait analysis is a complex science, all clinicians should feel comfortable evaluating for common abnormalities. The patient with hip pain may present with an antalgic gait. The severity of the limp should be classified as mild, moderate, or severe. Mild limps can only be detected by trained
occurs when the pelvis tilts toward the unsupported side during one-legged stance. This test is best performed with the examiner behind the patient. Causes of abductor weakness are numerous and may include a contracted or shortened gluteus medius, coxa vara, fracture, dysplasia, neurologic conditions (e.g., superior gluteal nerve injury, radiculopathy, poliomyelitis, myelomeningocele, spinal cord lesions), and slipped capital femoral epiphysis.

The patient is then asked to lay supine on the examination table. The range of motion of both hips should be evaluated by recording flexion, extension, adduction, abduction, internal rotation in extension, and external rotation in extension. Hip extension is best evaluated with the patient in the prone position. Normal range of motion values include 100 to 135 degrees for flexion (knee should be flexed to relax the hamstrings), 15 to 30 degrees for extension, 0 to 30 degrees for adduction, 0 to 40 degrees for abduction, 0 to 40 degrees for internal rotation, and 0 to 60 degrees for external rotation. Motion is often limited in cases of deformity (such as limited internal rotation in slipped capital femoral epiphysis) and advanced osteoarthritis. Internal rotation and abduction are usually the first motions to be limited in osteoarthritis. Motion will be painful in patients with synovitis as well. Areas that are painful should be palpated.

A series of special tests can be performed to evaluate for subtle muscle contractures and limitation of motion. The presence of a hip flexion contracture is common in patients with moderate to severe hip pathology and can be quantified with the Thomas test (Figure 48-24). This test is performed by having the patient bring his or her thighs to their chest while in the supine position. This allows for flattening of the spine, and the hip to be evaluated is allowed to extend to neutral. If the patient is unable to reach neutral, the amount of flexion contracture is recorded. The Ober test measures tightness of the iliotibial band. The patient lies on the unaffected side and the examiner helps the patient abduct the hip with the hip extended and the knee

Figure 48-22  Diagram of the bony landmarks on the pelvis that can be palpated during physical examination.

observers. Moderate limps will be noticed by the patient. A severe limp will be readily apparent and have a significant impact on speed of ambulation. Common causes of limp include pain and abductor (gluteus medius and gluteus minimus) weakness. Differentiating between these two etiologies of limp is an important part of the physical examination.

The patient with abductor dysfunction will likely have an abductor, or Trendelenburg lurch. With a Trendelenburg lurch, the patient compensates for abductor dysfunction by leaning over the involved hip to shift the body’s center of gravity in that direction (Figure 48-23). If the patient has a Trendelenburg lurch, we proceed to evaluate for a Trendelenburg sign. A positive Trendelenburg sign occurs when the pelvis tilts toward the unsupported side during one-legged stance. This test is best performed with the examiner behind the patient. Causes of abductor weakness are numerous and may include a contracted or shortened gluteus medius, coxa vara, fracture, dysplasia, neurologic conditions (e.g., superior gluteal nerve injury, radiculopathy, poliomyelitis, myelomeningocele, spinal cord lesions), and slipped capital femoral epiphysis.

The patient is then asked to lay supine on the examination table. The range of motion of both hips should be evaluated by recording flexion, extension, adduction, abduction, internal rotation in extension, and external rotation in extension. Hip extension is best evaluated with the patient in the prone position. Normal range of motion values include 100 to 135 degrees for flexion (knee should be flexed to relax the hamstrings), 15 to 30 degrees for extension, 0 to 30 degrees for adduction, 0 to 40 degrees for abduction, 0 to 40 degrees for internal rotation, and 0 to 60 degrees for external rotation. Motion is often limited in cases of deformity (such as limited internal rotation in slipped capital femoral epiphysis) and advanced osteoarthritis. Internal rotation and abduction are usually the first motions to be limited in osteoarthritis. Motion will be painful in patients with synovitis as well. Areas that are painful should be palpated.

A series of special tests can be performed to evaluate for subtle muscle contractures and limitation of motion. The presence of a hip flexion contracture is common in patients with moderate to severe hip pathology and can be quantified with the Thomas test (Figure 48-24). This test is performed by having the patient bring his or her thighs to their chest while in the supine position. This allows for flattening of the spine, and the hip to be evaluated is allowed to extend to neutral. If the patient is unable to reach neutral, the amount of flexion contracture is recorded. The Ober test measures tightness of the iliotibial band. The patient lies on the unaffected side and the examiner helps the patient abduct the hip with the hip extended and the knee

Figure 48-22  Diagram of the bony landmarks on the pelvis that can be palpated during physical examination.

observers. Moderate limps will be noticed by the patient. A severe limp will be readily apparent and have a significant impact on speed of ambulation. Common causes of limp include pain and abductor (gluteus medius and gluteus minimus) weakness. Differentiating between these two etiologies of limp is an important part of the physical examination.

The patient with abductor dysfunction will likely have an abductor, or Trendelenburg lurch. With a Trendelenburg lurch, the patient compensates for abductor dysfunction by leaning over the involved hip to shift the body’s center of gravity in that direction (Figure 48-23). If the patient has a Trendelenburg lurch, we proceed to evaluate for a Trendelenburg sign. A positive Trendelenburg sign occurs when the pelvis tilts toward the unsupported side during one-legged stance. This test is best performed with the examiner behind the patient. Causes of abductor weakness are numerous and may include a contracted or shortened gluteus medius, coxa vara, fracture, dysplasia, neurologic conditions (e.g., superior gluteal nerve injury, radiculopathy, poliomyelitis, myelomeningocele, spinal cord lesions), and slipped capital femoral epiphysis.

The patient is then asked to lay supine on the examination table. The range of motion of both hips should be evaluated by recording flexion, extension, adduction, abduction, internal rotation in extension, and external rotation in extension. Hip extension is best evaluated with the patient in the prone position. Normal range of motion values include 100 to 135 degrees for flexion (knee should be flexed to relax the hamstrings), 15 to 30 degrees for extension, 0 to 30 degrees for adduction, 0 to 40 degrees for abduction, 0 to 40 degrees for internal rotation, and 0 to 60 degrees for external rotation. Motion is often limited in cases of deformity (such as limited internal rotation in slipped capital femoral epiphysis) and advanced osteoarthritis. Internal rotation and abduction are usually the first motions to be limited in osteoarthritis. Motion will be painful in patients with synovitis as well. Areas that are painful should be palpated.

A series of special tests can be performed to evaluate for subtle muscle contractures and limitation of motion. The presence of a hip flexion contracture is common in patients with moderate to severe hip pathology and can be quantified with the Thomas test (Figure 48-24). This test is performed by having the patient bring his or her thighs to their chest while in the supine position. This allows for flattening of the spine, and the hip to be evaluated is allowed to extend to neutral. If the patient is unable to reach neutral, the amount of flexion contracture is recorded. The Ober test measures tightness of the iliotibial band. The patient lies on the unaffected side and the examiner helps the patient abduct the hip with the hip extended and the knee

Figure 48-22  Diagram of the bony landmarks on the pelvis that can be palpated during physical examination.

Figure 48-23  Physical examination of abductor function: A, Normal single-legged stance. B, Positive Trendelenburg lurch and negative Trendelenburg sign. C, Positive Trendelenburg lurch with pelvic obliquity and leaning over the involved hip to shift the body’s center of gravity.

Figure 48-24  In the Thomas test, a hip flexion contracture is measured by flexing the contralateral hip to eliminate compensatory lumbar lordosis. The ipsilateral hip is then allowed to extend with gravity. The angle between the examination table and the thigh is the degree of flexion contracture.
flexed to 90 degrees. The leg is slowly released from abduction to neutral, and the hip will remain abducted if there is contracture of the iliotibial band. Ely’s test will detect a tight rectus femoris. The knee is passively flexed with the patient in the prone position. If the rectus femoris is tight, the ipsilateral hip will spontaneously flex. If the rectus femoris is normal, the hip will remain flush with the examination table.

Patients will occasionally complain of a “snapping” sensation in their hip. Although it may be difficult for the clinician to reproduce snapping, patients may be able to demonstrate this by flexing and internally rotating their hip. Extra-articular causes of hip snapping include a thickened iliotibial band snapping over the greater trochanter, the iliopsoas tendon gliding over the ilipectineal eminence, the long head of the biceps tendon rubbing on the ischial tuberosity, and the iliofemoral ligament rubbing on the femoral head. Intra-articular causes of snapping hip syndrome include loose bodies and large labral tears.

In addition to using blocks with the patient standing, leg lengths can be measured while the patient is in the supine position (Figure 48-25). The apparent leg length is the distance from the umbilicus to the medial malleolus. The true leg length is measured from the anterior superior iliac spine to the medial malleolus. Pelvic obliquity and abduction/adduction of the hip will create an apparent leg-length discrepancy.

Sacroiliac disease should be included in the differential diagnosis of hip pain. Although multiple provocative tests have been described to elicit sacroiliac disease, the flexion in abduction and external rotation (FABER) test (also known as Patrick’s test) can help distinguish between hip and sacroiliac joint pathology. With the patient supine, the clinician has the patient place his or her hip in the flexion, abduction, and externally rotated position. The clinician then presses the flexed knee and the contralateral anterior superior iliac spine toward the floor. Pain in the buttocks suggests sacroiliac joint disease, whereas pain in the groin points to hip pathology. If the sacroiliac joint is implicated, it is recommended that multiple other provocative tests be performed. It has been shown that by using a combination of the distraction, thigh thrust, compression, sacral thrust, Gaenslen’s, and FABER tests, sacroiliac joint pathology is the likely pain generator when three or more of the tests are positive.

The acetabular labrum is drawing attention as a previously underappreciated cause of hip pain. Clinical presentation of a labral tear of the acetabulum may be variable, and the diagnosis is often delayed. Patients usually see multiple providers before the diagnosis is confirmed. In a series of 66 patients with arthroscopically confirmed tears of the acetabular labrum, 92% of the patients complained of groin pain, 91% of the patients had activity-related pain, 71% of the patients complained of night pain, 86% of the patients described the pain as moderate to severe, and 95% of the patients had a positive impingement sign. The authors recommended that a diagnosis of acetabular labral tear be suspected in young, active patients complaining of groin pain with or without trauma. The positive impingement test helps confirm the diagnosis of labral tear. The test is positive if the patient experiences groin pain with the hip flexed, adducted, and internally rotated. The positive predictive value of this test has been shown to range from 0.91 to 1.00 in six different studies.

A thorough evaluation of the neurovascular system should be completed after the musculoskeletal portion of the physical examination for the hip or knee is completed. This should include palpation or Doppler evaluation of the femoral, popliteal, dorsalis pedis, and posterior tibial arteries, as indicated. Strength testing with resisted isometric movements for each muscle in the lower extremity is performed, with 5 being normal strength, 4 being full motion against gravity and against some resistance, 3 being fair motion against gravity, 2 being movement only with gravity eliminated, 1 being evidence of muscle contraction but no joint motion, and 0 being no evidence of contractility.

Figure 48-25 Measurement of leg lengths: A, The apparent leg length is the distance from the umbilicus to the medial malleolus. B, Pelvic obliquity causing an apparent leg-length discrepancy. C, The true leg length is the distance from the anterior superior iliac spine to the medial malleolus.
Sensation in the lower extremity should be evaluated by assessing for light touch and/or appreciation of pin prick in a dermatomal distribution. Patellar and ankle reflexes should be tested. Lastly, the examiner should test for any abnormal clonus and Babinski reflexes as indicated.

Imaging

Conventional Radiographs

Plain radiographs remain the primary diagnostic imaging tool for the evaluation of hip pathology. All other imaging modalities should be viewed as complementary to conventional radiographs. Our standard screening series includes a low anteroposterior (AP) pelvis (Figure 48-26), an AP hip (Figure 48-27), a frog-lateral view, and a cross-table lateral view. The frog-lateral view provides a lateral of the proximal femur and is useful for detecting femoral head collapse (as seen in osteonecrosis, Figure 48-28). Numerous other special radiographs of the hip exist including Judet 45-degree oblique views and the false profile view. Judet views allow for easier visualization of the anterior (obturator oblique) and posterior (iliac oblique) columns. The false profile view allows for evaluation of anterior bony coverage of the femoral head in cases of acetabular dysplasia. Developmental dysplasia of the hip (DDH) is common, and we do not recommend the routine use of any special views before referral to an orthopedic surgeon (Figure 48-29).

Computed Tomography

Computed tomography (CT) is used for assessment of acetabular fractures, acetabular nonunions, femoral head fractures, subtle femoral neck fractures, neoplasia, and bone stock in the revision total hip arthroplasty setting. Due to its limited soft tissue contrast, CT has largely been replaced by MRI for detailed evaluation of the soft tissues around the hip.

Nuclear Scintigraphy

The role of bone scanning in the evaluation of hip pathology is similar to its role in the assessment of knee pain. It should always be used in conjunction with other imaging modalities due to its limited specificity (Figure 48-30).

Magnetic Resonance Imaging

MRI provides unprecedented detail of the soft tissues around the hip joint. Its use is now common for diagnosis of osteonecrosis, labral pathology, neoplasia, effusion, synovitis, loose bodies, tendinitis, transient osteoporosis of the hip,
occult femoral neck fractures, bone edema, gluteus medius tendon avulsions, and nerve injury. MR arthrography of the hip joint is useful for identifying gluteus medius tendon avulsion after total hip arthroplasty (Figure 48-31) and for detecting labral tears. One study showed a 92% sensitivity for the detection of labral tears using MR arthrography.\textsuperscript{63} Delayed gadolinium-enhanced MRI of cartilage, a technique designed to measure early arthritis in the hip joint, is now being used clinically in the management of hip dysplasia.\textsuperscript{64} Despite the tremendous diagnostic capabilities of MRI, its ability to detect bony pathology is limited. As such, conventional radiographs remain the imaging modality of choice for the screening of hip pathology.
**Hip Arthrography**

Hip arthrography is useful for detecting avulsions of the gluteus medius tendon from the greater trochanter and for differentiating intra-articular hip pathology from lumbosacral disease. In one study, intra-articular anesthetic injection was 90% accurate in predicting intra-articular pathology as confirmed by hip arthroscopy. Anesthetic arthrogram of the hip has shown a 95% positive predictive value and a 67% negative predictive value for pain relief after total hip arthroplasty in patients with concurrent hip and lumbar osteoarthritides.

**Common Disorders in the Differential Diagnosis of Hip Pain**

Numerous common causes of hip pain exist, and a detailed discussion of these is beyond the scope of this chapter. The differential diagnosis of hip pain should include osteoarthritis (most frequently from developmental dysplasia, Legg-Calvé-Perthes disease, or slipped capital femoral epiphysis); inflammatory arthritis; osteonecrosis; fractures (acetabulum, femoral head, femoral neck, intertrochanteric, or subtrochanteric); trochanteric bursitis; femoroacetabular impingement; tears of the acetabular labrum; transient osteoporosis of the proximal femur; infection; snapping hip syndrome; osteitis pubis; neoplasia (osteosarcoma, chondrosarcoma, pigmented villonodular synovitis, osteochondromatosis, malignant fibrous histiocytoma, or metastases); inguinal hernia; or referred pain (lumbosacral spine, lower GI tract). This list can be efficiently narrowed down by taking a detailed history, performing a comprehensive examination of the musculoskeletal and nervous systems, and obtaining the appropriate imaging studies.

**References**

Foot and Ankle Pain
MARK D. PRICE • CHRISTOPHER P. CHIODO

The differential diagnosis for foot and ankle pain is vast. Localizing symptoms by anatomic region helps narrow this differential.

On physical examination, most structures in the foot and ankle are immediately subcutaneous and readily palpable.

Beyond medications, useful nonoperative treatments include bracing, shoewear modification, orthoses, and physical therapy.

Most surgical procedures in foot and ankle surgery fall into one of the following categories: arthrodesis, arthroplasty, corrective osteotomy, ostectomy, tendon débridement and transfer, and synovectomy. Patient compliance and soft tissue integrity are important factors when considering surgery.

Advances in medical management now make joint-sparing procedures possible in many patients with inflammatory arthritis who previously would have required arthrodesis.

Foot and ankle pain are independent risk factors for locomotor instability, impaired balance, and increased risk for falling, as well as compromised functional activities of daily living.1-5 Foot and ankle pain appear to affect approximately one in five middle-aged to older individuals. Interference with daily activities occurs in one-half to one-third of affected individuals but is rarely disabling outside the context of rheumatoid arthritis. Foot and ankle pain is significantly more common in women, a finding that has been attributed to gender-specific footwear.

CAUSES OF FOOT AND ANKLE PAIN

The differential diagnosis of foot and ankle pain is vast and includes conditions of tendons, ligaments, muscle, bone, joints, periarticular structures, nerves, and vessels, as well as referred pain (Table 49-1). The most common cause of pain of the foot and ankle is osteoarthritis (OA). Although OA is the most prevalent joint disease, its pathophysiology remains poorly understood. Research regarding foot and ankle OA in particular is limited by absence of a standard case definition. Ankle and foot OA results from damage and loss of the articular cartilage, which can cause inflammation, stiffness, pain, swelling, deformity, and limitation of function, such as walking or standing. Osteophyte formation can lead to impingement and further pain. In the foot, OA most commonly occurs in the big toe, the midfoot, and ankle. In the early stages, pain may occur only at the beginning and at the end of an activity, but as the condition progresses it can become constant, even at rest.

The ankle is a complex joint that is subjected to enormous forces during daily activities and in sports, especially running. It is also the joint most commonly injured in the human body. This combination of factors predisposes the ankle joint to degenerative changes, although the risk is lower than other weight-bearing joints, such as the hip and knee. The ankle also rarely develops arthritic changes without an identifiable cause. The most common cause of ankle OA is trauma and can develop following a fracture or repeated sprains. Other causes of OA are abnormal foot mechanics (flat and high-arched feet) and, rarely, systemic diseases such as hemochromatosis.

Foot and ankle pain is the presenting complaint in approximately 15% to 20% of newly diagnosed rheumatoid arthritis (RA) patients. Further, of those patients already diagnosed with RA, the prevalence of foot and ankle involvement has been estimated to be greater than 90%. Evaluation of the rheumatoid foot and ankle begins with a thorough history and physical examination. The location, timing, and duration of symptoms can help establish a specific diagnosis and help guide the subsequent course of treatment. Radiographs and advanced imaging modalities provide useful adjuncts in the evaluation of specific foot and ankle pathologies.

The treatment of the rheumatoid foot and ankle is aimed at both alleviating pain and preserving function (i.e., maintaining the ambulatory status of the patient). Initial nonoperative treatment includes medical management, physical therapy, shoewear modification, orthotics, and bracing. These measures provide substantial relief for many. For recalcitrant symptoms, surgical intervention may be necessary. Most surgical procedures fall into one of the following general categories: arthrodesis (joint fusion), arthroplasty (joint replacement), corrective osteotomy, ostectomy, and synovectomy (joint or tendon).

FUNCTIONAL ANATOMY AND BIOMECHANICS

The ankle, or tibiotalar joint, is composed of the articulation between the foot (talus) and the lower leg (distal tibia and fibula). Its primary motion is plantar flexion and dorsiflexion in the sagittal plane. In addition, the articulation between the distal tibia and fibula allows a lesser amount of internal and external rotation to occur in the axial, or transverse, plane.

The foot may be loosely divided into three anatomic regions: forefoot, midfoot, and hindfoot. The forefoot...
Table 49-1  Differential Diagnosis of Foot and Ankle Pain

<table>
<thead>
<tr>
<th>Tendon, Ligament, and Muscle</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastoc-soleus strain</td>
<td>Anterior talofibular ligament tear</td>
</tr>
<tr>
<td>Plantaris rupture</td>
<td>Calcaneofibular ligament tear</td>
</tr>
<tr>
<td>Deltoid ligament tear</td>
<td>Anterolateral impingement due to complete tear of anterior talofibular ligament and anterior inferior tibiofibular ligament</td>
</tr>
<tr>
<td>Syneosmatic impingement due to tear of synesosmotic ligament</td>
<td>Sinus tarsi syndrome (lateral hindfoot pain and instability due to injury of contents of the sinus and tarsal tunnel)</td>
</tr>
<tr>
<td>Achilles tendinitis</td>
<td>Achilles rupture</td>
</tr>
<tr>
<td>Plantar fasciitis</td>
<td>Posterior tibial tendon dysfunction</td>
</tr>
<tr>
<td>Flexor hallucis longus dysfunction</td>
<td>Tibialis anterior tendon tear</td>
</tr>
<tr>
<td>Peroneus brevis tendon tear</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bone</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fracture of talus</td>
<td>Lisfranc fracture-dislocation (fracture of the first metatarsal base with dislocation of medial cuneiform)</td>
</tr>
<tr>
<td>Calcaneal fracture</td>
<td>Metatarsal stress fracture</td>
</tr>
<tr>
<td>Navicular fractures</td>
<td>Freiberg's infraction (sclerosis and flattening of the second metatarsal head due to trauma or microtrauma)</td>
</tr>
<tr>
<td>Avascular necrosis of the talus</td>
<td>Avascular necrosis of the talus</td>
</tr>
<tr>
<td>Fracture of the phalanx</td>
<td>Fracture of the sesamoids</td>
</tr>
<tr>
<td>Fracture of the sesamoids</td>
<td>Sesamoiditis</td>
</tr>
<tr>
<td>Metatarsalgia</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Joint</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Osteoarthritis</td>
<td>Gout</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Other inflammatory arthritides</td>
</tr>
<tr>
<td>Charcot's joint</td>
<td>Osteochondral lesion of the talus</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Periarticular Structures</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Shin splint (periostal avulsion and periostitis at the insertion of the medial soleus due to repetitive overuse, such as in running and hiking)</td>
<td>Hallux rigidus</td>
</tr>
<tr>
<td>Hallux valgus</td>
<td>Ingrown toenail</td>
</tr>
<tr>
<td>Toe deformities</td>
<td>Turf toe (sprain of the first metatarsophalangeal joint due to hyperextension forces)</td>
</tr>
<tr>
<td>Plantar fasciitis</td>
<td>Plantar fibromatosis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nerves</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior tarsal tunnel syndrome (involvement of deep peroneal nerve under the superficial fascia of the ankle)</td>
<td>Morton's neuroma</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vessels</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Atherosclerosis</td>
<td>Compartment syndrome</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Referred Pain</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex regional pain syndrome</td>
<td></td>
</tr>
</tbody>
</table>

Consists of the toes and metatarsal bones, along with the metatarsophalangeal (MTP) and interphalangeal (IP) joints. The tarsometatarsal (TMT) joints connect the forefoot to the midfoot, which comprises the three cuneiform bones, the navicular, and the cuboid. Finally, the hindfoot, located below the ankle, consists of the talus and calcaneus. The joints of the hindfoot include the talocalcaneal (sustalar), talonavicular, and calcaneocuboid articulations.

Forefoot and midfoot motion is primarily plantarflexion and dorsiflexion in the sagittal plane, with some secondary pronation and supination in the coronal plane and abduction/adduction in the axial plane. Motion in the hindfoot is primarily composed of inversion/eversion in the coronal plane, with secondary internal/external rotation in the axial plane and plantarflexion/dorsiflexion in the sagittal plane.

Knowledge of these anatomic divisions is important because radiographs often demonstrate polyarticular disease in patients with RA. An intimate understanding of the local anatomy greatly aids in the establishment of an accurate diagnosis and formulation of an appropriate treatment plan.

**DIAGNOSTIC EVALUATION**

**Physical Examination**

A thorough physical examination of the foot and ankle begins with gait analysis, even if simply observing the patient enter the examination room. Normal human gait is divided into two phases. The stance phase is the weight-bearing portion of the gait cycle and comprises roughly 60% of normal walking. It begins with heel-strike and then extends through foot-flat to toe-off. Meanwhile, the swing phase of gait extends from toe-off to heel-strike and comprises the remaining 40% of the gait cycle.

Patients with an “antalgic” gait pattern will have a shortened stance phase on the side of the affected limb, as they attempt to more quickly transfer their weight to the non-painful limb. In addition to an antalgic gait, foot and ankle pain often results in the avoidance of ground contact with the painful part of the foot. A further problem noted in stance phase is dynamic collapse of the medial longitudinal arch, most apparent at foot-flat and toe-off.

During the swing phase of gait, a “steppage” gait may be noted. This is characterized by excessive hip and knee flexion to allow a patient’s foot to clear the ground in the setting of a footdrop. In patients with RA, it may be caused by attritional rupture of the anterior tibialis tendon, which is the main dorsiflexor of the ankle.

Following gait analysis, the foot and ankle are inspected, both with the patient sitting and standing. The location of swelling is usually well correlated with the joint(s) involved (e.g., ankle vs. talocalcaneal joint). Deformity should also be noted. Commonly seen deformities in patients with RA include hallux valgus, or bunion (Figure 49-1); hammertoes; and flatfoot deformity (characterized by hindfoot valgus/foofoot abduction). Callosities develop over regions of increased pressure and are associated with deformity and fat pad atrophy. Rheumatoid nodules can appear anywhere on the foot but are often found in areas of repetitive trauma (i.e., at the site of irritation from a tight shoe counter). Similarly, ulcerations appear in areas of repeated injury such...
including arthritis, synovitis, impingement, and osteochondral defect (OCD).

A more detailed description of these conditions and their correlation with anatomic location is provided later in the chapter and in Table 49-2.

Following inspection and palpation, range-of-motion analysis is performed. Passive range of motion of the ankle is normally between 10 and 20 degrees of dorsiflexion and 40 and 50 degrees of plantarflexion. Normal hindfoot inversion and eversion are approximately 20 and 10 degrees, respectively. The first MTP joint should have approximately 45 degrees of “plantarflexion” (flexion) and 70 to 90 degrees of “dorsiflexion” (extension). Deviations from these norms should be noted as part of the standard workup.

**Imaging**

Despite the abundant availability of advanced imaging modalities such as magnetic resonance imaging (MRI) and computed tomography (CT), radiographs remain the imaging mainstay in the evaluation of foot and ankle pain. Weight-bearing images should be obtained whenever possible because joint space narrowing and deformity may not be apparent in non-weight-bearing images. Standard images consist of weight-bearing anteroposterior, lateral, and oblique views of the foot and anteroposterior, mortise, and lateral views of the ankle. Further radiographic findings of RA include periarticular erosions and osteopenia.

MRI provides reliable imaging of soft tissue structures and can be a useful tool in the evaluation of the rheumatoid foot and ankle. Early in the course of RA, MRI allows one to look for signs of the disease such as synovitis, tenosynovitis, periarticular edema, and bursitis. Later, MRI is useful in assessing disease progression and extent of joint involvement, as well as in distinguishing between tendon rupture and tendinitis/tendinopathy (Figure 49-2).

CT scan and nuclear scintigraphy are also used in the evaluation of foot and ankle pain. For example, either method can be quite helpful in postoperative evaluation in fusion surgery. Ultrasound is gaining utility as a method to evaluate tendon integrity as well. However, the results are largely technique dependent with minimization of artifacts being of utmost concern. In addition, ultrasound is difficult

---

**Table 49-2   Anatomic Characteristics of Pain in the Foot and Ankle**

<table>
<thead>
<tr>
<th>Location</th>
<th>Dysfunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forefoot</td>
<td>Hallux valgus, hammertoes</td>
</tr>
<tr>
<td></td>
<td>Metatarsophalangeal arthritis/synovitis/instability</td>
</tr>
<tr>
<td>Midfoot</td>
<td>Plantar fasciitis, arthritis, synovitis (rare)</td>
</tr>
<tr>
<td>Hindfoot</td>
<td>Hindfoot valgus deformity, arthritis</td>
</tr>
<tr>
<td></td>
<td>Stress fracture</td>
</tr>
<tr>
<td>Anterior ankle</td>
<td>Arthritis, synovitis, impingement, osteochondral defect</td>
</tr>
<tr>
<td>Posterior ankle</td>
<td>Achilles tendinitis/tendinosis, bursitis</td>
</tr>
<tr>
<td>Posterolateral ankle</td>
<td>Peroneal tendinitis, tear, instability</td>
</tr>
<tr>
<td>Posteromedial ankle</td>
<td>Posterior tibial tendinitis/dysfunction</td>
</tr>
<tr>
<td></td>
<td>Flexor hallucis longus/flexor digitorum</td>
</tr>
<tr>
<td></td>
<td>longus tendinitis</td>
</tr>
<tr>
<td></td>
<td>Tarsal tunnel syndrome</td>
</tr>
</tbody>
</table>
Symptoms of “impingement.” Specifically, patients will note pain with ankle dorsiflexion, such as when they walk up stairs or an incline. On physical examination, there may be anterior tenderness and/or pain with terminal passive dorsiflexion. Anterior osteophyte formation may produce even more pronounced impingement symptoms, although such osteophytes are more commonly seen with OA and long-standing ankle instability.

Central Joint Pain

Two other causes of central ankle pain, stress fracture and osteochondral defect, should be considered. Stress fractures are commonly seen in patients with periarticular and generalized osteopenia. An osteochondral defect (OCD) is a focal defect in the articular cartilage and subchondral bone. These lesions are encountered more commonly in patients without inflammatory arthritis. In the setting of RA, their presence may represent an early manifestation of RA or a separate pathologic process.

Posterior Joint Pain

Posterior ankle pain usually originates from the Achilles tendon, its insertion onto the calcaneal tuberosity, and two associated bursae in this region. The Achilles tendon is the largest tendon in the body but lacks a true synovial lining. As such, isolated Achilles tendinitis is uncommon. In most instances, Achilles pain results from degenerative tendinosis, with or without an overlying tendinitis. Although associated spur formation is common, it is important to remember that Achilles spurs are a manifestation of a disease process. As such, surgeries directed at spur excision also frequently entail tendon débridement and reconstruction, as well as tendon transfer.

The Achilles tendon is protected by two distinct bursae. A more superficial bursa is immediately subcutaneous and becomes inflamed primarily with irritation from ill-fitting shoes with a tight counter (“pump bump”). The “retrocalcaneal” bursa is a larger structure that lies deep to the Achilles. Inflammation of this structure often accompanies Achilles tendinitis/tendinosis. It may also be irritated by an enlarged posterior superior calcaneal bursitis, sometimes referred to as “Haglund’s” deformity.

Medial and Lateral Ankle Pain

As with anterior, central, and posterior ankle pain, the origin of medial or lateral ankle pain is also anatomically based.

On the medial side, pain directly over the medial malleolus should alert the clinician to the possibility of a stress fracture. Pain anterior to the medial malleolus is usually articular in nature. Pain posterior to the medial malleolus is often caused by inflammation and/or degeneration of the posteromedial flexor tendons. These include the posterior tibial tendon and the flexor hallucis longus and flexor digitorum longus tendons. The posterior tibial tendon is the largest and strongest of the posteromedial flexor tendons. Its primary function is to invert the hindfoot and thus support the medial longitudinal arch of the foot. Long-standing synovitis and dysfunction of this tendon may ultimately lead to use in preoperative planning because important landmarks are often not included. Anesthetic arthrograms are an extremely useful adjunct in diagnosing foot and ankle pain in patients with RA. Given the complex and crowded geometry of the foot and the propensity of RA to affect multiple joints and tendons, it is often difficult to determine if the pain is articular and, if so, which joint is symptomatic. With an anesthetic arthrogram, a mixture of steroid, anesthetic, and contrast material is injected under fluoroscopic guidance into a suspect joint. This allows the clinician to more precisely determine whether or not the injected joint is a significant pain generator. Again, this is especially helpful in the foot and ankle, where multiple joints are in close proximity and may be simultaneously diseased.

COMMON CAUSES OF ANKLE PAIN

From a diagnostic standpoint, it is useful for the clinician to conceptualize ankle pain on the basis of anatomic location. This approach applies to patients with virtually any form of ankle or foot pain, and RA will be used to illustrate how to formulate a differential diagnosis and treatment plan as advances in medical management temper disease and “allow” patients to develop other, noninflammatory disorders.

Anterior Ankle Pain

Anterior ankle pain, in patients both with and without RA, is most often the result of intra-articular pathology. Anteriorly, the ankle joint is not shielded by the malleoli and is immediately subcutaneous. Further, the anterior extensor tendons are typically not prone to the development of tendinitis and tendinosis.

In early RA, synovitis can cause anterior joint line pain, swelling, and tenderness. Clinically, this results in symptoms of “impingement.” Specifically, patients will note pain with ankle dorsiflexion, such as when they walk up stairs or an incline. On physical examination, there may be anterior tenderness and/or pain with terminal passive dorsiflexion. Anterior osteophyte formation may produce even more pronounced impingement symptoms, although such osteophytes are more commonly seen with OA and long-standing ankle instability.
to collapse of the arch and the development of an acquired flatfoot deformity.

On the lateral side of the ankle, pain directly over the lateral malleolus may be caused by a stress fracture. This is especially relevant in the setting of hindfoot valgus and a flatfoot, which will increase fibular loading. Similar to the medial side, pain anterior to the lateral malleolus is usually articular in nature. Finally, pain posterior to this lateral malleolus is usually indicative of peroneal tendon pathology. In patients with RA, the peroneal tendons may be affected by tenosynovitis, longitudinal “split” tears, and chronic tendon instability. With the latter, the tendons sublux over the posterolateral edge of the fibula, causing pain as well as attritional tearing.

**COMMON CAUSES OF FOOT PAIN**

Typically, the forefoot is the most common site of involvement early in the course of diseases such as RA but also can occur in gout and OA. The pathogenesis of forefoot pain and deformity in the rheumatoid forefoot is inflammation and progressive synovitis that eventually leads to a capsular distention at the MTP joints and destruction of the plantar plates. Eventually it progresses to loss of collateral ligament stability and, finally, destruction of the articular cartilage and bone (Figure 49-3A and B). Clinically, this manifests as dorsal subluxation or dislocation of the lesser toe MTP joints with a hallux valgus deformity and presents as metatarsalgia.

In the lesser MTP joints, loss of stability leads to progressive deformity secondary to the various forces on the forefoot. Muscle imbalance and dorsiflexion forces at toe-off lead to progressive subluxation and even dorsal dislocation of the MTP joints. With this, the metatarsal head is prone to forming keratotic skin lesions that can ulcerate. Muscle imbalance can also lead to the development of painful hammertoe and claw toe deformities that can exert a plantar-directed force that further exacerbates symptomatic metatarsalgia. Lesser MTP joint subluxation has been reported to be as high as 70% with a concomitant incidence of pressure sores in approximately 30% of those patients.

In the hallux, RA can cause both articular erosions and loss of capsular integrity that often results in the development of a hallux valgus deformity, or bunion (see Figure 49-1). The progression of this deformity may be further accelerated by loss of support from the adjacent lesser MTP joints. The incidence of hallux valgus deformity in patients with RA has been estimated to be up to 70%.

The midfoot is a less common site of involvement in the rheumatoid foot. Radiographically there can be erosions; however, the prevalence of symptoms is often quite low. The most frequent site of involvement is the first tarsometatarsal (TMT) joint. The symptoms seen here, though, may not be from rheumatoid synovitis per se, as seen in the forefoot. Rather, pain may also be due to hindfoot and hallux valgus deformities that lead to increased stresses across the TMT joint. Eventually this increased stress can lead to dorsiflexion of the first TMT joint and resultant lesser toe TMT joint abduction and dorsiflexion, thereby leading to pain in the dorsomedial midfoot. In addition, progressive biomechanical changes can lead to OA of the midfoot TMT joints yielding discomfort and pain with weight bearing.

The three joints of the hindfoot (talonavicular, talocalcaneal, and calcaneocuboid) are commonly affected by RA. Although these joints are affected at different rates, the overall prevalence of hindfoot involvement in RA is between 21% and 29%. The talonavicular joint is most often affected, followed by the talocalcaneal and then calcaneocuboid joints. Further, the hindfoot becomes more symptomatic and involved the longer the duration of RA. The incidence of hindfoot deformity in those with RA less than 5 years has been estimated to be 8% and increases to 25% in those with RA longer than 5 years. Clinically, patients with talocalcaneal or calcaneocuboid involvement will complain of lateral hindfoot pain. Meanwhile, arthritis and synovitis of the talonavicular joint are manifested by dorsal or medial pain.

The deformity most often seen in patients with hindfoot RA is an acquired flatfoot deformity, characterized by heel valgus and forefoot abduction. This usually results from articular deformity and instability but may also be caused by tenosynovitis and tendinosis of the posterior tibial tendon, the main supporter of the longitudinal arch of the foot.

**NONOPERATIVE TREATMENT**

Medical management remains the cornerstone of treatment for many forms of foot and ankle arthritis. In fact, many of the current recommendations for operative treatment may soon be modified given the alteration of disease progression with current medical regimens for RA. The most common medical management still consists of nonsteroidal anti-inflammatory drugs (NSAIDs), steroids, and disease-modifying antirheumatic drugs (DMARDs) and, more recently, biologic therapies. Although each of these drug classes has done much to alleviate patient suffering, they are not without impact on the surgical management of rheumatic disease. There is concern about lower fusion rates in
the setting of NSAIDs and increased infection rates in the setting of steroids or DMARDs. Moreover, those who have been on chronic steroids are at risk for postoperative adrenal insufficiency and may require perioperative corticosteroids. Close communication and collaboration between the rheumatologist and surgeon are essential to good outcomes.

Footwear modification can often have profound benefits for patients. Shoes should be examined in the clinic to be sure that they can accommodate a patient's deformity. Patients often feel best in shoes with a deep, wide toe-box, a firm heel counter, and soft heel. Well-constructed walking or jogging shoes usually provide sufficient room for mild to moderate deformities. It is helpful to provide patients with a list of suitable manufacturers when making such recommendations.

Often it is necessary to prescribe a custom orthotic insert for those with more moderate deformities. Typically the insole of the shoe must be removed in order to make room for the orthotic. Again, most walking or jogging shoes will suffice. In general, custom orthotics can be divided into rigid, semirigid, and softer accommodative devices. Rigid and semirigid orthotics are usually used to correct supple deformities and should be used with caution in patients with RA. More commonly, patients benefit from accommodative orthotics (i.e., orthotics made of softer material that can be molded to "accommodate" a deformity). These can then be further modified by incorporating a "relief" under a deformity, thereby further unloading it. When sending patients for orthotics, it is best to provide the orthotist with a prescription that includes the patient's precise diagnosis (e.g., RA or OA with metatarsalgia), as well as the type of orthotic and any modifications desired (e.g., a "custom accommodative orthotic with a relief under the lesser metatarsal heads").

Finally, injections of a mixture of anesthetic and corticosteroid to areas of inflammation or bursitis are useful in the treatment of both inflammatory and noninflammatory conditions affecting the foot and ankle. In the foot and ankle, however, such injections must be judiciously employed. Most importantly, injections into and around tendons should be avoided. Due to the forces associated with weight bearing and ambulation, these tendons are under substantial load. The injection of a corticosteroid directly into or even near a tendon can adversely affect the biomechanical properties of the tendon, ultimately leading to rupture. A further precaution is to avoid corticosteroid injections into the lesser MTP when there is evidence of joint instability (manifested by valgus or varus deviation on radiographs or sagittal plane instability on physical examination). Such injections can lead to further attenuation of the joint capsule and can result in frank joint dislocation.

**OPERATIVE TREATMENT**

If symptoms persist despite nonoperative management, surgical intervention should be considered. Two important factors must be taken into account when deciding whether or not to proceed with surgery. First, the soft tissues and vascular status must be carefully assessed. Both may be compromised and could negatively affect outcome. Second, the ability of patients to comply with the postoperative regimen (e.g., being able to use crutches and not bear weight if necessary) must be considered. Even limited non-compliance can lead to a poor outcome, especially in fusion surgery.

As noted earlier, most surgical procedures fall into one of the following categories: arthrodesis (joint fusion), arthroplasty (joint replacement), corrective osteotomy, ostectomy, and synovectomy (joint or tendon).

**Arthrodesis**

Arthrodesis remains a surgical cornerstone for the rheumatoid foot and ankle. With an arthrodesis procedure, the two sides of the joint are roughened with a burr or small chisel. Next, the two bones to be fused are compressed and fixed together, usually with one or more screws (see Figure 49-3B). In the weeks and months following surgery, the body is “tricked” into thinking that there is a fracture present at the fusion site and heals this with bone. As such, the two bones become one and are considered fused. Fusion surgery offers reliable pain relief in the majority of patients. One obvious concern with fusion surgery is the loss of motion. For the patient, however, this usually results in only mild functional compromise. Further, to the untrained eye, there is remarkably little change in gait.

Commonly performed fusions in patients with RA include ankle arthrodesis, isolated hindfoot fusions, triple arthrodesis, midfoot arthrodesis, and arthrodesis of the first MTP joint. A triple arthrodesis involves fusion of the talocalcaneal, talonavicular, and calcaneocuboid joints. Together, these joints allow coronal plane motion and thereby are most important when walking on uneven ground.

Fusion remains the gold standard for patients with RA of the ankle. If there is minimal deformity and no loss of bone stock, ankle fusion surgery may be performed arthroscopically or through a “miniopen” approach. These techniques involve less soft tissue dissection and stripping, thereby minimizing loss of bony perfusion. Nevertheless, the time period for which the patient must avoid bearing weight (from 6 to 12 weeks) remains the same. The success rate of ankle fusion surgery in patients with RA is generally 85% or greater. Although the osteopenia associated with the disease can compromise fixation, it can also theoretically enhance fusion because there is less sclerotic subchondral bone.

In the hindfoot, fusion surgery may be performed on one or more of the three joints of this part of the foot (i.e., the talocalcaneal, talonaviclar, and calcaneocuboid joints). If only one of these joints is diseased, an isolated fusion of this joint is acceptable. This reduces surgical morbidity and the extent of the procedure. Nevertheless, with fusion of just one of the joints of the hindfoot, motion in the other joints is reduced. If more than one joint is diseased, a “double” or “triple” arthrodesis is necessary.

In the midfoot, fusion surgery results in negligible loss of motion because the joints of the midfoot normally have less than 10 degrees of motion. With both OA and inflammatory arthritis, symptomatology is most often limited to the medial (first through third) TMT joints. The lateral (fourth
and fifth) TMT joints are infrequently symptomatic, even in the setting of advanced radiographic changes.

Finally, in the forefoot, fusion surgery is indicated only for the first metatarsophalangeal (MTP) joint. This procedure is used for both arthrosis and advanced hallux valgus (bunion) deformities. When the first MTP joint is fused, it is positioned in a slightly dorsiflexed position to assist ambulation. With MTP fusion in 47 feet, Coughlin reported 96% good to excellent results and 100% fusion at an average 6.2-year follow-up.

In summary, fusion surgery generally provides reliable pain relief and a stable, plantigrade foot. Nevertheless, the loss of motion of the fused joint can lead to increased motion and altered biomechanics at adjacent joints. This ultimately may lead to arthritic changes in these joints. Further, fusion surgery may lead to subtle albeit real changes in gait. Finally, the minimal ramifications of fusing just one joint may become much greater in the setting of a subsequent fusion in either the ipsilateral or contralateral limb.

**Arthroplasty**

Concerns regarding fusion have driven many to work toward improving joint replacement surgery (arthroplasty) in the foot and ankle. Most notably, total ankle replacement surgery continues to evolve and remains a controversial topic among orthopedic surgeons. Although there are many who perform ankle replacement surgery, there are many that do not or that do so only on a limited basis. To this end, the U.S. Food and Drug Administration currently approves only four ankle prostheses for implantation. Long-term survival data as published for hip and knee arthroplasty are not yet available.

The main advantage of ankle arthroplasty is preservation of motion. Its main two disadvantages are technical complexity and the difficulty with subsequent fusion if the procedure fails. In general, ankle replacement surgery is indicated for middle-aged and elderly individuals with low functional demands and minimal deformity. Two other indications especially pertinent in patients with ankle arthritis include (1) bilateral disease and (2) concomitant ipsilateral hindfoot disease or preexisting arthodesis. The paradox of ankle replacement surgery remains as follows: Ankle replacement is contraindicated in young patients for whom preservation of motion is most important. On the other hand, arthroplasty is more commonly performed in older patients for whom preservation of motion is less important and who would do well with a fusion. Nevertheless, total ankle replacement surgery continues to evolve and reported success rates with modern designs continue to improve (Figure 49-4).

In the foot, arthroplasty is performed by some surgeons for the first MTP joint. The relevant literature is still somewhat conflicted, though. Although there were some encouraging early results with arthroplasty, other studies have shown high rates of implant failure and loosening secondary to synovitis from polymeric silicone (Silastic) particle wear. Further, advanced deformity, often present in patients with RA, is considered to be a relative contraindication to first MTP joint arthroplasty. Nevertheless, new implant designs may hold increased promise. In general, these implants are lower profile and resect less bone, which also makes it easier to perform a subsequent fusion, if necessary.

**Osteotomy**

Corrective osteotomies are used primarily for two reasons in the treatment of RA: to correct deformity and/or to redistribute forces on a joint or the terminal aspect of a bone.

Examples of osteotomies to correct deformity include calcaneal osteotomies for pes planovalgus and metatarsal osteotomies for hallux valgus. Previously, patients with RA and concomitant pes planovalgus or hallux valgus underwent fusion surgery. However, with advances in medical management of the disease, it is not unreasonable to attempt joint preservation surgery in patients who have mild to moderate disease, healthy soft tissues, and flexible deformities.

Examples of osteotomies to redistribute forces include tibial osteotomies in the setting of eccentric ankle arthritis and metatarsal osteotomies in the setting of metatarsalgia. Patients requiring surgery for ankle RA previously underwent fusion surgery only, while patients requiring surgery for metatarsalgia underwent metatarsal head resection. Again, however, advances in medical management of the disease allow joint preservation osteotomies to be considered. This is especially the case for metatarsalgia, which is common in patients with RA yet increasingly does not entail frank dislocation or articular erosion.

**Osteotomy**

Although more commonly seen in patients with OA, some patients with RA may present with symptoms of mechanical ankle impingement arising from anterior bone spurs. In cases without global joint destruction, surgical resection of
the spurs, or cheilectomy, is a reasonable treatment. Although no studies have examined cheilectomy in RA specifically, patients with less severe erosive changes tend to be more satisfied with the results of cheilectomy.  

Synovectomy

For those patients with inflammatory arthritis resistant to medical management and nonoperative treatment, synovectomy can provide a period of pain relief for many patients.36,37 It is thought that early synovectomy of either the affected joint or tendon may help halt the progress of joint destruction. Joint synovectomy is indicated in those who have failed medical management yet still have a relatively preserved articular surface. Otherwise, synovectomy of the affected tendons allows some preservation of function.

CONCLUSION

Foot and ankle pain is a prevalent and potentially debilitating problem. Unfortunately, many forms of arthritis, including RA, set up a vicious cycle of foot and ankle pain and biomechanics. Synovitis and articular erosions lead to both pain and deformity. A proper history and physical examination are essential for establishing an anatomic diagnosis. Although advanced imaging modalities such as MRI and CT can be useful as adjuncts, radiography remains the gold standard. Nonoperative modalities such as medications, bracing, physical therapy, orthotics, and footwear modification are able to relieve pain and maintain function for many. For recalcitrant symptoms, substantial relief are able to relieve pain and maintain function

References


The references for this chapter can also be found on www.expertconsult.com.
Hand and Wrist Pain

CARRIE R. SWIGART

KEY POINTS

- Patients with carpal tunnel syndrome typically present with nocturnal paresthesias associated with intermittent pain or paresthesia during the day.
- Ganglia are mucin-filled cysts arising from joint capsules or tendon sheaths. If they are particularly symptomatic, corticosteroid injections may be tried, but surgical excision may be necessary to effect a cure.
- Tendinitis of the extensor pollicis longus tendon can be particularly dangerous because of the risk of tendon rupture.
- De Quervain’s disease, inflammation of the extensor pollicis brevis and abductor pollicis longus tendons in the first dorsal extensor compartment, is common in women and is associated with repetitive hand activities such as caring for an infant.
- Painful osteoarthritis involving the carpometacarpal joint of the thumb can often be treated successfully by splinting.
- Trigger fingers, caused by thickening of the A1 retinacular pulley in the palm, usually can be treated by corticosteroid injections and splinting.

The multiple functions that the hand performs in daily life are usually taken for granted until they become affected by disease or injury. Depending on the nature of the disorder, patients have different capacities to adapt. Patients presenting with pain and dysfunction of the hand or wrist or both represent a wide spectrum, diverse in age, occupations, and avocations. These patients have a broad range of medical conditions that may or may not be related to their current problem. Each patient has a different story to tell about his or her hand and wrist and why he or she is seeking treatment. It is up to the clinician to sort out these various factors, some of which may seem confounding, and determine the most appropriate diagnosis and course of treatment.

This chapter presents guidelines that are useful in the evaluation of patients presenting with hand and wrist pain. Complete coverage of all of the various conditions that can affect the hand and wrist is beyond the scope of this chapter. Instead, the conditions discussed include the most common pathologies seen by general practitioners and hand surgeons. The conditions are grouped by their anatomic area to include pain localized to the volar, dorsal, radial, or ulnar wrist; the base of the thumb; and the palm and digits.

PATIENT EVALUATION

Anatomy

The complex anatomy of the hand and wrist involves many structures interacting in close proximity to one another. Several different diagnoses can manifest with similar symptom patterns despite varying pathologies. A precise knowledge of the anatomy of the hand and wrist often eliminates several diagnostic considerations on the basis of the physical examination alone. The history of the illness and the examination also help to narrow further investigation by enabling the physician to choose appropriate additional diagnostic tests better. Several common sites of pain in the hand and wrist and their corresponding leading diagnoses are illustrated in Figure 50-1. Pain in one location can have multiple etiologies depending on the patient profile and the history of the problem. A thorough review of the pertinent regional anatomy is important to help differentiate successfully the many possible causes of hand and wrist pain.

History

Important patient factors include age, sex, hand dominance, occupation, and hobbies or sports. When determining the history of the problem, a history of recent or distant trauma should be sought and an estimation of the severity of the trauma should be noted. Next, questions about the duration and frequency of the pain and the intensity and quality should be addressed. The pain of degenerative arthritis is often described as a localized “toothache”-type pain, which is always present at a low level and increases with activity, whereas the pain of tendinitis may be sharp, poorly localized, and present only with activity. Rheumatoid arthritis (RA) manifests initially with hand and wrist involvement in 25% of patients and is characterized by joint effusion, with bilateral hand and wrist involvement and morning stiffness. Nighttime symptoms of a burning-type pain in the hand and wrist that are exacerbated by arm position are often associated with nerve entrapment syndromes. Specific activities that either cause pain or alleviate it should also be noted. Arthritis at the base of the thumb, or first carpometacarpal joint, is often aggravated by opening jars, turning doorknobs, and doing needlework or other hobbies.

PHYSICAL EXAMINATION

A thorough examination of the involved extremity and comparison with the uninvolved extremity are essential. Attention should be paid to abnormalities of the more proximal joints of the elbow and shoulder and the cervical...
spine. As the differential diagnosis narrows, the examination should be tailored as needed to include or eliminate any possible systemic etiologies. As with other musculoskeletal examinations, a complete record of the range of motion of the involved joints and comparison measurements of the opposite side should be made. Any difference between active and passive motion should be noted. Careful palpation for the site of maximal tenderness is important in differentiating the source of pain and is particularly important when trying to exclude possible factors of secondary gain. Measurements of grip and pinch strength are also helpful in many situations as a diagnostic aid and a baseline measurement to follow for improvement. Many provocative maneuvers are useful in differentiating etiologies; these are discussed with the specific pathology with which they are associated.

**Imaging Studies**

Technologic advances have increased the availability of imaging studies for the hand and wrist. Improvements in magnetic resonance imaging (MRI) resolution using small joint coils allow more precise imaging of small structures in the hand and wrist. Advancements in ultrasound technology have allowed this tool to be increasingly used in the diagnosis of musculoskeletal complaints. With the multitude of ancillary studies available, it is important to be selective in using these to establish or refute diagnostic possibilities. In this era of cost containment, imaging studies should be used most often to confirm a diagnosis rather than to find one. An understanding of the advantages and limitations of each study is necessary to enable using them to their fullest potential.

Plain radiographs are the easiest and most readily available study that can be obtained in most offices. A routine hand or wrist series including anteroposterior, lateral, and oblique views is a useful screening tool but often lacks the specificity required. Depending on the suspected diagnosis, there are many available special views. These are discussed with the specific diagnoses to which they pertain later in this chapter.

If further detail of the bony anatomy is required, computed tomography (CT) is the best available tool today. The most common uses for CT in the hand and wrist include evaluation of intra-articular fractures of the distal radius and metacarpals, scaphoid fractures and nonunions, and intraosseous cysts or tumors.1,2

Advances in ultrasound and MRI technology have enhanced the ability to evaluate the soft tissue structures of the hand and wrist. Smaller ultrasound probes with higher resolution have made it possible to visualize and differentiate structures such as flexor tendons, ganglion cysts, and ligaments. Doppler ultrasound can help to differentiate vascular disorders of the hand. MRI technology is constantly improving and allowing for new uses in the hand and wrist. By altering the parameters of this test, information about anatomy and physiology can be obtained.3 Specific uses of these tests and others such as arthrography and bone scans are addressed with the diagnoses for which they are most useful.

### Additional Diagnostic Tests

#### Neurodiagnostic Tests

Neurodiagnostic tests including nerve conduction studies and electromyography are useful in the diagnosis of suspected neurologic disorders of the upper extremity. Specifying the type and nature of examination required enhances the information gained by these studies. If a nerve compression syndrome such as carpal or cubital tunnel syndrome is suspected, nerve conduction studies may be sufficient without the added cost and patient discomfort of formal electromyography testing. Nerve conduction studies evaluate the speed of conduction of motor and sensory nerves across a set distance at a specific location and compare this with established normal values. A decrease in the speed of nerve conduction, as evidenced by an increase in the latency, is seen with localized nerve compression and is shown in several different nerves concomitantly in demyelinating diseases such as multiple sclerosis. When more severe nerve injuries are suspected or if there is clinical evidence of muscle weakness or atrophy, an electromyogram can be useful to delineate better the extent of the process or rule out a myopathic process.4
Injections and Aspirations

The use of injections and aspirations can be therapeutic and diagnostic. A so-called lidocaine challenge can be used to discriminate between different diagnoses when placed precisely in one joint or painful area. Corticosteroids can be given selectively in conjunction with the local anesthetic for more lasting relief and in some cases can be curative. Some of the most common sites for injection are the A1 pulley region of the finger for trigger finger, the carpal canal for carpal tunnel syndrome, and the first dorsal compartment of the wrist for de Quervain’s disease.

Aspiration of joints or other fluid collections such as ganglia can yield vital diagnostic information and can be therapeutic. If infection is suspected, aspiration should be used to obtain a sample of joint fluid for Gram stain, cell count, and culture. Diagnoses such as gout and pseudogout can be confirmed by crystal analysis under polarized light. Many ganglia and retinacular cysts can be treated temporarily or permanently with simple aspiration.

Arthroscopy

Direct visualization of a joint via arthroscopy can be an invaluable diagnostic tool. Despite the increasing sensitivity of imaging techniques such as MRI, arthroscopy provides a dynamic evaluation that static imaging cannot provide. Since the first published report of a series of cases by Roth and colleagues in 1988, it has become the “gold standard” for evaluation of chronic wrist pain. With new surgical techniques being developed, surgeons often can proceed directly to the definitive treatment using arthroscopy entirely or in part.

COMMON ETIOLOGIES FOR HAND AND WRIST PAIN

Wrist Pain: Palmar

Carpal Tunnel Syndrome

Carpal tunnel syndrome (CTS) is the most commonly diagnosed compression neuropathy in the upper extremity. It usually occurs as an isolated phenomenon, but symptoms of CTS can accompany many systemic diseases such as congestive heart failure, multiple myeloma, and tuberculosis. More commonly, CTS is associated with conditions such as pregnancy, diabetes, obesity, rheumatoid arthritis, and gout.

The classic constellation of symptoms consists of nocturnal paresthesias in the affected digits; paresthesias or hypoesthesias in the thumb, index, and long fingers; and weakness or clumsiness of the hand. Patients often complain of forearm and elbow pain that is aggravated by activities but is poorly localized and aching in nature. Occasionally, more proximal symptoms such as shoulder pain are the main presenting complaint. Past reports have indicated a 3:1 prevalence of CTS in women. Approximately half of patients are 40 to 60 years old, although CTS has occasionally been diagnosed in children.

The diagnosis of CTS is usually clinical. Tinel’s sign, shown by radiating paresthesias in the median nerve distribution with gentle percussion over the volar wrist, indicates nerve irritation. Reproduction of symptoms with wrist flexion, as described by Phalen, and with the carpal compression test, as described by Durkan, has been shown to be more specific. Decreased sensibility and thenar atrophy are late signs seen in advanced median nerve entrapment. Bilateral electrodiagnostic tests, specifically nerve conduction velocity testing, should be used to confirm the diagnosis, particularly in patients claiming a compensable injury or in patients with atypical signs or symptoms. Prolonged motor and sensory latencies across the carpal canal confirm pathologic compression of the median nerve.

In patients with classic clinical findings, a study found that CTS could be diagnosed with a high degree of accuracy on clinical grounds alone and that the addition of electrodiagnostic tests did not increase the accuracy. When attempting to differentiate CTS from more proximal nerve entrapments such as cervical root compression or thoracic outlet syndrome, the addition of electromyography of the cervical paraspinal muscles and proximal conduction tests (H reflex, f waves) can be useful.

Conservative treatment for CTS consists of splinting of the wrist in neutral position and consideration of oral nonsteroidal anti-inflammatory drugs (NSAIDs) for pain control. Splinting should be used sparingly during the workday to prevent secondary muscle weakness and fatigue but is best prescribed to prevent provocative wrist positioning at night. The splint should not hold the wrist in extension beyond 10 degrees (Figure 50-2). Although splinting may be beneficial for relief of symptoms in cases of mild compression, its long-term effectiveness is limited. The use of vitamin B6 (100 to 200 mg/day) has been helpful in some cases, but its efficacy has not been confirmed in a randomized trial. The popularity of injections of corticosteroid in the treatment of CTS has waxed and waned over the last half century. Although it has been shown to be quite effective in the short term, the long-term efficacy is mixed. Also, injections have been associated with exacerbation of the condition and permanent median nerve injury if performed incorrectly. For these reasons, injections are most often indicated in cases when the condition is thought to be temporary such as with pregnancy or if surgery must be deferred because of a medical condition or major life event.

Figure 50-2 Typical night splint used in the treatment of carpal tunnel syndrome.
Ulnar Nerve Entrapment: Cubital Tunnel Syndrome

Entrapment of the ulnar nerve as it passes through the cubital tunnel just posterior to the medial epicondyle of the elbow can manifest with symptoms localized to the ulnar border of the hand. Medial forearm pain and irritability of the ulnar nerve at the elbow may be associated as well. Presenting symptoms usually consist of paresthesias or numbness or both in the small and ring fingers. Percussion of the nerve in the cubital tunnel elicits Tinel’s sign. Prolonged elbow flexion reproduces the symptoms. In contrast to carpal tunnel syndrome, it is not unusual for patients to present with early atrophy of the intrinsics, most easily appreciated in the first dorsal interosseous muscle.

Electrodiagnostic studies can help to confirm the diagnosis and differentiate cubital tunnel syndrome from more distal compression of the ulnar nerve in Guyon’s canal (see later). If malalignment of the elbow is present or the patient relates a history of childhood trauma, radiographs should be obtained to rule out a supracondylar or epicondylar malunion. So-called tardy ulnar nerve palsy can develop years after a supracondylar fracture of the elbow.67

Conservative treatment includes strategies to help the patient avoid having the elbow flexed for prolonged periods, particularly at night. Soft, or semirigid, elbow splints prevent elbow flexion beyond 50 to 70 degrees. Medial elbow pads can be used if the patient’s job or hobbies require resting the medial elbow on a hard surface. NSAIDs can be beneficial in acute or traumatic cases. Surgical decompression of the nerve is indicated if a patient fails to obtain relief from splinting and activity modification or if there is clinical or electrodiagnostic evidence of muscle denervation.

Ulnar Nerve Entrapment: Guyon’s Canal

In 1861 Guyon58 published a description of the contents of an anatomic canal at the wrist. The distal branches of the ulnar nerve and the ulnar artery pass through this space. As it exits the canal, the ulnar nerve divides into its sensory and motor branches. Compression of the nerve within or proximal to the canal usually manifests with a combination of sensory and motor symptoms in the ulnar nerve distribution. Patients complain of numbness and paresthesias of the palmar aspect of the ring and small fingers. Motor symptoms are usually described as a cramping weakness with grasping and pinching. As with median neuropathy, atrophy of the intrinsics and objective sensory loss are late findings.

In contrast to carpal tunnel syndrome, in which patients usually have an ill-defined onset of symptoms, ulnar nerve compression in the canal of Guyon is often of more acute onset. It can be associated with repeated blunt trauma, a fracture of the hamate or the metacarpal bases, or occasionally a fracture of the distal radius.62,63 Space-occupying lesions such as a ganglion, lipoma, or anomalous muscle can also cause compression.64,65 Because of the difference in etiology, this nerve entrapment syndrome is often not amenable to conservative treatment. If there is an anatomic lesion such as a fracture or a mass, this must be addressed. If repetitive blunt trauma is the cause, without associated fracture or arterial thrombosis, splinting and activity modification can alleviate the symptoms.

Flexor Carpi Radialis and Flexor Carpi Ulnaris Tendinitis

Similar to other tendinopathies around the wrist, irritation of the wrist flexors occurs with stress of the wrist in a particular position. Activities that require forced wrist flexion for prolonged periods or with repetition put patients at risk for inflammation around the flexor carpi radialis tendon69 or the flexor carpi ulnaris tendon or both. The condition manifests with tenderness along the course of the tendon, especially near its insertion. Wrist flexion against resistance with radial or ulnar deviation reproduces the symptoms. Treatment consists of splinting and rest, elimination of activities that cause pain, and oral NSAIDs. Injection of corticosteroid into the flexor carpi radialis or flexor carpi ulnaris sheath may be curative. Sharp pain, associated with an intense inflammatory localized reaction, is suggestive of calcific tendinitis and is most commonly seen around the flexor carpi ulnaris tendon.70,71 If calcific tendinitis is suspected, a plain radiograph can be useful in confirming the diagnosis but the calcification may not become apparent for 7 to 10 days after the onset of symptoms.

Hamate Fracture

An uncommon and underdiagnosed etiology of palmar pain in young, active individuals is a fracture of the hook of the hamate. These fractures can occur from a fall on an extended wrist, a “dubbed” golf shot, or from forcefully striking a ball with a club or bat. Plain radiographs of the wrist are usually
read as normal. The condition should be established and treated expeditiously because it may lead to ulnar nerve entrapment, ulnar artery thrombosis, or rupture of flexor tendons. Pain in the base of the palm overlying the hamate is the most common presenting symptom. Often, the pain is present only with the activity that caused the fracture such as driving a golf ball or swinging a bat. Because of the proximity of the ulnar nerve, patients also can have sensory and motor symptoms of distal ulnar neuropathy. Occasionally, in the acute setting, vascular complaints such as cold intolerance or frank ischemia from ulnar artery thrombosis can be the presenting condition.

A carpal tunnel view, obtained with the wrist in a hyperextended position, may show the fracture (Figure 50-4A). Alternatively, a selective CT scan through the hamate is a more accurate way to confirm the diagnosis (Figure 50-4B). If diagnosed within 2 to 3 weeks of injury, casting should be attempted to allow the fracture to heal. If this fails or if the fracture is diagnosed late, surgical treatment is indicated, and most authors favor excision of the hook followed by a gradual return to activities.

**Wrist Pain: Dorsal**

**Ganglion**

Ganglia account for 50% to 70% of all soft tissue tumors of the hand and wrist. Of these, 60% to 70% occur around the dorsal wrist. These mucin-filled cysts usually arise from an adjacent joint capsule or tendon sheath. The most common site of origin is the scapholunate ligament, and the main body of the cyst may be located elsewhere on the dorsum of the wrist and attached to this ligament by a long pedicle. Although most ganglia occur as a well-circumscribed and obvious soft mass, some are subtler and are evident only with the wrist in marked volar flexion. As a result of their characteristic appearance, ganglia are not often misdiagnosed but should be differentiated from the less well-demarcated swelling of extensor tenosynovitis, lipomas, and other hand tumors. Plain radiographs are usually normal but occasionally show an intraosseous cyst or an osteoarthritic joint. Some ganglia may not be clinically apparent and are known as “occult” ganglia. Ultrasound and MRI have been shown to be useful in the diagnosis of these ganglia.

Not all ganglia are painful. Patients may present with complaints of wrist weakness or simply because of the cosmetic appearance of the cyst. In approximately 10% of cases, there is evidence of associated trauma to the wrist. The ganglia may appear suddenly or develop over many months. Intermittent complete resorption followed by reappearance months or years later is common.

Most conservative measures such as splinting and rest have only a temporary effect on ganglia. They tend to diminish in size with rest and enlarge with increased activity. Spontaneous rupture is common, and at one time attempting to rupture the cyst with a heavy object such as a large book was recommended as treatment. Aspiration can be performed but has mixed results because of the thick gelatinous nature of the fluid within the cyst. Even if adequate decompression of the cyst can be achieved, recaccumulation of the fluid usually occurs. Aspiration in conjunction with irrigation or injection of corticosteroids can be effective in alleviating the symptoms for varying periods of time.

Occasionally, a ganglion can become so large that it can interfere with the function of the wrist by limiting the motion, especially in extension. Pressure of the mass on the terminal branches of the posterior interosseous nerve may be painful. Excision is generally curative but may result in short-term stiffness and some loss of terminal flexion secondary to surgical scarring. Occasionally, a patient desires excision of the cyst for cosmetic reasons. With proper excision, recurrence is less than 10%, but if the dissection is incomplete and fails to identify the origin of the cyst, recurrence rates can be 50%. Arthroscopic resection has been shown to be a safe and effective method of treating dorsal wrist ganglia.

**Carpal Boss**

Often confused with a dorsal ganglion, the carpal boss is a bony, nonmobile prominence on the dorsum of the wrist. It is an osteoarthritic spur that forms at the second or third carpometacarpal joints. The boss is most evident with the wrist in volar flexion. Patients usually present with pain and localized tenderness over the prominence. The condition is twice as common in women as in men, and most patients are in their 20s to 30s. It is not unusual for a small ganglion to be associated with the boss. Radiographs are best taken with the hand and wrist in 30 to 40 degrees of supination and 20 to 30 degrees of ulnar deviation to put the bony prominence on profile (the “carpal boss view”).
Conservative treatment consists of rest, immobilization, NSAIDs, and occasionally injection with corticosteroids. If persistently painful despite these measures, surgical excision of the boss may be necessary but is associated with a prolonged recovery and continued symptoms in a high percentage of patients.

**Extensor Tendinopathies**

The extensor pollicis longus (EPL) tendon can be irritated as it passes around Lister’s tubercle. This condition, in contrast to other tendinopathies around the wrist, carries a significant risk of tendon rupture. Early diagnosis and sometimes urgent operative treatment are necessary to prevent this complication. Localized pain, swelling, and tenderness are the hallmarks of this condition, and similar to other tendinopathies, initial treatment consists of decreased activity and splinting. A short course of oral anti-inflammatory medication can be useful in decreasing symptoms. Diagnostic injections with lidocaine can help to differentiate the condition from other causes of wrist pain, but corticosteroid injections are not routinely used in this condition because of a propensity for the EPL to rupture in chronic cases.

Commonly, a patient may present with a rupture of the EPL without antecedent pain or swelling. There is a well-known association of EPL rupture with fractures of the distal radius that likely occurs owing to a relative “watershed zone” of vascular supply within its tight retinacular sheath. Tendon rupture most often occurs with minimally displaced or nondisplaced fractures and can occur several weeks or months after the original injury. Individuals with RA and systemic lupus erythematosus are especially prone to rupture of the EPL and other tendons.

**Kienböck’s Disease**

Kienböck’s disease is so named for Kienböck, who first described in 1910 what he postulated were avascular changes in the lunate. Nearly a century later, the cause of this disease remains unclear; it is likely multifactorial. Kienböck’s disease should be suspected when a young adult presents with pain and stiffness of the wrist and swelling and tenderness around the region of the dorsal lunate. There is an increased propensity of the disease among patients with an ulna that is anatomically shorter than the radius (so-called ulnar negative variance). Radiographs are needed to confirm and stage the process. Kienböck’s disease is staged by the degree of fragmentation and collapse of the lunate, associated osteoarthritis, and carpal collapse in a system originally proposed by Stahl.

In this system, the earliest sign of the disease is a linear or compression fracture in the lunate. Later stages show sclerosis of the lunate, followed by lunate collapse and a loss of carpal height. In the final stage the carpus shows signs of diffuse osteoarthritis with complete collapse and fragmentation of the lunate (Figure 50-5). With the increased sensitivity of MRI, it is possible to identify avascular changes within the lunate before they become evident on plain radiographs. This is referred to as “stage zero” Kienböck’s disease.

The treatment for Kienböck’s disease is largely surgical. Depending on the stage of the disease and the postulated etiology, several surgical procedures have been described. In early stages of the disease, when there is little lunate collapse and no osteoarthritis, the goal of surgery is to “unload” the lunate by redistributing articular contact forces and allow it to revascularize. The most common procedure is a radial shortening osteotomy, performed to neutralize ulnar variance. In later stages, various intercarpal arthrodeses have been used to readjust and maintain carpal height and alignment. Microsurgical techniques have been used more recently to revascularize the lunate with promising early results.

**Scapholunate Interosseous Ligament Injury**

The interosseous ligament between the scaphoid and the lunate is a stout structure, especially dorsally, and usually

---

**Figure 50-5** Advanced Kienböck’s disease, showing carpal collapse, intercarpal and radiocarpal arthrosis, and fragmentation of the lunate. A, Posteranterior view. B, Lateral view.
The radius and ulna must remain congruent through a 190-degree arc. Limitation of motion and pain with pronation and supination are consistent with a tear of the supporting ligaments and resultant distal radioulnar joint (DRUJ) instability. If a sufficient portion of the stability has been lost, the ulna appears clinically dislocated or subluxated, and there is severe limitation of forearm rotation. Lateral radiographs of the wrist in neutral and full pronation and supination are not generally specific enough to confirm ulnar subluxation. To evaluate better the congruency of the DRUJ through its range of motion and to assess for subtle subluxations, CT can be performed on both wrists simultaneously in positions of neutral, full pronation, and full supination.

Tears of the TFCC may manifest with painful clicking during wrist rotation. Patients generally have localized tenderness on the midaxial border of the wrist and directly beneath the extensor carpi ulnaris tendon. If forced ulnar deviation of the wrist or gripping or both reproduce the patient’s symptoms, a degenerative tear of the central portion of the TFCC is more likely. The degenerative tear is frequently a component of the ulnocarpal impaction syndrome, a condition associated with higher than normal loads on the ulnar carpus secondary to a congenitally positive ulnar variance.

Plain radiographs are most useful in determining ulnar variance and for ruling out fractures or arthritis as a cause of ulnar wrist pain. Because of the variable relationship of the radius and ulna depending on forearm rotation, it is important to take standardized films when measuring ulnar variance.

Early surgical intervention is recommended with the goals of maintaining carpal alignment and preventing an otherwise inevitable progression to carpal collapse and degenerative arthritis.

**Gout and Inflammatory Arthritis**

All of the inflammatory arthritides including the crystal arthropathies can manifest as dorsal wrist pain. Approximately 25% of patients with a diagnosis of RA present initially with hand and wrist symptoms. The reader is referred to Chapters 94 to 96 for further details.

**Wrist Pain: Ulnar**

**Triangular Fibrocartilage Complex Injury and Ulnocarpal Impaction Syndrome**

One of the most complex and confusing areas of the wrist from a diagnostic standpoint is the articulation of the ulna with the carpus. The triangular fibrocartilage complex (TFCC), so named by Palmer and Werner, comprises the articular disk itself and the immediately surrounding ulnocarpal ligaments. It can be injured by a variety of acute and chronic mechanisms. Hyperpronation and hypersupination of the carpus during forceful activities are the usual causes of acute injuries, whereas repetitive pronation and supination more often cause attritional changes in the TFCC. Careful physical examination is important to determine the origin of the pain and to try to discover the maneuver or wrist position that most closely reproduces the symptoms.

The radius and ulna must remain congruent through a 190-degree arc. Limitation of motion and pain with pronation and supination are consistent with a tear of the supporting ligaments and resultant distal radioulnar joint (DRUJ) instability. If a sufficient portion of the stability has been lost, the ulna appears clinically dislocated or subluxated, and there is severe limitation of forearm rotation. Lateral radiographs of the wrist in neutral and full pronation and supination are not generally specific enough to confirm ulnar subluxation. To evaluate better the congruency of the DRUJ through its range of motion and to assess for subtle subluxations, CT can be performed on both wrists simultaneously in positions of neutral, full pronation, and full supination.

Tears of the TFCC may manifest with painful clicking during wrist rotation. Patients generally have localized tenderness on the midaxial border of the wrist and directly beneath the extensor carpi ulnaris tendon. If forced ulnar deviation of the wrist or gripping or both reproduce the patient’s symptoms, a degenerative tear of the central portion of the TFCC is more likely. The degenerative tear is frequently a component of the ulnocarpal impaction syndrome, a condition associated with higher than normal loads on the ulnar carpus secondary to a congenitally positive ulnar variance.

Plain radiographs are most useful in determining ulnar variance and for ruling out fractures or arthritis as a cause of ulnar wrist pain. Because of the variable relationship of the radius and ulna depending on forearm rotation, it is important to take standardized films when measuring ulnar variance. A posteroanterior view of the wrist with the shoulder abducted to 90 degrees and the elbow flexed to 90 degrees shows the DRUJ in neutral forearm rotation and is easily reproducible. Because the ulna lengthens relative to the radius during power grip, a radiograph in
the same position during maximal grip best shows impaction of the ulna on the carpus.

Ancillary studies for TFCC tears include three-compartmental arthrography and MRI. In arthrography, sequential injections of radiopaque dye are performed into the carpal joint, midcarpal joint, and DRUJ. The test is considered positive when the dye is seen leaking from one compartment to another. The site of the leak determines the location of the torn structure.110 Several studies have shown, however, that there are age-related attritional tears, which occur in the TFCC and other ligamentous structures of the wrist.111-113 Technologic advancements in MRI have improved the ability to visualize and diagnose abnormalities in the TFCC. MRI can be combined with arthrography to visualize better the TFCC and the intrinsic wrist ligaments. Peripheral detachments and central degenerative tears of the TFCC can be visualized. MRI remains highly operator dependent and technique dependent, and the studies should be interpreted in the context of the findings on physical examination.114

Patients presenting with pain localized to the ulnar side of the wrist often respond to simple splinting and rest. This conservative treatment and NSAIDs can be used effectively while a workup is in progress. A course of rest and splinting, followed by a gradual return to activities, may completely alleviate ulnar-sided symptoms.

Despite the advancements in imaging techniques, there is often no substitute for direct visualization of the ulnocarpal joint or DRUJ or both. Arthroscopy has become an invaluable diagnostic and surgical tool. Tears of the TFCC can be visualized, and their clinical significance better determined. Arthroscopy, done in conjunction with fluoroscopy, can assess for instability of the DRUJ or intercarpal joints or both. Several surgical procedures can now be performed entirely or in part through the arthroscope.115,116

\textbf{Extensor Carpi Ulnaris Tendinitis and Subluxation}

The extensor carpi ulnaris tendon can become irritated with forced pronation/supination activities such as putting topspin on a tennis ball. In severe cases the tendon can begin to sublux around the ulnar head as its restraining dorsal retinaculum becomes increasingly lax. Patients complain of pain with forcible rotation of the forearm, and sometimes there is an associated snapping of the extensor carpi ulnaris tendon. Early treatment consists of immobilization of the wrist and forearm to prevent rotation. Anti-inflammatory medication can help to decrease the inflammation more quickly. After an adequate period of rest, if the acute inflammation resolves, but the extensor carpi ulnaris tendon continues to be unstable, surgery may be indicated to reconstruct or release the sheath at the wrist.

\textbf{Lunotriquetral Ligament Injury}

Tears in the short, stout intraosseous ligament connecting the lunate and the triquetrum are uncommon and often difficult to diagnose. As with the aforementioned diagnoses, patients present with ulnar-sided wrist pain usually worsened by either pronation or supination. Forceful translation of the triquetrum against the lunate causes pain in affected individuals. If diagnosed within 3 to 4 weeks of injury, a short arm cast allows healing and eliminates symptoms. Chronic tears may lead to advanced carpal instability and collapse. MRI or wrist arthroscopy or both may be necessary to make the diagnosis. Treatment is predicated on the staging of instability and ranges from simple casting for acute instability to ligament reconstruction or intercarpal fusion for more advanced cases.

\textbf{Pisotriquetral Arthritis}

Degenerative changes in the pisotriquetral articulation are usually posttraumatic in nature. Patients may recall a fall onto the extended wrist with direct trauma to the ulnar side of the palm. Affected patients present with pain during passive wrist hyperextension and exacerbation with flexion against resistance. With palpation of the pisotriquetral joint, there is tenderness and often crepitus. As with many joints, splinting, NSAIDs, and occasionally injection with corticosteroid and lidocaine are the mainstays of conservative treatment. If this is inadequate to control the symptoms, surgical resection of the pisiform is indicated.

\textbf{Wrist Pain: Radial and Base of Thumb}

\textbf{De Quervain’s Disease and Intersection Syndrome}

One of the most common sites of tendon irritation around the wrist is in the first dorsal extensor compartment, a phenomenon known as de Quervain’s disease. The tendons involved are the extensor pollicis brevis and the abductor pollicis longus. At the level of the radial styloid, these two tendons pass through an osteoligamentous tunnel composed of a shallow groove in the radius and an overlying ligament. Anatomic studies have shown that a high percentage of patients have a divided first dorsal compartment, and this can account for failure of conservative treatment and injections.117-119

Patients with de Quervain’s disease are typically women in their 30s and 40s, although men and women can develop the condition at any age. This is the most common tendinopathy to develop in postpartum women because of the specific hand and wrist position requirements in the care of an infant. Any activity requiring repeated thumb abduction and extension in combination with wrist radial and ulnar deviation can aggravate this problem. Patients complain of pain along the course of these tendons with grasping activities. Clinically, there is tenderness along the affected compartment and there may be swelling over the radial styloid. In severe cases a cracking sound can be elicited with movement of the involved tendons. Finkelstein’s test of forced ulnar deviation of the wrist with the thumb clasped in the fist is pathognomonic of the condition.120,121

A less common condition that may occur in the same general location in the wrist is intersection syndrome. Although initially attributed to friction between the first and second dorsal compartment tendons, Grundberg and Reagan122 subsequently showed that the condition represented a tendinopathy of the radial wrist extensors within the second dorsal compartment.

The primary treatment for de Quervain’s disease and intersection syndrome is rest with splinting. For
de Quervain’s disease, the wrist should be held in slight extension and the thumb abducted in a thumb spica splint to the level of the interphalangeal joint. Immobilization of the wrist alone, in approximately 15 degrees of extension, is usually adequate for intersection syndrome. The addition of a 2- to 4-week course of anti-inflammatory medication also can be helpful. Phonophoresis with a cortisone cream and injection of the compartment with cortisone are second-line treatments if immobilization alone fails to give adequate relief. Injection of corticosteroid into the affected first dorsal compartment is curative for de Quervain’s disease in approximately 75% of patients. Surgery may be indicated for patients who do not respond to a course of conservative treatment including injection. For de Quervain’s disease and intersection syndrome, surgery consists of releasing the stenotic retinacular sheath of the involved compartment.

**Basal Joint Arthropathy**

Inflammation and pain related to the carpometacarpal joint of the thumb are common and can occur at any age. In younger patients, instability secondary to ligamentous laxity is associated with joint subluxation and abnormal cartilage wear and may lead to pain with mechanical activities. In women older than 45 years, studies show 25% have radiographic evidence of degeneration of the basal joint. Patients generally present with pain at the base of the thumb, worsened by pinch and highly dexterous activities. They often report difficulty with tasks such as opening jars and bottles, turning doorknobs and keys, and other activities of daily living. The thumb carpometacarpal joint may be swollen and subluxed and is generally tender to palpation. The joint should be assessed for the presence of increased laxity by manual subluxation of the base of the metacarpal out of the trapezial “saddle” with radial and volar force. With advanced degenerative disease, crepitus is sometimes appreciated.

Radiographs should be obtained to determine the stage of the disease. The addition of a basal joint posteroanterior stress film, in which the patient presses the tips of the thumbs together firmly with the nail plates facing up, is helpful in assessing joint subluxation (Figure 50-8). The most commonly used staging system was developed by Eaton and Glickel and is based on the degree of involvement of the metacarpal joint and whether or not the scaphotrapezial joint is involved. Advancing stages show increased subluxation of the basal joint, with development of joint space narrowing, osteophytes, and subchondral cysts.

Regardless of the stage of the disease, the first line of treatment is immobilization of the thumb metacarpal, leaving the interphalangeal joint free. Splinting has been shown to alleviate the symptoms of carpometacarpal joint inflammation in more than 50% of patients. NSAIDs can be a useful adjunct. Injections of corticosteroid are effective, usually for just a limited time. Although therapy for thenar muscle strengthening has been advocated, especially in early stages, its benefits are minimal and it can occasionally aggravate the problem.

Many patients are able to manage their symptoms with a combination of splinting, medications, corticosteroid injections, and activity modification. The most effective splints are those that are custom made of a moldable plastic material. They may be hand based as shown in Figure 50-9 or forearm based to immobilize the wrist as well. If these various nonoperative treatments are insufficient, surgery may be indicated in young patients to reconstruct the ligaments that stabilize the metacarpal base. In patients with advanced degenerative changes and whose symptoms continue to interfere sufficiently with their daily activities, surgery is indicated to replace the joint with a prosthetic device or to excise the trapezium and reconstruct the soft tissue supports.
Volar Ganglion

Another common location for ganglia is the radial side of the volar wrist. Ganglia typically originate from the scaphotrapezial joint but become superficial and are clinically evident at or near the distal wrist crease over the flexor carpi radialis tendon. Volar ganglia can occur in close proximity to the radial artery and should be differentiated from a radial artery aneurysm. Aspiration, if attempted, should be performed carefully to avoid vascular injury, and surgery should be preceded by performance of an Allen test to document patent ulnar arterial flow. Volar ganglia are associated with a higher recurrence rate and a higher complication rate than their dorsal counterparts.129

Scaphoid Fracture and Nonunion

Occasionally, a young or middle-aged patient presents with a nonunited scaphoid fracture without recollection of a traumatic incident. When evaluating a relatively young patient with pain at the base of the thumb, wrist swelling in the region of the anatomic snuffbox, and a decreased range of motion of the wrist, plain radiographs and a specialized ulnar-deviation ‘navicular’ radiograph should be obtained to rule out scaphoid pathology. In patients in whom a scaphoid nonunion has been present for a significant period, secondary changes in carpal alignment and joint degeneration have usually occurred. Although splint or cast immobilization can be tried, surgical repair of the scaphoid or other wrist salvage procedure is usually required.

Palm

Trigger Finger

Painful clicking and locking of the digits in flexion is one of the most common causes of pain in the hand. This condition, caused by a thickening of the A1 retinacular pulley in the palm, is commonly known as trigger finger. The thumb is the most commonly affected digit, followed by the ring and long fingers.129 Patients may present with isolated activity-related pain in the proximal interphalangeal joint without frank clicking or locking. Early clicking is felt as a snapping sensation during digital motion and is frequently worse on awakening. As the condition progresses, the digital range of motion can be reduced and secondary proximal interphalangeal joint contractures develop. The final stage is a locked trigger finger that cannot be straightened actively.

Primary trigger finger is the most common type, found most often in middle-aged individuals. Triggering of the thumb is four times more frequent in women than in men.5 Secondary triggering is seen in association with such diseases as RA, diabetes, and gout. In this type, trigger fingers are often multiple and can coexist with other stenosing tendinopathies such as de Quervain’s disease or CTS. Congenital or developmental triggering can be identified in children and is much less common. Similar to its presentation in adults, the thumb is most commonly affected, but in contrast to adults, triggering often presents with the interphalangeal joint locked in flexion.

Nonoperative treatment of this condition consists primarily of splinting and local steroid injections. Splinting is most effective at night to prevent the digit from locking. In adults, injection of steroid into the tendon sheath has been shown to be quite effective (Figure 50-10).5,6,130 Injection is used infrequently in infants or children. When nonoperative treatments fail to give lasting relief, surgical treatment consists of longitudinal division of the A1 pulley at the level of the metacarpal head. It is a simple procedure that yields reliable and permanent results with few complications.

Retinacular Cysts

Retinacular ganglion cysts can occur in conjunction with a triggering digit or in isolation. They are located at the base of the digit over the A1 pulley as a discrete, firm, pea-sized nodule. They originate from the flexor tendon sheath or annular pulleys and contain synovial fluid. Patients usually complain of pain with gripping objects or with direct pressure over the cyst. A retinacular cyst is most easily treated initially by needle decompression, with care to avoid injury to the sensory nerves that lie immediately adjacent to the flexor tendon and associated cyst. Approximately 50% recur after aspiration, and surgical resection may be required.

Digits

Mallet Finger

Mallet finger refers to a loss of terminal extension of the distal interphalangeal joint of the digit and can be classified as bony or soft tissue depending on where the disruption in the extensor mechanism occurred. Mallet fingers can occur with minimal trauma such as tucking in bed sheets and may not be recalled by the patient. This sometimes leads to a delay in diagnosis and treatment. When a patient presents with a digit that droops at the distal interphalangeal joint and cannot be actively extended, but has full passive motion, a radiograph should be obtained to determine if there is an associated fracture of the distal phalanx. An extension splint is the treatment of choice for bony and soft
tissue mallet fingers. The distal interphalangeal joint should be held in full extension, and care should be taken not to force the distal interphalangeal joint into hyperextension to prevent dorsal skin ischemia and necrosis. Splinting is employed full time for 6 weeks. The patient should not remove the splint for showering or any other activity but may change the splint carefully for skin care, provided that the joint is maintained in extension throughout. Proximal interphalangeal flexion exercises are initiated from the onset and are important to help reset the tension in the extensor mechanism. Gentle distal interphalangeal flexion exercises are begun at 6 weeks, and splinting is decreased to nighttime between 6 and 8 weeks. Patients usually can expect a small extension lag, on the order of 5 degrees, and a return of most of their flexion.

**Osteoarthritis of the Digits**

Osteoarthritis of the interphalangeal joints is extremely common in older patients and is most often manifested as Heberden’s nodes of the distal interphalangeal joint. Despite gross deformities, pain and dysfunction may be minimal. A mucous cyst may appear in association with degenerative arthritis. Mucous cysts appear on the dorsum of the joint and can cause nail growth deformities owing to pressure on the germinal matrix (Figure 50-11). The changes in nail growth may precede clinical detection of the cyst. These cysts should not be aspirated with a needle because of the close proximity of the distal interphalangeal joint and the risk of secondary joint infection. Treatment consists of distal interphalangeal joint immobilization to control symptoms or surgical excision of the cyst and in particular the underlying osteophytic spurs.

**Tumors**

Benign bone tumors such as simple bone cysts and enchondromas are common in the phalanges. These usually cause no symptoms and frequently are diagnosed as incidental findings on routine hand radiographs. Enchondromas are most commonly located in the metaphysis of the proximal phalanx and may lead to fracture with minimal trauma as a result of weakening of the bone structure. If a pathologic fracture occurs, nonoperative treatment is indicated until the fracture heals. The bone tumor subsequently can be addressed with curettage and bone grafting. Occasionally, because of malalignment, earlier surgical intervention becomes necessary.

Many soft tissue tumors can occur in the hand and digits. Some common benign tumors are giant cell tumors of the tendon sheath, lipomas, and glomus tumors. Lipomas and giant cell tumors of the tendon sheath manifest clinically as painless, slow-growing masses in the palm and digits. Surgical excision is necessary for diagnosis. Glomus tumors arise from the pericytes in the fingertip or subungual area and typically present with intermittent sharp pain in the fingertip. These vascular tumors become intensely symptomatic when the hand is exposed to cold temperatures, owing to abnormal arteriovenous shunting through the hypertrophic glomus system. Surgical excision is generally curative and should be preceded by MRI to rule out multifocal sites.

**Infection**

The most common infection in the hand is the paronychia. It involves the fold of tissue surrounding the fingernail. *Staphylococcus aureus* is the usual pathogen, introduced by a hangnail, a manicure instrument, or nail biting. Patients present with an exquisitely painful and erythematous swelling involving a part of the nail fold. Occasionally, the infection can progress to surround the nail in a horseshoe fashion. Treatment consists of local soaks or surgical drainage. Larger or more chronic infections require surgical drainage.

An infection of the distal pulp of the fingertip, known as a felon, is a particular problem in diabetic patients. This infection differs from other subcutaneous infections because of the vertical fibrous septa that divide and stabilize the pulp of the fingertip. Often patients have had some recent penetrating injury in the area. Because of the tightly constrained area of the infection, patients present with an intensely painful fingertip. There may be an area of “pointing” over the abscess. Surgical drainage is required followed by soaks and oral antibiotics, and intravenous antibiotics are generally recommended in diabetic patients.

Although similar in appearance to a paronychia, herpetic whitlow is caused by herpes simplex virus and must be differentiated from other fingertip infections because of a radically different treatment protocol. Whitlow was common among dental hygienists before the widespread use of gloves for all health care workers. As with bacterial infections, the area becomes painful and erythematous; local tenderness is much less severe, however. Diagnosis is by clinical presentation and history. If seen early, vesicles can be ruptured for fluid analysis and viral culture. Nonoperative treatment with oral antiviral agents is recommended.

---

**Figure 50-11** Dorsal view of a digit with an as yet clinically inapparent mucous cyst and the corresponding groove deformity of the nail plate.
Other hand and digit infections such as suppurrative flexor tenosynovitis, deep space infections of the palm, pyogenic arthritis, infections from bite wounds, and osteomyelitis should be evaluated initially with radiographs of the hand and appropriate blood work. If possible, antibiotics should be withheld until definitive cultures are obtained from the affected area. Antibiotics should be administered intravenously, and the hand and wrist should be immobilized. Most infections of this nature require surgical drainage for definitive treatment.

Selected References

99. Chuinard RG, Zeman SC: Kienbock's disease: an analysis and ratio-
100. Mino DE, Palmer AK, Levinsohn EM: The role of radiography and computerized
101. Mino DE, Palmer AK, Levinsohn EM: Radiography and computerized
104. Finkelstein H: Stenosing tendovaginitis at the radial styloid process, J Bone Joint Surg
106. Mino DE, Palmer AK, Levinsohn EM: The role of radiography and computerized
107. Mino DE, Palmer AK, Levinsohn EM: Radiography and computerized
References

43. Phalen GS: Spontaneous compression of the median nerve at the wrist, JAMA 145:1128, 1951.
Temporomandibular Joint Pain

DANIEL M. LASKIN

KEY POINTS

Temporomandibular joint (TMJ) pain must be distinguished from the pain that more commonly arises from the muscles of mastication (myofascial pain), which can produce similar signs and symptoms.

TMJ pain also must be distinguished from pain coming from the ear or parotid gland.

TMJ pain and masticatory muscle pain generally are accompanied by limitation of mouth opening, but not pain arising from the ear or parotid gland.

Most major systemic arthropathies may also involve the TMJ and thereby give rise to pain and limited jaw movement.

Displacement of the intra-articular disk in the TMJ produces pain that is accompanied by a clicking or popping sound or sudden onset of jaw locking.

Pain in the temporomandibular joint (TMJ) region, a commonly encountered symptom, affects more than 10 million Americans. Because of its diverse causes, however, considerable difficulty is often involved in proper diagnosis and treatment. Owing to the proximity of the ear and parotid gland and the similar nature of pain in these areas, pathologic conditions involving these structures are often confused with conditions arising in the TMJ. Pain occurring in the adjacent muscles of mastication, also a frequently encountered situation, not only is similar to TMJ pain in character and location, but also is associated with jaw dysfunction, a common finding with painful conditions directly involving the TMJ. For these reasons, knowledge of the various painful conditions occurring in the TMJ region is essential in establishing a correct diagnosis.

Because patients with primary TMJ disease often have secondary myofascial pain in the muscles of mastication, and because patients with primary myofascial pain problems in the masticatory muscles can develop secondary TMJ disease, the generally accepted term used to describe this overlapping group of conditions is temporomandibular disorders. These conditions are subdivided for purposes of diagnosis and treatment into conditions that primarily involve the TMJ (TMJ problems) and conditions that primarily involve the muscles of mastication (myofascial pain and dysfunction [MPD], masticatory myalgia). From a diagnostic standpoint, it is important to consider the numerous conditions that mimic the temporomandibular disorders or MPD by producing similar signs and symptoms (Tables 51-1 and 51-2).

Table 51-3 lists the various pathologic entities that commonly involve the TMJ. Although a variety of conditions are known, only three types are considered to generally produce pain: the various arthritides, derangements of the intra-articular disk, and certain neoplasms.

ARTHRITIS OF THE TEMPOROMANDIBULAR JOINT

Arthritis is the most common painful condition affecting the TMJ. Although osteoarthritis and rheumatoid arthritis are encountered most frequently, cases of infectious arthritis, metabolic arthritis, and presentation as part of the spondyloarthropathies are also seen in practice. Traumatic arthritis is another relatively common occurrence.

Osteoarthritis

Osteoarthritis is the most common type of arthritis involving the TMJ and the most frequent cause of pain in that region. Clinical symptoms of the disease have been reported in 16% of the general population, but radiographic features have been found in 44% of asymptomatic individuals. Although the TMJ is not a weight-bearing joint in the same sense as the joints of the long bones, the stresses associated with such parafunctional habits as clenching and grinding of the teeth are sufficient to contribute to similar degenerative changes in some patients. Acute and chronic trauma and derangements of the intra-articular disk also are common causes of secondary degenerative arthritis.

Clinical Findings

Primary osteoarthritis, which usually is seen in older individuals, is insidious in its onset; it generally produces only mild discomfort, and individuals rarely complain about the condition. Secondary osteoarthritis usually occurs in younger patients (20 to 40 years old) and tends to be painful. In contrast to primary degenerative joint disease and rheumatoid arthritis, it often is limited to only one TMJ, although it may become bilateral in the late stages, and involvement of other joints is uncommon. The condition is characterized by TMJ pain that is increased by function, joint tenderness, limitation of mouth opening, and occasional clicking and popping sounds. In the late stages, crepitation may be noted in the joint.

Imaging Findings

The earliest radiologic feature of osteoarthritis of the TMJ, whether primary or secondary, is subchondral sclerosis in the mandibular condyle. If the condition progresses,
Table 51-1  Differential Diagnosis of Nonarticular Conditions Mimicking Temporomandibular Joint Pain or Myofascial Pain in the Masticatory Muscles

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Jaw Limitation</th>
<th>Muscle Tenderness</th>
<th>Diagnostic Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulpitis</td>
<td>No</td>
<td>No</td>
<td>Mild to severe ache or throbbing; intermittent or constant; aggravated by thermal change; eliminated by dental anesthesia; positive radiographic findings</td>
</tr>
<tr>
<td>Pericoronitis</td>
<td>Yes</td>
<td>Possible</td>
<td>Persistent mild to severe ache; difficulty swallowing; possible fever; local inflammation; relieved with dental anesthesia</td>
</tr>
<tr>
<td>Otitis media</td>
<td>No</td>
<td>No</td>
<td>Moderate to severe earache; constant pain; fever; usually history of upper respiratory infection; no temporomandibular joint tenderness</td>
</tr>
<tr>
<td>Parotitis</td>
<td>Yes</td>
<td>No</td>
<td>Constant aching pain, worse when eating; pressure feeling; absent salivary flow; ear lobe elevated; suppuration from duct</td>
</tr>
<tr>
<td>Sinusitis</td>
<td>No</td>
<td>No</td>
<td>Constant aching or throbbing; worse with change of head position; nasal discharge; often maxillary molar pain not relieved by dental anesthesia</td>
</tr>
<tr>
<td>Trigeminal neuralgia</td>
<td>No</td>
<td>No</td>
<td>Sharp stabbing pain of short duration; trigger zone; pain follows nerve pathway; older age group; often relieved by dental anesthesia</td>
</tr>
<tr>
<td>Atypical (vascular) neuralgia</td>
<td>No</td>
<td>No</td>
<td>Diffuse throbbing or burning pain of long duration; often associated autonomic symptoms; no relief with dental anesthesia</td>
</tr>
<tr>
<td>Temporal arteritis</td>
<td>No</td>
<td>No</td>
<td>Constant throbbing preauricular pain; artery prominent and tender; low-grade fever; may have visual problems; elevated erythrocyte sedimentation rate</td>
</tr>
<tr>
<td>Trotter’s syndrome</td>
<td>Yes</td>
<td>No</td>
<td>Aching pain in ear, side of face, and lower jaw; deafness; nasal obstruction; cervical lymphadenopathy</td>
</tr>
<tr>
<td>Eagle’s syndrome</td>
<td>No</td>
<td>No</td>
<td>Mild to sharp stabbing pain in ear, throat, and retromandible; provoked by swallowing, turning head, carotid compression; usually post tonsillectomy; styloid process &gt;2.5 cm</td>
</tr>
</tbody>
</table>


Condylar flattening and marginal lipping may be noted. In the later stages, erosion of the cortical plate, osteophyte formation, or both may occur. Breakdown of the subcortical bone occasionally may result in the formation of bone cysts. Although changes in the articular fossa generally are not as severe as changes in the condyle, cortical erosion sometimes can be seen. Narrowing of the joint space also occurs in the late stages; this is indicative of concomitant degenerative changes in the intra-articular disk. Although changes in the TMJ usually can be seen on plain radiographs, sagittal and coronal computed tomography (CT) scans are the preferred modality for imaging the bony structures.

Table 51-2  Differential Diagnosis of Nonarticular Conditions Producing Limitation of Mandibular Movement

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Jaw Limitation</th>
<th>Muscle Tenderness</th>
<th>Diagnostic Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odontogenic infection</td>
<td>Yes</td>
<td>Yes</td>
<td>Fever; swelling; positive radiographic findings; tooth tender to percussion; pain relieved and movement improved with dental anesthesia</td>
</tr>
<tr>
<td>Nonodontogenic infection</td>
<td>Yes</td>
<td>Yes</td>
<td>Fever; swelling; negative dental findings on radiograph; dental anesthesia may not relieve pain or improve jaw movement</td>
</tr>
<tr>
<td>Myositis</td>
<td>Yes</td>
<td>Yes</td>
<td>Sudden onset; jaw movement associated with pain; areas of muscle tenderness; usually no fever</td>
</tr>
<tr>
<td>Myositis ossificans</td>
<td>No</td>
<td>No</td>
<td>Palpable nodules seen as radiopaque areas on radiograph; involvement of nonmasticatory muscles</td>
</tr>
<tr>
<td>Neoplasia</td>
<td>Possible</td>
<td>Possible</td>
<td>Palpable mass; regional nodes may be enlarged; may have paresthesia; radiograph may show bone involvement</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>No</td>
<td>No</td>
<td>Skin hard and atrophic; mask-like facies; paresthesias; arthritic joint pain; widening of periodontal ligament</td>
</tr>
<tr>
<td>Hysteria</td>
<td>No</td>
<td>No</td>
<td>Sudden onset after psychological trauma; no physical findings; jaw opens easily under general anesthesia</td>
</tr>
<tr>
<td>Tetanus</td>
<td>Yes</td>
<td>No</td>
<td>Recent wound; stiffness of neck; difficulty swallowing; spasm of facial muscles; headache</td>
</tr>
<tr>
<td>Extrapyramidal reaction</td>
<td>No</td>
<td>No</td>
<td>Patient on antipsychotic drug or phenothiazine tranquilizer; hypertonic movement; lip smacking; spontaneous chewing motions</td>
</tr>
<tr>
<td>Depressed zygomatic arch</td>
<td>Possible</td>
<td>No</td>
<td>History of trauma; facial depression; positive radiographic findings</td>
</tr>
<tr>
<td>Osteochondroma coronoid</td>
<td>No</td>
<td>No</td>
<td>Gradual limitation; jaw may deviate to unaffected side; possible clicking sound on jaw movement; positive radiographic findings</td>
</tr>
</tbody>
</table>

Diagnosis

The diagnosis of osteoarthritis is based on the patient’s history and clinical and radiographic findings. A history of trauma or parafunctional oral habits is often reported. Involvement is generally unilateral, and no significant changes are observed in any of the other joints. The pain tends to be well localized, and the TMJ is often tender to palpation.

Treatment

Treatment of degenerative arthritis of the TMJ is usually medical, as in other joints of the body. It involves the use of nonsteroidal anti-inflammatory drugs, application of heat, eating a soft diet, limitation of jaw function, and use of a bite appliance to control parafunction if the patient has a chronic habit of clenching or grinding the teeth. Arthrocentesis has also been shown to be helpful.4,5 Physical therapy with thermal agents, ultrasound, and iontophoresis also can be beneficial, and isotonic and isometric exercises are used to improve joint stability after acute symptoms have subsided. The use of intra-articular steroid injections is controversial; they should be used only in patients with acute symptoms that do not respond to other forms of medical management. Because of the potentially damaging effects of long-acting steroids,4,6 they should be limited to no more than three or four single injections given at 3-month intervals. Intra-articular injection of high-molecular-weight sodium hyaluronate given twice, 2 weeks apart, has been shown to have essentially the same therapeutic effect as a steroid injection, without the potential adverse effects.5,7

When the acute symptoms have been controlled, therapy is directed toward control of factors possibly contributing to the degenerative process. Unfavorable loading of the joint is eliminated by replacement of missing teeth to establish a good, functional occlusion; by correction of any severe dental malrelationships through orthodontics or orthognathic surgery; and by continued use of a bite appliance at night to control teeth-clenching or teeth-grinding habits.5,8

In patients in whom medical management for 3 to 6 months fails to relieve the symptoms, surgical management may be indicated. Surgery involves removal of only the minimal amount of bone necessary to produce a smooth articular surface. Unnecessary removal of the entire cortical plate, as occurs with the so-called condylar shave procedure or high condylotomy, can lead to continuation of the resorptive process in some instances, and should be avoided if possible.

Rheumatoid Arthritis

More than 50% of patients with rheumatoid arthritis have involvement of the TMJ.3 Although the TMJ may be affected early in the course of the disease, other joints in the body usually are involved first. The general female-to-male ratio is 3:1. TMJ involvement may also characterize juvenile inflammatory arthritis. In children, destruction of the mandibular condyle by the disease process results in growth retardation and facial deformity characterized by a severely retruded chin. Fibrous or bony ankylosis is a possible sequel at all ages.

Clinical Findings

Patients with rheumatoid arthritis of the TMJ have bilateral pain, tenderness, swelling in the preauricular region, and limitation of mandibular movement. These symptoms are characterized by periods of exacerbation and remission. Joint stiffness and pain are usually worse in the morning and decrease during the day. The limitation in mandibular movement worsens as the disease progresses; the patient also may develop an anterior open bite.

Imaging Findings

Although radiographic changes may not be noted in the early stages of the disease, about 50% to 80% of patients show bilateral evidence of demineralization, condylar flattening, and bone erosion as the disease progresses, so the articular surface appears irregular and ragged. Erosion of the glenoid fossa also is seen sometimes. Narrowing of the joint space is caused by destruction of the intra-articular disk. With continued destruction of the condyle, loss of ramus height can lead to contact of only the posterior teeth and an anterior open bite.

Diagnosis

Rheumatoid arthritis is diagnosed on the basis of the history, clinical and radiographic findings, and confirmatory laboratory tests. Distinguishing features for rheumatoid arthritis and degenerative arthritis of the TMJ are shown in Table 51-3.

Treatment

Treatment of rheumatoid arthritis of the TMJ is similar to that provided for other joints.7,10 Anti-inflammatory drugs are used during the acute phases, and mild jaw exercises are used to prevent excessive loss of motion when acute symptoms subside. In severe cases, disease-modifying drugs, such as methotrexate, and biologic agents, including etanercept, infliximab, adalimumab, certolizumab, golimumab, abatacept, tocilizumab, and rituximab, may be used pending systemic presentation. Orthognathic surgery may be necessary in patients with an anterior open bite after the disease goes into remission, or in patients in whom ankylosis develops after that condition is corrected.

Spondyloarthropathies

In addition to the adult and juvenile forms of rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, and reactive arthritis also can involve the TMJ.6-13

Psoriatic Arthritis

Psoriatic arthritis occurs in approximately one-third of patients who have cutaneous psoriasis. It can have a sudden onset, can be episodic in nature, and may show spontaneous
Table 51-3  Differential Diagnosis of Temporomandibular Joint (TMJ) Diseases

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Jaw Limitation</th>
<th>Muscle Tenderness</th>
<th>Diagnostic Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agenesis</td>
<td>No</td>
<td>Yes</td>
<td>Congenital; usually unilateral; mandible deviates to affected side; unaffected side long and flat; severe malocclusion; often ear abnormalities; radiograph shows condylar deficiency</td>
</tr>
<tr>
<td>Condylar hypoplasia</td>
<td>No</td>
<td>No</td>
<td>Congenital or acquired; affected side has short mandibular body and ramus, fullness of face; deviation of chin; body of mandible elongated and face flat on unaffected side; malocclusion; radiograph shows condylar deformity, antegonial notchin</td>
</tr>
<tr>
<td>Condylar hyperplasia</td>
<td>No</td>
<td>No</td>
<td>Facial asymmetry with deviation of chin to unaffected side; cross-bite malocclusion; prognathic appearance; lower border of mandible often convex on affected side; radiograph shows symmetric enlargement of condyle</td>
</tr>
<tr>
<td>Neoplasia</td>
<td>Possible</td>
<td>Yes</td>
<td>Mandible may deviate to affected side; radiographs show enlarged, irregularly shaped condyle or bone destruction, depending on type of tumor; unilateral condition</td>
</tr>
<tr>
<td>Infectious arthritis</td>
<td>Yes</td>
<td>No</td>
<td>Signs of infection; may be part of systemic disease; radiograph may be normal early, later can show bone destruction; fluctuance may be present; pus may be obtained on aspiration; usually unilateral</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Yes</td>
<td>Yes</td>
<td>Signs of inflammation; findings in other joints (hands, wrists, feet, elbows, ankles); positive laboratory test results; retarded mandibular growth in children; anterior open bite; radiograph shows bone destruction; usually bilateral</td>
</tr>
<tr>
<td>Spondyloarthropathies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
<td>Yes</td>
<td>Yes</td>
<td>Presence of cutaneous psoriasis; nail dystrophy; involvement of distal interphalangeal joints; radiograph shows condylar erosion; negative for rheumatoid factor</td>
</tr>
<tr>
<td>Ankylosing spondylitis</td>
<td>Yes</td>
<td>Yes</td>
<td>Frequent involvement of the spine and sacroiliac joint; extra-articular manifestations of spondylitis include iritis, anterior uveitis, aortic insufficiency, and conduction defects; erosive condylar changes; TMJ ankylosis may occur</td>
</tr>
<tr>
<td>Metabolic arthritis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gout</td>
<td>Yes</td>
<td>Yes</td>
<td>Usually sudden onset; often monoarticular; commonly involves great toe, ankle, and wrist; joint swollen, red, and tender; increased serum uric acid; late radiographic changes</td>
</tr>
<tr>
<td>Pseudogout</td>
<td>Yes</td>
<td>Yes</td>
<td>Generally unilateral; TMJ may be only joint involved; joint frequently swollen; presence of intra-articular calcification; may be a history of trauma</td>
</tr>
<tr>
<td>Traumatic arthritis</td>
<td>Yes</td>
<td>Yes</td>
<td>History of trauma; radiograph normal except for possible widening of joint space; local tenderness; usually unilateral</td>
</tr>
<tr>
<td>Degenerative arthritis</td>
<td>Yes</td>
<td>Yes</td>
<td>Unilateral joint tenderness; often crepitus; TMJ may be only joint involved; radiograph may be normal or show condylar flattening, lipping, spurring, or erosion</td>
</tr>
<tr>
<td>Ankylosis</td>
<td>No</td>
<td>Yes</td>
<td>Usually unilateral, but can be bilateral; may be history of trauma; young patient may show retarded mandibular growth; radiographs show loss of normal joint architecture</td>
</tr>
<tr>
<td>Internal disk degeneration</td>
<td>Yes</td>
<td>Yes</td>
<td>Pain exacerbated by function; clicking on opening or opening limited to &lt;25 mm with no click; positive magnetic resonance imaging findings; may be history of trauma; usually unilateral</td>
</tr>
</tbody>
</table>


remission.\(^9,12\) Often only one TMJ is involved. Symptoms include TMJ pain and tenderness, restricted jaw movement, and crepititation, mimicking the symptoms of rheumatoid arthritis.\(^9,12\) Radiographic changes are nonspecific and cannot be distinguished easily from those of other types of arthritis, particularly rheumatoid arthritis and ankylosing spondylitis.\(^9,12\) They usually involve erosive changes in the condyle and glenoid fossa associated with extreme narrowing of the joint space.\(^11,15,16\) In severe cases, ankylosis may develop, reflected occasionally in new bone formation at earlier stages.\(^9,12,17\)

The diagnosis usually is based on the triad of psoriasis, radiographic evidence of erosive arthritis, and a negative serologic test for rheumatoid factor. Even in the presence of a rash, however, the diagnosis cannot be absolutely confirmed. The differential diagnosis always should include rheumatoid arthritis, reactive arthritis, ankylosing spondylitis, and gout.

Treatment of psoriatic arthritis of the TMJ is similar to that of rheumatoid arthritis, and is driven essentially by the imperative to treat the systemic inflammatory disease process.\(^13,18,21\) Surgery is necessary if ankylosis occurs.

**Ankylosing Spondylitis**

About one-third of patients with ankylosing spondylitis develop TMJ involvement several years after onset of the disease. Pain and limitation of jaw movement are the most common symptoms, and ankylosis can develop in advanced cases.\(^9,11,14,22\) On radiographic examination, about 30% of patients show erosive changes in the condyle and fossa and narrowing of the joint space.\(^15,23\) In long-standing cases, a
more florid osteophytic response is sometimes seen during quiescent periods. The severity of the changes seems to be related to the severity of the disease. Treatment of ankylosing spondylitis of the TMJ is generally medical and is part of the total management of the patient. Physical therapy is used to improve jaw mobility, and bite appliances are used, when indicated, to reduce parafunctional stress on the joint. If ankylosis develops, surgery is the treatment of choice.24

**Reactive Arthritis**

Reactive arthritis of the temporomandibular joint is more common in males than in females. It is characterized by recurrent pain, swelling, and limitation of mouth opening.25 Radiographically, condylar erosion may be evident.26 Treatment is similar to that of the other seronegative spondyloarthropathies, consisting of nonsteroidal anti-inflammatory drugs, intra-articular steroids, and disease-modifying drugs. If a specific triggering bacterial infection can be identified, an appropriate antibiotic should be prescribed.

**Traumatic Arthritis**

Acute trauma to the mandible that does not result in a fracture can still produce injury to the TMJ. When this occurs in a child, it is essential to warn the parents about the possibility of future retardation of mandibular growth and associated facial deformity resulting from damage to the articular cartilage, which is an important growth site.16,27

Traumatic arthritis is characterized by TMJ pain and tenderness and limitation of jaw movement. The resultant inflammation and occasional hemarthrosis can lead to loss of tooth contact on the affected side. Frequently, bruises or lacerations are apparent at the site of the initial injury. No radiographic changes may be seen, or widening of the joint space may be produced by intra-articular edema or hemorrhage. In some instances, radiographs may show an intra-capsular fracture that was not recognized on clinical examination.

Treatment of traumatic arthritis consists of the use of nonsteroidal anti-inflammatory drugs, application of heat, a soft diet, and initial restriction of jaw movement. When acute symptoms subside, range-of-motion exercises should be used to avoid fibrous ankylosis.

**Infectious Arthritis**

Infectious arthritis rarely involves the TMJ. Although it can affect the joint as part of such systemic diseases as gonorrhea, syphilis, tuberculosis, and Lyme disease,15,18,22,29 the most common way is by direct extension of an adjacent infection of dental, parotid gland, or otic origin.19,30 Occasionally, it also may occur from localization of blood-borne organisms in the joint after a traumatic injury or by direct involvement through a penetrating wound.20,30 The most common pathogens are Staphylococcus aureus, Haemophilus influenzae, and Streptococcus species.31

**Clinical Findings**

Infectious arthritis generally results in unilateral pain, tenderness, swelling, and redness in the region of the TMJ. Chills, fever, sweating, and systemic findings characteristic of the specific type of infection also are present. Often the teeth cannot be occluded because of swelling within the joint. In pyogenic forms of infectious arthritis, fluctuation may be noted in the joint region. Patients with Lyme disease show characteristic skin lesions and often positive serology.18,29

**Imaging Findings**

Radiographic findings are usually normal in early stages of the disease because of lack of bony involvement, but the intra-articular accumulation of pus or inflammatory exudate may cause separation of articulating surfaces, which can be detected on magnetic resonance imaging (MRI). Later, depending on the severity and chronicity of the infection, varying degrees of bony destruction, ranging from damage to the articular surface of the mandibular condyle to extensive osteomyelitis, may be seen. In the late stages, fibrous or bony ankylosis may occur. In children, infectious arthritis can affect growth of the condyle, resulting in facial asymmetry.

**Treatment**

Treatment of infectious arthritis includes the use of appropriate antibiotics, proper hydration, control of pain, and limitation of jaw movement. Arthrocentesis with Ringer’s solution one to three times weekly until acute symptoms subside has also been recommended.31 Suppurative infections may require aspiration, incision, and drainage, or sequestrectomy. When bone loss has been extensive, reconstructive procedures may be necessary. In children in whom mandibular growth has been affected, a costochondral graft can be used to correct facial asymmetry and re-establish growth of the mandible.

**Metabolic Arthritis**

Metabolic arthritis, which can accompany gout or pseudogout (calcium pyrophosphate dehydrate arthropathy), is rare in the TMJ.21,33

**Gout**

Gouty arthritis of the TMJ occurs most frequently in men older than 40 years and usually is preceded by involvement of one or more joints of the feet or hands. The attack usually occurs suddenly, and the joint becomes swollen, painful, red, and tender. Recovery may occur in a few days, and remission can last for months to years.

When attacks are infrequent, radiographic changes may not be noted for a long time. Because so few cases have been reported, the precise radiographic changes that occur have not been well documented. Calcified areas in the disk, destruction of the hard tissues of the joint, condylar exostoses and spurring, and the presence of tophi have been described.21,31-35 The initial approach to treatment of gout involving the TMJ is medical. If symptoms are not controlled, however, surgical debridement of the joint and arthroplasty may be indicated.
**Pseudogout**

Calcium pyrophosphate dehydrate arthropathy (pseudogout) in the TMJ clinically mimics gout, and the mandibular condyle may show degenerative and erosive changes radiographically. In the primary form, which usually is seen in older patients, intra-articular calcification is noted (chondrocalcinosis), and diffuse calcification occurs in the intra-articular disk. Similar changes are seen in the secondary form, but it occurs in younger patients and frequently is preceded by a history of trauma. Just as in gout of the TMJ, the initial treatment of pseudogout is medical, and surgery is reserved for patients in whom such treatment is ineffective.

**INTERNAL DERANGEMENTS**

Internal derangements are a common cause of pain in the TMJ. They represent a disturbance in the normal anatomic relationship between the intra-articular disk and the condyle, resulting in interference with the smooth movement of the joint.

**Clinical Findings**

Three stages of internal derangement have been identified: (1) a painless incoordination phase, in which a momentary catching sensation is felt during mouth opening; (2) anterior disk displacement with reduction into the normal position during mouth opening, which is characterized by a clicking or popping sound (Figure 51-1); and (3) anterior disk displacement without reduction on attempted mouth opening, which is characterized by restriction of jaw movement, or locking (Figure 51-2). Joint pain in patients with anterior disk displacement, with or without reduction, is caused by condylar compression of the highly innervated retrodiskal tissue that occupies the glenoid fossa as the intra-articular disk assumes a more forward position, and by the accompanying inflammation.

**Etiology**

The three main causes of internal derangement of the intra-articular disk are trauma, abnormal functional loading of the joint, and degenerative joint disease. It has been suggested that spasm in the lateral pterygoid muscle, a portion of which attaches to the anterior aspect of the disk, can lead to a disk derangement, but evidence for this theory is circumstantial. Although some clinicians believe that occlusal factors also play a role in causing internal derangements, no conclusive studies have shown such a relationship.

Acute macrotrauma is probably the most common cause of internal derangement. Among the incidents that have been implicated are a blow to the jaw, endotracheal intubation, cervical traction, and iatrogenic stretching of the joint during dental or oral surgical procedures. Although whiplash injuries frequently have been implicated in the development of internal derangement, a study of 155 patients with this type of injury showed that only 1 developed clicking in the TMJ immediately after the automobile accident. At 1 month of follow-up, two additional patients of the 129 contacted experienced clicking, but at 1 year, no additional patients of the 104 contacted had developed clicking. Although internal derangements of the TMJ can be caused by a whiplash injury, the incidence seems to be low.
Whether a patient merely develops alterations in the articular surface leading to a catching or binding sensation, anterior disk displacement with reduction on mouth opening (clicking or popping), or anterior disk displacement without reduction during mouth opening (locking) after trauma to the TMJ depends on the severity of the injury. Although associated traumatic arthritis causes pain during function in each of these instances, the pain is more severe in the last two conditions because of compression of retrodiskal tissue, which is now located in the articular zone.

Functional overloading of the TMJ associated with the habit of chronic teeth clenching is another frequent cause of internal derangements. Although the TMJ is constructed for eccentric movements, it is not constructed for the constant isometric loading and unloading that occurs during this activity. Such parafunction affects the lubrication of the joint and alters the articular surfaces, introducing friction between the disk and the condyle that leads to degenerative changes in the articular surfaces and results in gradual anterior displacement of the disk.26,28,40,42

Degenerative joint disease may precede the development of an internal derangement, or it may occur after the development of an internal derangement. In the first instance, changes in the character of the articulating surfaces result in an inability of the parts to glide smoothly over each other, gradually leading to forward displacement of the disk, which normally rotates posteriorly during mouth opening. In the second instance, the displaced disk results in an altered relationship between articulating components of the joint, which leads to degenerative changes in these structures. In patients in whom the condition causing the degenerative joint disease is still active, whether primarily or secondarily, the condition and the disk derangement must be treated for the problem to be resolved completely.

**Imaging Findings**

Depending on the cause of the internal derangement and its duration, radiographs may or may not show any evidence of degenerative joint disease. Magnetic resonance imaging shows anterior disk displacement in the closed mouth position, however, as well as a return to a normal disk relationship during mouth opening in patients with clicking and popping; in patients with locking, the disk remains in the anterior position on attempted mouth opening, and movement of the condyle is limited. A small group of patients with locking show the intra-articular disk in normal position when the teeth are in occlusion, rather than anteriorly displaced, and no change in disk position occurs when the patient attempts to open the mouth.29,43 In such cases, adhesion of the disk to the articular eminence prevents translation of the condyle. These patients differ from those with anteriorly displaced, nonreducing disks in that they do not have a history of TMJ clicking preceding the sudden onset of locking.

**Treatment**

Initial treatment of patients with painful clicking or popping in the TMJ consists of a nonsteroidal anti-inflammatory drug; a soft, nonchewy diet; and use of a bite-opening appliance to reduce compression of retrodiskal tissue (Figure 51-3). A muscle relaxant drug can be added to the regimen if the patient has associated myofacial pain. When the pain has stopped, no further treatment is necessary, although joint noise still may be present. A long-term follow-up study (1 to 15 years) of 190 patients with a history of clicking treated by such conservative nonsurgical modalities, which are not directed specifically to the problems of joint noise or disk displacement, showed that the condition worsened in only 1%, indicating that it is permissible to observe individuals with painless clicking as long as they remain otherwise asymptomatic.40,44 However, in those patients who have teeth-clenching and -grinding habits, use of a bite appliance is indicated to control these habits.

In patients with pain and clicking in the TMJ that is unresponsive to nonsurgical management, the disk should be repositioned arthroscopically or by open surgery (diskoplasty). Patients with parafunctional habits should continue the use of a bite appliance when sleeping. In patients with locking (anterior disk displacement without reduction), whether painful or not, treatment is urgent because if the condition is left untreated for a long time, subsequent management can be complicated by further degenerative changes in the disk and condyle that make disk salvage (diskoplasty) impossible. Initial treatment involves joint lavage and lysis of adhesions arthroscopically or by arthrocentesis. The latter involves establishment of inlet and outlet portals in the upper joint space with hypodermic needles, irrigation with lactated Ringer’s solution to remove inflammatory tissue breakdown products and cytokines, and lysis of adhesions by hydraulic distention and manual manipulation of the joint (Figure 51-4).31,45

The results of arthrocentesis parallel those achieved with arthroscopic lysis and lavage, and the procedure is less invasive. Although neither of these procedures restores the disk to its normal position, they do restore disk and joint mobility, and they reduce pain and improve function in most
tumor, and synovial hemangioma have been reported. Malignant tumors of the TMJ are even rarer, with infrequent reports of fibrosarcoma, chondrosarcoma, synovial fibrosarcoma, osteosarcoma, malignant fibrous histiocytoma, malignant schwannoma, leiomyosarcoma, and multiple myeloma. The TMJ also can be invaded by neoplasms from the cheek, the parotid gland, the external auditory canal, and the adjacent ramus of the mandible. Metastasis to the condyle from distant neoplasms in the breast, lung, prostate, colon, thyroid gland, liver, stomach, and kidney has been described.

Tumors of the TMJ can cause pain, limitation of jaw movement, deviation of the mandible to the affected side on attempted mouth opening, and difficulty in occluding the teeth. Depending on the nature of the condition, radiographs may show bony deformation, apposition, or resorption. A biopsy is necessary to establish the definitive diagnosis.

MYOFASCIAL PAIN AND DYSFUNCTION

Myofascial pain and dysfunction (MPD), or masticatory myalgia, is considered to be a psychophysiologic disease that primarily involves the muscles of mastication, and not the TMJ. Women are affected more frequently than men; the ratio in various reports ranges from 3:1 to 5:1. Although the condition can occur in children, the incidence seems to be greatest in adults 20 to 40 years old. MPD frequently is confused with painful conditions affecting the TMJ, such as degenerative arthritis or internal derangements, because patients with primary MPD can develop these diseases secondarily, and patients with primary joint disease can develop secondary MPD. Enhanced understanding of the causes and pathogenesis of this condition makes its diagnosis easier and its treatment more effective.

Etiology

Psychological stress has been suggested as an important contributing factor in the development of MPD (psychophysiologic theory). It is hypothesized that in most patients, stress-related, centrally induced increases in muscle activity, frequently combined with the presence of parafunctional habits such as clenching or grinding of the teeth, may result in associated muscle fatigue, pain, and limited mouth opening. However, similar symptoms occasionally have been seen to result from muscle overextension, muscle overcontraction, or trauma (Figure 51-5). A counter-theory (the pain adaptation theory of Lund) has been proposed to suggest that pain in the masticatory muscles leads to a reduction rather than an increase in muscle activity as a protective mechanism, and this causes the limitation in mouth opening; however, this theory does not explain the origin of the pain. Despite extensive research, the cause of myofascial pain and dysfunction remains unknown.

Clinical Findings

Pain of unilateral origin is the most common symptom of MPD. In contrast to the pain associated with joint disease, which is well localized, the pain of muscle origin is more
diffuse. The patient generally is unable to identify accurately the specific site involved; this can serve as an important diagnostic criterion in distinguishing between muscle and joint disorders.

Depending on the muscle involved, the pain associated with MPD may be described by the patient in various ways. The masseter is the muscle most frequently involved, and the patient usually refers to the pain as a jaw ache. The temporalis is the next most commonly involved muscle; it produces pain on the side of the head, which is interpreted by the patient as a headache. Involvement of the lateral pterygoid muscle produces earache or a deep pain behind the ear, whereas medial pterygoid involvement causes discomfort on swallowing and the feeling of a painful, swollen gland beneath the angle of the mandible. Medial pterygoid involvement also can cause stuffiness or a full feeling in the ear.

The pain associated with MPD is usually constant, but it is often more severe on arising in the morning or may worsen gradually as the day progresses. Pain generally is exacerbated by jaw function, especially during such activities as eating and excessive talking. Myofascial pain tends to be regional rather than local, and patients with a long-standing problem may complain that pain in the facial region has spread to the cervical area and later to the shoulders and back.

Tenderness in the muscles of mastication, another common finding, can be used to confirm the source of the pain in muscles that are accessible to palpation (masseter, temporalis, and medial pterygoid). Although muscle tenderness usually is not reported by the patient, this symptom can be elicited easily by the examiner. The most frequent sites of tenderness are near the angle of the mandible, in the belly and the posterosuperior aspect of the masseter, in the anterior temporal region, and over the temporal crest on the anterior aspect of the coronoid process. The location of some of the tender areas suggests that tendons may be a source of pain and tenderness.

Limitation of mandibular movement is the third cardinal symptom of MPD. It manifests as an inability to open the mouth as wide as usual and as a deviation of the mandible to the affected side when mouth opening is attempted. Lateral excursion to the unaffected side is reduced. The limitation of mandibular movement usually is correlated with the amount of pain present.

A clicking or popping sound in the TMJ is another finding in some patients with MPD. This is not a cardinal sign, however, because it occurs only in patients with a chronic teeth-clenching habit, which gradually produces frictional changes in the joint and subsequent disk displacement.\(^\text{26,40}\) The presence of joint sounds alone is insufficient to allow a diagnosis of MPD. Joint sounds must be accompanied by myofascial pain and tenderness in the masticatory muscles that began before the onset of the joint noise. Such patients must be distinguished from patients with a primary internal derangement, in whom muscle splinting produces myofascial pain and tenderness after the onset of the joint noise. The history and differences in physical findings are helpful in making this distinction.

In addition to having the three cardinal symptoms of pain, muscle tenderness, and limitation of mouth opening, patients with MPD usually have no clinical or radiographic evidence of pathologic changes in the TMJ. These negative characteristics are important in establishing the diagnosis because they confirm that the primary site of the problem is not the articular structures.

**Diagnosis**

Because the cardinal signs and symptoms of MPD are similar to those produced by such organic problems involving the TMJ as degenerative joint disease and internal disk derangement and by a variety of nonarticular conditions (see Tables 51-1 and 51-2), diagnosis of this condition can be difficult, requiring a careful history and a thorough clinical evaluation. Periapical radiographs of the teeth and screening radiographs (transcranial, transpharyngeal, or panoramic) of the TMJs can be helpful in eliminating dental problems or gross joint disease. If screening views of the TMJs show some abnormality, CT scans are usually advisable for confirmation. MRI also can be useful in determining the position of the disk when an internal derangement of the TMJ is being considered. Depending on the suspected condition, other radiographic views of the head and neck and scintigraphy may be needed to establish a final diagnosis.

Certain laboratory tests may be helpful in some instances. These include a complete blood cell count if an infection is suspected; serum calcium, phosphorus, and alkaline phosphatase measurements for possible bone disease; serum uric acid determination for gout; serum creatinine and creatine kinase levels as indicators of muscle disease; and erythrocyte sedimentation rate, rheumatoid factor, latex fixation, and antinuclear antibody tests for suspected rheumatoid arthritis. Electromyography can be used to evaluate muscle function. Psychological evaluation and psychometric testing are good research tools, but they have little diagnostic value other than in identifying the presence of associated abnormal behavioral characteristics.

A condition that sometimes is confused with myofascial pain is fibromyalgia, particularly when MPD involves several regions in addition to the face. Although a small subset of patients with MPD eventually may develop fibromyalgia, these are probably distinct conditions.\(^\text{31,57}\) Table 51-4 lists...
becomes symptom free, the medications are stopped first, and wearing the bite appliance is discontinued next. If the patient has a return of symptoms, and the appliance is worn only at night, its use can be continued indefinitely.

Patients who do not respond to the use of a bite appliance are entered into phase III treatment for 4 to 6 weeks. In this phase, physical therapy (heat, massage, ultrasound, electrogalvanic stimulation) or relaxation therapy (electromyographic biofeedback, conditioned relaxation) is added to the regimen. No evidence shows that one form of treatment is better than another, and either can be used

the distinguishing characteristics of myofascial pain versus fibromyalgia.

**Treatment**

Treatment of MPD is divided into four phases. When a definitive diagnosis is made, phase I therapy should be started (Figure 51-6). Phase I therapy initially involves providing the patient with some understanding of the problem. Because patients often have difficulty accepting a psychophysiologic explanation for their condition, the discussion should deal with the issue of muscle fatigue as the cause of the pain and dysfunction, delaying consideration of the role of stress and psychological factors until the symptoms have improved, and the patient’s confidence has been gained. Relating the symptoms to the specific masticatory muscles from which they arise helps the patient understand the reason for the type and location of the pain—headache from the temporalis muscle, jaw ache from the maseter muscle, discomfort on swallowing and stuffiness in the ear from the medial pterygoid muscle, and earache and pain behind the eye from the lateral pterygoid muscle.

In addition to the initial explanation, the patient should be counseled regarding home therapy; this includes recommendations about avoidance of clenching and grinding of the teeth, eating a soft diet, use of moist heat and massage on the masticatory muscles, and limitation of jaw movement. A nonsteroidal anti-inflammatory drug should be prescribed for the pain. In patients who have problems sleeping, a small dose of amitriptyline at bedtime is helpful in improving sleep and reducing parafunction.

About 50% of these patients experience resolution of their symptoms within 2 to 4 weeks with phase I therapy. For patients whose symptoms persist, phase II therapy is initiated. Home therapy and medications are continued, and a bite appliance is made for the patient. Although numerous types have been used, the Hawley-type maxillary appliance is probably most effective because it prevents contact of the posterior teeth and prevents most forms of parafunctional activity (Figure 51-7). The appliance generally is worn at night, but it can be worn for 5 to 6 hours during the day, if necessary. The appliance should not be worn continuously, however, because the posterior teeth may supraerupt in some patients.

With phase II therapy, another 20% to 25% of patients become symptom free in 2 to 4 weeks. When the patient
first. If one is unsuccessful, the other can be tried. Phase III therapy usually helps another 10% to 15% of patients.

If all of these approaches fail, and no question arises about the correctness of the diagnosis, psychological counseling is recommended. This counseling involves helping patients identify possible stresses in their lives and learning to cope with such situations. If the diagnosis is in doubt, the patient should be referred first for appropriate dental and neurologic consultation and re-evaluation. Another alternative is to refer patients with recalcitrant MPD to a TMJ center or pain clinic because such patients generally require a multidisciplinary approach for successful treatment.

SUMMARY

Successful management of patients with temporomandibular disorders depends on establishing an accurate diagnosis and using proper therapy based on an understanding of the cause of the condition being treated. Of particular importance is separating patients with MPD, who constitute the major group encountered and who are not surgical candidates, from patients with TMJ disease, who frequently require surgical treatment. Even in the latter group, many commonly encountered conditions, such as arthritis and internal disk derangements, often respond to nonsurgical therapy, and this type of treatment should be given a fair trial before more aggressive management is considered.

References


Websites
www.nicd.nih.gov—General information, clinical trials, and sponsored research in TMJ and related areas.
www.aao.org—General information about TMJ surgery.
www.tmj.org—Advocate group that provides general information for patients.

The references for this chapter can also be found on www.expertconsult.com.
Fibromyalgia

FREDERICK WOLFE • JOHANNES J. RASKER

Fibromyalgia is a controversial disorder. Certain aspects of the controversies surrounding fibromyalgia reflect scientific disagreements about categorization, pathophysiology, and treatment (Figure 52-1). But another important reason for controversy is that the diagnosis carries with it profound societal consequences. Whether fibromyalgia “exists” or is “real” or should be valued matters a great deal to patients, payers, pension systems, researchers, professional and patient organizations, and pharmaceutical companies. Fibromyalgia is a clinical syndrome that is defined by the presence of generalized pain, fatigue, unrefreshed sleep, multiple somatic symptoms, cognitive problems, and other symptoms, often including depression. Symptoms important to the fibromyalgia case definition are shown in Figure 52-2 in order of their importance. The 2010 American College of Rheumatology (ACR) preliminary diagnostic criteria for fibromyalgia require the presence of widespread pain and multiple symptoms (Table 52-1). The more restrictive ACR 1990 classification criteria require the presence of widespread pain plus the presence of tenderness on palpation in at least 11 of 18 specified “tender point” sites. Fibromyalgia can be diagnosed in the presence of other medical conditions and is never a diagnosis of exclusion. However, concomitant disorders associated with musculoskeletal pain and fatigue will always need to be identified.

THE FIBROMYALGIA CONSTRUCT

One Syndrome or One of Many?

The central features of fibromyalgia that were noted earlier are also found in illnesses such as chronic fatigue syndrome, irritable bowel syndrome, headache syndromes, and multiple chemical sensitivities, among many others (see Figure 52-1). Taken together, these syndromes have been called functional somatic syndromes (FSS). Because the symptom content of the syndromes is similar, as are the treatments and the demographic characteristics of patients who have the disorders, it has been suggested by many that a single diagnostic term, rather than individual syndrome names, should be used for diagnosis. These suggestions derive mostly from the psychiatric literature. Terms suggested include FSS and bodily pain disorder. However, if fibromyalgia is just a name given to the disorder primarily by rheumatologists, but does not differ essentially from other somatic syndromes, then fibromyalgia does not exist as a separate syndrome. Fibromyalgia versus FSS creates a series of problems. FSS connotes a strong psychologic component, which is undesirable to patients, pharmaceutical companies, and medical researchers. In addition, there is the logical inconsistency in which regulatory authorities such as the U.S. Food and Drug Administration (FDA) approve treatments for the select fibromyalgia indication, when fibromyalgia is not different from other FSS. The FDA mandate strengthens the position of fibromyalgia as a “separate” disease, although there is little evidence that it is such an entity.

A Separate Syndrome or Part of a Continuum?

Fibromyalgia is properly considered to lie “at the extreme end of the spectrum of polysymptomatic distress.” Fibromyalgia diagnosis depends on splitting the distress continuum, placing on one side of the divide those with fibromyalgia and on the other side all other persons. Polysymptomatic distress refers to problems in many symptom areas—pain, fatigue, sleep disturbance, functional impairment, psychologic status, and so on. Because symptoms are correlated, persons with high levels of one symptom will...
Figure 52-1  Fibromyalgia controversies. CSS, central sensitivity syndrome; FM, fibromyalgia; FSS, functional somatic syndromes; TMJ, temporomandibular joint.

Figure 52-2  Symptoms that differentiate patients who satisfy American College of Rheumatology 1990 criteria from other rheumatic disease patients with noninflammatory rheumatic pain disorders sorted by strength of association. The two figures represent different measures of association. Higher scores mean stronger associations. (C), categorical variable; IBS, irritable bowel syndrome; WPI, Widespread Pain Index.
tend to have high levels of other symptoms. As an aggregate concept, polysymptomatic distress cannot be measured directly but can be approximated with the use of surrogate variables. One such surrogate measure of polysymptomatic distress is the fibromyalgia symptom scale—also called the fibromyalgianess scale.12 This scale represents the summation of the Widespread Pain Index (the number of body sites reported as painful) and characteristic fibromyalgia symptoms used in the ACR 2010 preliminary diagnostic criteria.3 In patients with various rheumatic diseases followed in the U.S. National Data Bank for Rheumatic Diseases,13 the upper part of Figure 52-3 shows the relation between the scale and the Short Form-36 (SF-36) Physical Component Summary (PCS) score, the EQ5D Quality of Life score, and the patient’s estimate of global severity. A value of 13 on the fibromyalgianess scale best divides fibromyalgia-positive and fibromyalgia-negative patients.14 It can also be seen that the Widespread Pain Index alone is similarly associated with these three measures of illness severity (see Figure 52-3, lower part). About 2% to 4% of the adult population meets criteria for fibromyalgia. One can sense in the figure the distribution of polysymptomatic distress and its correlation with quality of life. Note that polysymptomatic distress is a quantity that exists in all persons, not just in those with fibromyalgia, though it is greater in those with fibromyalgia.

The higher the score on the polysymptomatic distress scale and the Widespread Pain Index, the more likely we are to find evidence of social disadvantage such as lower income, less education, and childhood mistreatment, and we will also find more psychologic distress and abnormality; it appears that these factors play a role in the development of fibromyalgia-like symptoms and symptom intensification.14 Fibromyalgianess differs from other measures of polysymptomatic distress by the centrality of musculoskeletal pain because nonarticular musculoskeletal pain is a central component of the scale.

To define fibromyalgia by criteria, we must, in effect, draw a line on the distress continuum and say that those beyond this line have fibromyalgia and those before it do not. In the ACR 2010 criteria (see Table 52-1),3 the cut point is identified by the extent of widespread pain and fibromyalgia symptoms. In the 1990 criteria (Table 52-2), the cut point is represented by a combination of tender points and widespread pain. Both cut points, though aided by data analyses, are determined by committees. There is nothing intrinsic in the polysymptomatic distress scale that tells us where the dividing point is. But in the general population a PCS score of 50 represents the population mean, with each standard deviation representing 10 units. At a fibromyalgianess score of 13, patients designated as

### Table 52-1  American College of Rheumatology 2010 Preliminary Diagnostic Criteria for Fibromyalgia3

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A patient satisfies diagnostic criteria for fibromyalgia if the following 3 conditions are met:</td>
<td></td>
</tr>
<tr>
<td>1. Widespread Pain Index (WPI) ≥ 7 and Symptom Severity Score ≥ 5 or</td>
<td></td>
</tr>
<tr>
<td>WPI between 3 and 6 and Symptom Severity Score ≥ 9.</td>
<td></td>
</tr>
<tr>
<td>2. Symptoms have been present at a similar level for at least 3 months.</td>
<td></td>
</tr>
<tr>
<td>3. The patient does not have a disorder that would otherwise explain the pain.</td>
<td></td>
</tr>
</tbody>
</table>

### Ascertainment

1. WPI: Note the number areas in which the patient has had pain over the past week. In how many areas has the patient had pain? Score will be between 0 and 19.

<table>
<thead>
<tr>
<th>Part of Body</th>
<th>Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulder girdle, Lt.</td>
<td>1</td>
</tr>
<tr>
<td>Shoulder girdle, Rt.</td>
<td>2</td>
</tr>
<tr>
<td>Upper arm, Lt.</td>
<td>3</td>
</tr>
<tr>
<td>Upper arm, Rt.</td>
<td>4</td>
</tr>
<tr>
<td>Lower arm, Lt.</td>
<td>5</td>
</tr>
<tr>
<td>Lower arm, Rt.</td>
<td>6</td>
</tr>
<tr>
<td>Hip (buttock, trochanter), Lt.</td>
<td>7</td>
</tr>
<tr>
<td>Hip (buttock, trochanter), Rt.</td>
<td>8</td>
</tr>
<tr>
<td>Jaw, Lt.</td>
<td>9</td>
</tr>
<tr>
<td>Jaw, Rt.</td>
<td>10</td>
</tr>
<tr>
<td>Upper leg, Lt.</td>
<td>11</td>
</tr>
<tr>
<td>Lower leg, Lt.</td>
<td>12</td>
</tr>
<tr>
<td>Lower leg, Rt.</td>
<td>13</td>
</tr>
<tr>
<td>Chest</td>
<td>14</td>
</tr>
<tr>
<td>Neck</td>
<td>15</td>
</tr>
<tr>
<td>Abdomen</td>
<td>16</td>
</tr>
</tbody>
</table>

2. Symptom Severity Score:
   - Fatigue
   - Waking unrefreshed
   - Cognitive symptoms

For the each of the three symptoms above, indicate the level of severity over the past week using the following scale:

- 0 = No problem
- 1 = Slight or mild problems; generally mild or intermittent
- 2 = Moderate; considerable problems; often present and/or at a moderate level
- 3 = Severe: pervasive, continuous, life-disturbing problems

Considering somatic symptoms* in general, indicate whether the patient has:

- 0 = No symptoms
- 1 = Few symptoms
- 2 = A moderate number
- 3 = A great deal of symptoms

The Symptom Severity Score is the sum of the severity of the three symptoms (fatigue, waking unrefreshed, cognitive symptoms) plus the extent (severity) of somatic symptoms in general. The final score is between 0 and 12.

*For reference purposes, here is a list of somatic symptoms that might be considered: muscle pain, irritable bowel syndrome, fatigue/tiredness, thinking or remembering problem, muscle weakness, headache, pain/cramps in abdomen, numbness/tingling, dizziness, insomnia, depression, constipation, pain in upper abdomen, nausea, nervousness, chest pain, blurred vision, fever, diarrhea, dry mouth, itching, wheezing, Raynaud’s phenomenon, hives/welts, ringing in ears, vomiting, heartburn, oral ulcers, loss/change in taste, seizures, dry eyes, shortness of breath, loss of appetite, rash, sun sensitivity, hearing difficulties, easy bruising, hair loss, frequent urination, painful urination, and bladder spasms.
fibromyalgia have a PCS score about 2 standard deviations below the mean (see Figure 52-3, upper left). Thus fibromyalgia diagnosis identifies persons with substantially reduced quality of life—those at the “extreme end of the spectrum of polysymptomatic distress.”

### Social Construction and Medicalization

Many physicians doubt the existence of fibromyalgia as a separate entity, considering instead that it is primarily a psychologic illness—not a “real disease” (see Figures 52-1 and 52-4). Epidemiologic and clinical studies give no support to the idea that fibromyalgia is a distinct entity; instead, they support the contrary idea that fibromyalgia represents the end of a spectrum of polysymptomatic distress.

Ilnesses exist within societies, and their existence and phenotype are often a function of the degree of acceptance of the disorder. The idea and consequences of fibromyalgia as a socially constructed, medicalized disorder has been discussed at length. An illness may be considered to be socially constructed when it is at least in large part the consequence

### Table 52-2 1990 American College of Rheumatology Criteria for the Classification of Fibromyalgia

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Definition</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. History of Widespread Pain</td>
<td>Pain is considered widespread when all of the following are present: pain in the left side of the body, pain in the right side of the body, pain above the waist, and pain below the waist. In addition, axial skeletal pain (cervical spine or anterior chest or thoracic spine or low back) must be present. In this definition, shoulder and buttock pain is considered as pain for each involved side. “Low back” pain is considered lower segment pain.</td>
<td></td>
</tr>
<tr>
<td>2. Pain in 11 of 18 Tender Point Sites on Digital Palpation</td>
<td>Pain, on digital palpation, must be present in at least 11 of the following 18 sites: Occiput: bilateral, at the suboccipital muscle insertions. Low cervical: bilateral, at the anterior aspects of the intertransverse spaces at C5-C7. Trapezius: bilateral, at the midpoint of the upper border. Supraspinatus: bilateral, at origins, above the scapula spine near the medial border. Second rib: bilateral, at the second costochondral junctions, just lateral to the junctions on upper surfaces. Lateral epicondyle: bilateral, 2 cm distal to the epicondyles. Gastrocnemius: bilateral, in upper outer quadrants of buttocks in anterior fold of muscle. Greater trochanter: bilateral, posterior to the trochanteric prominence. Knee: bilateral, at the medial fat pad proximal to the joint line. Digital palpation should be performed with an approximate force of 4 kg.</td>
<td>For a tender point to be considered “positive,” the subject must state that the palpation was painful. “Tender” is not to be considered “painful.”</td>
</tr>
</tbody>
</table>

*For classification purposes, patients will be said to have fibromyalgia if both criteria are satisfied. Widespread pain must have been present for at least 3 months. The presence of a second clinical disorder does not exclude the diagnosis of fibromyalgia.

CHAPTER 52 | FIBROMYALGIA

2 Primarily a psychosomatic illness.

2 The diagnosis of a “valid” fibromyalgia provides entry to medical insurance and treatment and is grounds for work disability and pension. Extensive networks of patient organizations throughout the world work toward this purpose.

2 The second pillar of medicalization in fibromyalgia is the pharmaceutical industry. Direct-to-patient advertising is ubiquitous. Often deceptive, it seeks to expand the definition of fibromyalgia, entice persons with pain and fatigue into the diagnosis, and strongly promote its treatments as effective.

2 Industry financially supports patient and professional organization, medical education and symposia,2,28 and advertising in professional and lay journals. Virtually all major authors of fibromyalgia studies have received pharmaceutical company support. The influence of drug companies has increased dramatically in the past two decades to the extent that “… companies are having an increasing impact on the boundaries of the normal and the pathological, becoming active agents of social control.”

23 Although “medicalization is now more driven by commercial and market interests than by professional claim-makers,”23 physicians and professional organizations remain the importance sources of scientific support, and National Institutes of Health grants for fibromyalgia research have become common.

The more fibromyalgia is seen as a “real disease” with strong criteria, reliable assessments and professional support,
the more the claims of patients will be taken seriously, the more drugs will be sold and consumed, and the more financial and intellectual support will come to researchers and professional organizations. By contrast, perceiving fibromyalgia as an FSS and part of a continuum with assessments that are not always reliable weakens support for the syndrome and those who have it. The issues that surround fibromyalgia are not scientific ones—it is widely agreed that pain and suffering are real; instead, the issues are social, political, and financial.

Historical Development

Attempts to characterize and diagnose fibromyalgia have gone through several changes in conceptualization. The earliest roots of the syndrome can be found in the nineteenth century perception of abnormal connective tissue and muscles. In various forms, this concept held sway until the late 1970s when a new emphasis on sleep disturbance and tender points led to proposed clinical criteria that included sleep disturbance and tenderness to palpation at 12 of 14 selected sites. In the early 1980s, most of the other fibromyalgia-associated symptoms were identified, and unofficial criteria were proposed that combined symptoms with tenderness. With the publication of the ACR criteria in 1990, the fibromyalgia case definition was reduced to generalized pain and the presence of multiple tender points. The 1990 criteria, supported by the imprimatur of the ACR, gave the syndrome official sanction. The criteria publication, professional opposition to fibromyalgia solidified and has continued until the present. In 2010 the ACR preliminary diagnostic criteria were published. These criteria expanded the case definition and criteria items to include widespread pain and multiple symptoms including fatigue, disturbed sleep, cognitive symptoms, and multiple somatic symptoms. Despite scientific data questioning the validity of fibromyalgia as a distinct entity, fibromyalgia has become a dominant paradigm, supported strongly by funding and influence of the pharmaceutical industry. At present there are currently several opposing views of fibromyalgia. One view holds that it is not distinct and is a part of FSS, with a functional symptom being defined as one that “after appropriate medical assessment, cannot be explained in terms of a conventionally defined medical disease.” The second view, and the dominant paradigm, supports the concept that fibromyalgia is a disorder of “... aberrant central pain transmission ... [in which] purely behavioral or psychologic factors are not primarily responsible for the pain and tenderness ...” A third view holds that fibromyalgia is the end of a spectrum of polysymptomatic distress and is not a distinct entity. Perhaps, not surprisingly, these views are not mutually exclusive.

**CLINICAL FEATURES**

Fibromyalgia is characterized by high levels of pain, sleep disturbance, and fatigue combined with a general increase in medical symptoms (Table 52-3) including problems of memory or thinking and often psychologic distress. Individuals with the syndrome are unusually sensitive to digital pressure (tender points) in certain body areas. Clinically, fibromyalgia is often identified or suspected by the inexplicability and severity of symptoms and by their number. The most common defining symptom is that of generalized pain (“pain all over”), and pain all over, or widespread pain, is a requirement of the 1990 and 2010 criteria. The clinician may be surprised by the extent and severity of symptoms (see Table 52-3 and Figure 52-3) and surprised at unexpected emotional distress. Fibromyalgia has a quality of inexplicability and unexpectedness.

Upper and lower back pain is the most common pain problem (>80%). Many patients, at the clinical interview, emphasize only a few areas of pain. Questions specifically directed to other areas may elicit reports of pain that were not stated spontaneously. Patients with fibromyalgia may complain of greater pain in an osteoarthritic joint than patients without fibromyalgia. Although musculoskeletal pain is central to fibromyalgia, patients may be more concerned about fatigue or memory problems.

Fibromyalgia patients perform more poorly in formal cognitive testing than age-matched controls. In the National Data Bank for Rheumatic Diseases in 2006, 66% of 2784 fibromyalgia patients complained of memory or thinking problems compared with 31% of 24,479 patients with other rheumatic conditions. The most common symptoms, found

<table>
<thead>
<tr>
<th>Symptom</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep problems</td>
<td>89.1</td>
</tr>
<tr>
<td>Fatigue or tiredness</td>
<td>88.6</td>
</tr>
<tr>
<td>Muscle pain</td>
<td>85.2</td>
</tr>
<tr>
<td>Muscle weakness</td>
<td>70.2</td>
</tr>
<tr>
<td>Paresthesias</td>
<td>67.6</td>
</tr>
<tr>
<td>Cognitive problems</td>
<td>66.3</td>
</tr>
<tr>
<td>Headache</td>
<td>64.7</td>
</tr>
<tr>
<td>Dry mouth</td>
<td>53.3</td>
</tr>
<tr>
<td>Insomnia</td>
<td>51.8</td>
</tr>
<tr>
<td>Easy bruising</td>
<td>49.1</td>
</tr>
<tr>
<td>Dry eyes</td>
<td>47.5</td>
</tr>
<tr>
<td>Depression</td>
<td>47.5</td>
</tr>
<tr>
<td>Blurred vision</td>
<td>47.0</td>
</tr>
<tr>
<td>Irritable bowel syndrome</td>
<td>46.3</td>
</tr>
<tr>
<td>Heartburn</td>
<td>44.4</td>
</tr>
<tr>
<td>Itching</td>
<td>44.3</td>
</tr>
<tr>
<td>Dizziness</td>
<td>42.1</td>
</tr>
<tr>
<td>Constipation</td>
<td>41.9</td>
</tr>
<tr>
<td>Pain/cramps in the abdomen</td>
<td>41.5</td>
</tr>
<tr>
<td>Ringing in ears</td>
<td>41.4</td>
</tr>
<tr>
<td>Pain in upper abdomen</td>
<td>40.3</td>
</tr>
<tr>
<td>Nervousness</td>
<td>39.7</td>
</tr>
<tr>
<td>Nausea</td>
<td>37.7</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>33.6</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>32.3</td>
</tr>
<tr>
<td>Hearing difficulties</td>
<td>29.8</td>
</tr>
<tr>
<td>Hair loss</td>
<td>23.6</td>
</tr>
<tr>
<td>Oral ulcers</td>
<td>22.4</td>
</tr>
<tr>
<td>Wheezing</td>
<td>21.4</td>
</tr>
<tr>
<td>Loss of appetite</td>
<td>21.1</td>
</tr>
<tr>
<td>Raynaud’s phenomenon</td>
<td>20.1</td>
</tr>
<tr>
<td>Chest pain</td>
<td>19.2</td>
</tr>
<tr>
<td>Rash</td>
<td>17.1</td>
</tr>
<tr>
<td>Sun sensitivity</td>
<td>16.7</td>
</tr>
<tr>
<td>Loss/change of taste</td>
<td>14.4</td>
</tr>
<tr>
<td>Fever</td>
<td>13.4</td>
</tr>
<tr>
<td>Hives/welts</td>
<td>9.3</td>
</tr>
<tr>
<td>Vomiting</td>
<td>9.1</td>
</tr>
<tr>
<td>Seizures</td>
<td>1.7</td>
</tr>
</tbody>
</table>
in more than two-thirds of patients, are sleep problems, fatigue, muscle pain, paresthesias, and cognitive problems (see Table 52-3). In addition, the prevalence of other important symptoms is as follows: headache, 65%; depression, 48%; and irritable bowel syndrome, 46% (see Table 52-3). A high count of symptoms is characteristic of fibromyalgia and is frequently a key item in the 2010 diagnostic criteria to diagnosis (see Figure 52-2). Fibromyalgia is also associated with increased reporting of comorbid conditions. The typical picture of fibromyalgia emphasizes certain symptoms (pain, fatigue, sleep disturbance, cognitive problems) and an abundance of symptoms and comorbidities. Given the high levels of symptom variables and membership at the tail of the pain-distress continuum, it is not surprising that evidence of psychosocial disruption and high rates of lifetime psychiatric illness are found.

Fibromyalgia occurs frequently in other rheumatic disorders including rheumatoid arthritis, osteoarthritis, and systemic lupus erythematosus, in which the prevalence of fibromyalgia exceeds 20%. The clues to identifying fibromyalgia in the presence of other painful disorders are location of pain (nonarticular), continued pain and distress despite objective improvement in the concomitant disorder, and unusual fatigue.

**ASSESSMENT AND DIAGNOSIS OF A PATIENT WITH FIBROMYALGIA**

**Diagnosis and Diagnostic Criteria**

A number of approaches to fibromyalgia diagnosis are available. To treat patients, recognition of the degree of pain, fatigue, and other symptoms is necessary, but a specific diagnostic term is not. Chronic pain syndrome, FSS, or fibromyalgia will all suffice for a diagnostic term in most settings. But in countries such as the United States, chronic pain syndrome or FSS often may not be sufficient for access to insurance reimbursement or pension systems. In addition, direct-to-patient advertising may influence diagnostic terminology and diagnosis toward fibromyalgia.

Today, two sets of valid criteria for fibromyalgia are used in most of the world, although country-specific criteria also exist. The approach to fibromyalgia diagnosis should differ according to the setting and the physician's underlying beliefs about fibromyalgia acceptability. The 1990 ACR classification criteria (see Table 52-2) require the presence of widespread pain and the identification of pain on palpation at 11 or more of 18 tender points. Until 2010, with the publication of the ACR preliminary diagnostic criteria, the 1990 criteria was the only method for an official diagnosis. The 2010 diagnostic criteria are easier in some ways and more difficult in others. They are easier because they eliminate the tender point examination that may be difficult for some examiners. The 2010 criteria are more difficult because they require a thorough symptom evaluation. One advantage of the 2010 criteria is that the examiner/interviewer becomes much more familiar with the spectrum and degree of the patient's problem. But for the criteria to work correctly, the interviewer must be comprehensive and thorough. The 2010 criteria provide two scales to evaluate the degree of polysymptomatic distress: the symptom severity scale and the fibromyalgianess scale, both of which are discussed earlier. The fibromyalgianess scale has the advantage that it is a continuous measure of polysymptomatic distress. It is suitable for use in all patients whether or not they satisfy fibromyalgia criteria now or have satisfied them in the past. The scale is also useful when the physician or examiner does not believe in the fibromyalgia concept because it does not require a criteria diagnosis to be useful. The ACR 2010 criteria have been modified by the authors so that self-report forms can be used. Although these self-report, form-based criteria can be useful for survey and clinical research, they have not been endorsed by the ACR and they should never be used for clinical diagnosis.

**Primary, Secondary, and Secondary-Concomitant Fibromyalgia**

Fibromyalgia is sometimes divided into primary, secondary, and secondary-concomitant fibromyalgia. The term primary fibromyalgia is most often used when there is not another condition with symptoms that could explain fibromyalgia symptoms. This division between primary and secondary fibromyalgia is artificial, however. Back pain in older individuals when age-related radiographic changes are present might be considered secondary fibromyalgia, whereas the same symptoms in younger individuals might be considered primary fibromyalgia. The ACR 1990 criteria study showed no difference between primary and secondary fibromyalgia with regard to symptoms and diagnosis. The usefulness of primary fibromyalgia occurs in clinical trials, in which it is desirable to ensure those symptoms are not coming from another well-established illness. A fibromyalgia diagnosis implies understanding of issues such as pain, fatigue, sleep, and cognitive and emotional problems. When fibromyalgia is considered only in patients without other musculoskeletal conditions, the “benefit” of fibromyalgia diagnosis—its consideration of symptom issues and extent of pain—is lost. If fibromyalgia is to be diagnosed or considered, such consideration should be applied to all patients. As noted earlier, fibromyalgia is never a diagnosis of exclusion. When fibromyalgia is diagnosed in the presence of another condition, treatment is indicated for one or both disorders, as determined clinically.

**Differential Diagnosis**

The primary symptoms of fibromyalgia, widespread pain and fatigue, can be found in many medical disorders. Similarly, fibromyalgia can coexist with other medical conditions. The proper approach to avoiding misdiagnosis is to ascertain the presence or absence of fibromyalgia and then to determine whether other disorders with widespread pain and fatigue are present. Practically, the categories are fibromyalgia AND other disorders, fibromyalgia AND NOT other disorders, and other disorders AND NOT fibromyalgia. Conditions with fibromyalgia-like features include polymyalgia rheumatica, polymyositis, lupus, cervical spine disorders, hypermobility syndromes, endocrine and paraneoplastic disorders, and forms of polyarthritis including rheumatoid arthritis and ankylosing spondylitis. When differential diagnosis is problematic, it is because the other medical condition is difficult to diagnose or has not been
evaluated properly. The clue to understanding a patient’s illness is thoroughly evaluating the patient with a careful history, physical examination, and laboratory evaluation.

**ASSESSMENT OF FIBROMYALGIA SEVERITY**

**Self-Report Measures**

The ACR 2010 preliminary diagnostic criteria provided a new measure of fibromyalgia severity, the Symptom Severity Score (see Table 52-1). Used in the diagnosis of fibromyalgia, this scale also functions as a measure of the severity of fibromyalgia symptoms and can be useful independently of the criteria. Another scale that is an effective measure is the fibromyalgianess scale. It is the sum of the two items used in the 2010 criteria, the Widespread Pain Index and the Symptom Severity Score. It is suitable for use in all patients, regardless of fibromyalgia status, thereby integrating fibromyalgianess and fibromyalgia symptoms into general patient care.

Symptom severity, physical function, and work status are the key status and outcome variables in fibromyalgia, as in other rheumatic disorders. Assessments that can be useful routinely to clinicians include measurements of pain, fatigue, physical function, sleep quality, anxiety, depression, and work status. At minimum, assessments should include visual analog scales (VAS) for pain and fatigue and a measure of functional status. Function can be assessed by one of the family of health assessment questionnaires including the Health Assessment Questionnaire (HAQ), the Health Assessment Questionnaire–II (HAQ-II), and the Multidimensional Health Assessment Questionnaire (MDHAQ). The HAQ is a 33-item questionnaire; the function scale of the HAQ-II and MDHAQ is a 10-item questionnaire. Simple scales for the assessment of anxiety, depression, and sleep disturbance also can be added. For simplicity and ease of administration, however, we recommend VAS assessments of pain and fatigue and either the HAQ-II or MDHAQ.

The Fibromyalgia Impact Questionnaire (FIQ) is a widely used 21-item research assessment scale that addresses all of the key fibromyalgia variables and can be used in clinical care. The limitation of the FIQ is that it is suitable only for use in fibromyalgia patients, whereas the previously mentioned health assessment questionnaires are useful and have been used across the entire range of rheumatic disorders. In addition, the FIQ total scale has no simple interpretation.

Functional questionnaire results have reduced validity among fibromyalgia patients. Compared with patients with rheumatoid arthritis and ankylosing spondylitis, there was striking discordance between observed and questionnaire-reported activities in patients with fibromyalgia. This discordance limits slightly the usefulness of functional questionnaires and alters their interpretation: Results may represent perceived rather than actual functional difficulties.

**Research Questionnaires**

The Outcome Measures in Rheumatoid Arthritis Clinical Trial committee has recommended research domains and questionnaires for fibromyalgia clinical trials. These domains include pain, fatigue, sleep, depression, physical function, quality of life and multidimensional function, patient’s global impression of change, tenderness, dyscognition, anxiety, and stiffness. The recommendations include the use of the FIQ and the Medical Outcomes Scale SF-36. A recent study using observational data has shown that pain, HAQ, and fatigue explained more than 50% of fibromyalgia severity variance and that the main determinants of global severity and health-related quality of life in fibromyalgia are pain, function, and fatigue. On the basis of the ACR 2010 preliminary diagnostic criteria, criteria and survey assessments have been developed. The Symptom Intensity Scale, which combines the Widespread Pain Index and a VAS fatigue scale, is another self-report measure of fibromyalgia severity that is suitable for clinical and survey research.

**Physical Measures**

With the exception of the performance of the tender point examination, the physical examination of a patient suspected to have fibromyalgia does not differ from the examination of any other rheumatic disease patient or pain patient. Measurement of pain threshold by the tender point examination is the only routinely useful physical measurement. Although helpful for diagnosis using the ACR 1990 classification criteria (see Table 52-2), the tender point count is poorly correlated with other fibromyalgia symptoms and with change in symptom severity among fibromyalgia patients. Patients may improve or worsen substantially without important differences in the tender point count.

**How to Perform the Tender Point Examination**

Fibromyalgia patients have a lower threshold for pain than do subjects without fibromyalgia. In the clinic, two methods exist by which tenderness can be elicited and measured—digital palpation and dolorimetry. Tender point sites represent specific areas of muscle, tendon, and fat pads that are much more tender to palpation than surrounding sites. Sites selected as part of ACR 1990 criteria represent tender point sites that best discriminate between patients with and without fibromyalgia. To test for pain with digital palpation, the ACR 1990 criteria indicate that the examiner should press the tender point site with an approximate force of 4 kg. Usually the second and third fingers or the thumb is used for palpation, and a rolling motion is helpful in eliciting tenderness. The amount of force that the examiner uses is important because a large force would elicit pain in a subject without fibromyalgia, whereas a small force may miss tenderness. The amount of force that does not elicit tenderness in an individual without fibromyalgia (just below the pressure pain threshold) is the correct force to use. In practice, less force is required in smaller, thinner, less-muscled individuals. The pressure used by the examiner and the examiner's interpretation of the patient's response can influence results of palpation. The best and most appropriate way to perform the tender point count is to ask the patient if the palpation is painful, accepting only a “yes” as a positive reply, regardless of facial expression or body
movement. Specifically, the frequently heard comment of patients to the digital examiner’s question regarding pain, “It’s tender,” is a negative rather than a positive response and should be followed by another question such as, “Yes, but is it painful?”

**Limitations to the Tender Point Examination**

Although the tender point examination can provide clinically useful information when properly performed, it can be influenced by external factors. Physicians who believe the patient does or does not have fibromyalgia can influence the results by the amount of pressure applied. The meaning and use of the examination are widely known among physicians, patients, and patient support groups; in some circumstances where a positive or negative examination would seem to be desirable (e.g., in a disability or medicolegal examination), results might differ from those obtained during a routine examination. In addition, the tender point examination is inherently inaccurate around the “diagnostic” tender point count of 11.

**Epidemiology**

Most of the information about fibromyalgia is based on sampling using the ACR 1990 criteria. Fibromyalgia is diagnosed more frequently in women (9:1 ratio) in clinical studies. However, in population-based studies the female-to-male ratio is lower. A recent five-country European study noted the female-to-male ratio to be about 1.7:1, though a U.S. study found a ratio of 6.8:1 and the ratio varies from high to low in other countries. 

Using ACR criteria, the prevalence of fibromyalgia in the adult general population is generally similar across the world. The prevalence of fibromyalgia in Wichita, Kansas, was 3.4% among women, 0.5% among men, and 2% overall; among women in New York City, it was 3.7%. In Ontario, Canada, the estimated prevalence was 4.9%; among women, 1.6% among men, and 3.3% overall. The prevalence of fibromyalgia in these studies increased with age until about age 70, after which it decreased slightly. Outside of North America, reports indicate the prevalence in five European countries was 4.7% and 2.9% according to different screening methods; in studies in Bangladesh it was 5.3% to 7.5% in women and 0.2% to 1.4% in men; in North Pakistan, it was 2.1% overall; in Italy, it was 2.2%; in Turkey, it was 3.6% for ages 20 to 64; in Brazil, it was 2.5%; and in Southwest Sweden, it was 1.3%.

The prevalence of fibromyalgia in children in three studies was 1.2%, 1.4%, and 6.2%. At a follow-up time of 1 year, approximately 25% of individuals meeting ACR criteria initially still satisfied the criteria. These data should not be interpreted as evidence of prognosis because some individuals not meeting criteria initially meet them at the 1-year follow-up. Instead, the data suggest that the concept of fibromyalgia in children may be dubious, particularly when dependent on tender point assessment. The prevalence of fibromyalgia is generally greater in clinical settings than in epidemiologic studies. It was noted to be 5.7% in general medical clinics and 2.1% in family practice settings. In rheumatology clinics, fibromyalgia prevalence was expectedly higher: 12% to 20% of new patients.

The ACR 2010 criteria should result in changes in the sex ratio because men have higher pain thresholds and are therefore less likely to be diagnosed as having fibromyalgia than women when the 1990 criteria are used. The proportion of men with fibromyalgia in the community in a large German population study was 40.3%. This study included criteria that used the Regional Pain Scale and measurement of fatigue. Diagnosis by this method yields results that are similar to survey modifications on the ACR 2010 preliminary criteria. The overall prevalence in the German study was 3.8%. Additional studies are necessary to determine the prevalence of fibromyalgia when the 2010 criteria are used.

**Etiology and Pathophysiology**

In the 30-year period following the establishment of the fibromyalgia case definition and criteria, there have been substantial advances in understanding mechanisms associated with fibromyalgia pain and other symptoms. Although most of the recent study data are robust, the interpretation of the data is often questionable and misleading. Because these research data form the basis of “scientific” support for fibromyalgia, the objections should be considered carefully and seriously. We outline some of the objection before providing the research data themselves.

1. Research data treat fibromyalgia as a disease associated with at least 11 tender points (ACR 1990 criteria definition), but it is exceedingly unlikely that the observed pathophysiologic abnormalities are confined to greater than or equal to 11 tender points because the body of clinical and epidemiologic evidence does not support a dichotomous condition. It seems likely that observed abnormalities are also found in nonfibromyalgia patients. Studies need to be performed to determine the distribution of the observed abnormalities in pain patients not satisfying the fibromyalgia classification criteria definition.

2. Almost all of the data linking the observed abnormalities to fibromyalgia are correlational, but they are often interpreted causally—a direction of causality that may be wrong. The causal path in fibromyalgia may be complex. All human processes and sensations are expressed biologically. It would be surprising not to find associations.

3. Even assuming causal associations, the explanatory power of these associations have not been described and may be weak. The noted associations do not necessarily predict development of fibromyalgia.

4. The pathogenetic associations attributed to fibromyalgia are noted in other disorders.

5. The literature of fibromyalgia pathogenesis is filled with inadequate proofs because authors have drawn strong conclusions from limited correlative data.

6. Selection of patients and controls can be a problem. Specifically, patients may be too “good” and control subjects represent “healthy controls” rather than other pain patients. Healthy controls will always be different from patients with illnesses.
Muscules and Microtrauma

Originally thought to be important in pathogenesis, muscle and tendon disorders have fallen out of favor because they do not explain adequately the systemic symptoms of fibromyalgia. In addition, changes found in muscle biopsy specimens are non-specific, consistent with many types of muscle damage, ranging from ischemia to simple deconditioning, and are not different from changes found in individuals without fibromyalgia.

Genetic and Familial Factors

Compared with patients with rheumatoid arthritis, fibromyalgia aggregated strongly in families: the odds ratio measuring the odds of fibromyalgia in a relative of a proband with fibromyalgia versus the odds of fibromyalgia in a relative of a proband with rheumatoid arthritis was 8.5. Genetic factors may predispose individuals to fibromyalgia. Patients with chronic widespread pain and fibromyalgia have been found to have low gene expression for the proinflammatory cytokines interleukin-4 and interleukin-10 and reduced levels of serum concentrations compared with controls. These findings might indicate a role for cytokines in the pathophysiology of fibromyalgia or as a sequel of chronic pain and its treatment. However, a study of 31,318 twins in the Swedish Twin Registry suggested that the co-occurrence of FSS in women can be best explained by affective and sensory components in common to all these syndromes, as well as by unique influences specific to each of them, suggesting a complex view of the multifactorial pathogenesis of these illnesses.

Psychosocial Factors

Psychosocial factors, which include reduced education, nonmarried status, lower household income, smoking, and obesity, have been identified in many studies. The chicken or egg question remains. There has been disagreement as to whether psychiatric abnormalities represent reactions to chronic pain or whether the symptoms of fibromyalgia are a reflection of psychiatric disturbance. Psychiatric disorders may interact with the neuroendocrine system as part of a stress reaction. The most common psychiatric conditions observed in patients with fibromyalgia include depression, dysthymia, panic disorder, and simple phobia. In the National Data Bank for Rheumatic Diseases 64% of patients report prior depression, and 8% report mental illness. Fibromyalgia also occurs in patients without significant psychiatric problems, however. Some individuals with fibromyalgia satisfy the American Psychiatric Association criteria for somatoform disorders (DSM 307.80 and 307.89).

Sleep Disturbance

Fibromyalgia patients often report unrefreshing and non-restorative sleep. Electroencephalographic abnormalities initially were thought to play a major role in the pathogenesis of fibromyalgia, but it is now clear that such abnormalities are non-specific findings. Sleep electroencephalographic studies show abnormalities of delta wave or stage 4 sleep by repeated alpha wave intrusion. Similar abnormalities are found in healthy individuals and in individuals with emotional stress, fever, osteoarthritis, rheumatoid arthritis, and Sjogren’s syndrome.

Stress-Related Neuroendocrine Dysfunction

Stress responses and endocrine axes are disturbed in fibromyalgia, but many of these changes are commonly seen in patients who have known external sources of chronic pain. It is unclear whether these endocrine disturbances in fibromyalgia are primary to the disorder or are secondary to the pain or distress associated with fibromyalgia. Patients with fibromyalgia report more past stressful life events and more daily stressful hassles than patients with rheumatoid arthritis or pain-free healthy controls. Similarly, fibromyalgia is associated with increased reports of virus and other infections (Epstein-Barr virus, parvovirus, Lyme disease); hormonal alterations such as hypothyroidism; and catastrophic events where the patient is the victim of actions of others (e.g., war, car accidents) but not natural disaster, preceding fibromyalgia; and a higher frequency of sexual abuse in childhood. Work-related psychologic factors such as work demands and factors such as job control, social support, and psychologic distress are associated with reporting of musculoskeletal pain, particularly when pain is reported at multiple sites.

Primary Neuroendocrine Dysregulation

Primary neuroendocrine dysregulation found in fibromyalgia can be divided into changes in the two major stress systems: the hypothalamic-pituitary-adrenal axis and the autonomous nervous system. In fibromyalgia, almost all hormonal feedback mechanisms controlled by the hypothalamus are disrupted. After stimulation of the hypothalamic-pituitary-adrenal axis with exogenous corticotropin-releasing hormone or by insulin-induced hypoglycemia, an exaggerated pituitary adrenocorticotropic hormone release has been observed with relative adrenal hyporesponsiveness. Serum thyroid hormone levels are normal, but after intravenous injection of thyrotropin-releasing hormone, patients with primary fibromyalgia responded with a significantly reduced secretion of thyrotropin and thyroid hormones. Growth hormone is secreted during stage 4 sleep and is important for muscle repair and strength. Low levels might explain extended periods of muscle pain after exertion in fibromyalgia patients. Serum growth hormone levels and levels of somatomedin C (insulin-like growth factor-I) have often been reported to be low, but results are inconsistent. It is possible that physical deconditioning, related to avoidance of physical activities because of pain, could lead to more fatigue, stiffness, and, via altered growth hormone metabolism, sleep disturbance.

Autonomus Nervous System

Sympathetic function in fibromyalgia patients has been reported as low, normal, or functionally high. There is a derangement of sympathetic tone and reaction in some patients, being high or low, depending on the situation. One
Abnormal Pain Processing

There are major differences between the sexes with respect to analgesic responses, across all animal species. This may explain the decreased pain tolerance in women with fibromyalgia compared with men. Patients with fibromyalgia have reduced pain tolerance to stimuli that are normally not painful such as pressure, heat, and electric pulse, at the classic tender points and control points (alldynia). They also perceive pain as being more intense and extending for a longer time (hyperalgesia).

This abnormal sensory pain processing could be explained by increased pain facilitation and reduced pain-inhibiting mechanisms on the spinal and cerebral levels. Fibromyalgia patients also displayed abnormal temporal summation of pain after a series of thermal stimulations, called “wind-up.”

The concentration of substance P, a neuromodulator of pain, in the cerebrospinal fluid was threefold greater in fibromyalgia patients than in controls. Substance P may play a role in spreading of muscle pain. This elevation of substance P is not specific to fibromyalgia, however, and has been shown in patients with pain due to other causes. Measures of pain intensity in fibromyalgia patients are correlated with levels of metabolites of the excitatory amino acid neurotransmitters glutamate and aspartate. Sensitization of nociceptive neurons in the spinal dorsal horn by hyperexcitable receptors such as the glutamate receptor N-methyl-D-aspartate could be one of the mechanisms responsible for pain in fibromyalgia.

Decreased Pain Inhibition

Pain inhibitory pathways, descending from the cortex, limbic system, hypothalamus, thalamus, and brain stem, modulate the activity of spinal nociceptive neurons. In fibromyalgia patients, regional blood flow seems to be reduced in the most important pain processing areas in the brain, the thalamus and caudatum, compared with controls.

Serotonin is a neurotransmitter in the descending inhibitory pathways that inhibits release of substance P and excitatory amino acids from the terminals of primary afferent neurons. Serotonin also regulates nonrapid eye movement sleep. Low levels of serotonin metabolites have been reported in the cerebrospinal fluid and serum of patients with fibromyalgia and low back pain. Serotonin antibodies are found in fibromyalgia patients four times as frequently as in controls. Although serotonin antibodies have no diagnostic relevance, they could potentially play a role in pathogenesis. The role of serotonin in the pathophysiology of fibromyalgia is unclear. Drugs that affect serotonin metabolism or action do not have a dramatic effect.

Concentrations of enkephalins in the cerebrospinal fluid are roughly twice as high in fibromyalgia and idiopathic low back pain patients, consistent (but not pathognomonic) with the hypothesis that there is increased release of endogenous mu opioid ligands in fibromyalgia, leading to high baseline occupancy of the receptors. This is consistent with the anecdotal clinical experience that opioids are generally ineffective analgesics in patients with fibromyalgia.

CENTRAL NERVOUS SYSTEM (CNS) INVOLVEMENT IN FIBROMYALGIA SYNDROME

Proton Magnetic Resonance Spectroscopy and Functional Brain Imaging in Assessment of CNS Involvement in Fibromyalgia

Why Search in the Brain For an Explanation of the Riddle of Fibromyalgia?

Fibromyalgia is complex and variably expressed but almost always features some degree of pain amplification. Interestingly, this hyperalgesia is not confined to pressure stimuli but also involves heightened responses to heat, noise, and smell, suggesting an important role for central pain processing abnormalities. Although the pathology of fibromyalgia is poorly understood, a growing body of evidence suggests involvement of the CNS. The hippocampus is a brain center that is sensitive to the effects of stress exposure and has been demonstrated to be affected in a variety of disorders that, like fibromyalgia, began with a stressful experience.

Ultimately, there is central sensitization to pain in which low-intensity stimuli in peripheral tissues such as skin and muscle generate an exaggerated nociceptive response that is interpreted centrally as pain. The central mechanisms underlying this amplified pain perception have been explored using a number of advanced imaging techniques that aim to localize and characterize abnormalities in specific areas of the brain called the pain “matrix.”

ADVANCED IMAGING TECHNIQUES

Studies with single photon emission computed tomography, using injected radioactive compounds in the bloodstream that decay over time, have reported an abnormal reduction of regional cerebral blood flow in thalamic and caudate nuclei of patients with fibromyalgia during rest. In two other studies that used functional magnetic resonance imaging, fibromyalgia patients exhibited enhanced responses to painful and nonpainful stimulation in multiple areas of the brain such as the somatosensory cortices, insula, putamen, anterior cingulate cortex, and cerebellum, as compared with healthy control subjects. These findings were consistent with a left shift in the stimulus-response function, which is characteristic of centrally mediated hyperalgesia and reduced noxious threshold to sensory stimuli.

Hippocampus Dysfunction in Fibromyalgia and Neurometabolic Assessment by Proton Magnetic Resonance Spectroscopy

The hippocampus plays crucial roles in maintenance of cognitive functions, sleep regulation, and pain perception, and in studies using single-voxel magnetic resonance
spectroscopy, metabolic dysfunction of the hippocampus was found in fibromyalgia patients. Others found proton magnetic resonance spectroscopy abnormalities at the basal ganglia and the supraventricular white matter and right dorsolateral prefrontal cortex. Gray matter loss in fibromyalgia patients was suggested by magnetic resonance voxel-based morphometric analysis. In this study fibromyalgia patients had significantly less total gray matter volume and showed a 3.3 times greater age-associated decrease in gray matter than healthy controls. The longer the individuals had had fibromyalgia, the greater the gray matter loss, with each year of fibromyalgia being equivalent to 9.5 times the loss in normal aging. In addition, fibromyalgia patients demonstrated significantly less gray matter density than healthy controls in several brain regions including the cingulate, insular, and medial frontal cortices and parahippocampal gyri. In particular, fibromyalgia appears to be associated with an acceleration of age-related changes in the very substance of the brain. Moreover, the regions in which objective changes are demonstrated may be functionally linked to core features of the disorder including affective disturbances and chronic widespread pain.

**Sensory Gating and Reduced Brain Habituation to Somatosensory Stimulation in Patients with Fibromyalgia**

The attenuation effect of the event-related brain responses following stimulus repetition in healthy subjects is a well-known psychophysiologic phenomenon called sensory gating. Montoya and colleagues examined brain activity elicited by repetitive nonpainful stimulation in patients with fibromyalgia in order to determine possible psychophysiologic abnormalities in their ability to inhibit irrelevant sensory information. Their findings suggest that in fibromyalgia patients, there is abnormal information processing, which may be characterized by a lack of inhibitory control to repetitive nonpainful somatosensory information during stimulus coding and cognitive evaluation. These data further extend previous findings of an abnormal brain processing of nonpainful somatosensory information, rather than a generalized information processing dysfunction, in patients with fibromyalgia.

**Deficits of Nociceptive Information Processing**

In this regard, findings of Montoya and colleagues add to a growing literature in which fibromyalgia patients have been shown to have some deficits of nociceptive information processing relative to healthy controls such as enhanced sensitivity to repetitive pain pressure, abnormal maintenance of pain sensations after repetitive thermal stimulation, or deficits in the endogenous pain inhibitory system. In another work Wood and colleagues investigated presynaptic dopaminergic function in six female fibromyalgia patients in comparison with eight age- and sex-matched controls as assessed by positron emission tomography (PET) with 6-fluoro-l-DOPA as a tracer. Their findings indicate a disruption of presynaptic dopamine activity wherein dopamine plays a putative role in natural analgesia. Harris and colleagues demonstrated by PET decreased mu opioid receptor availability in fibromyalgia.

Further, it has been suggested that hyperalgesia and allodynia in fibromyalgia, as well as in other chronic pain states, are the behavioral consequences of central sensitization. Thus it would be possible that the observed disruption of the inhibitory brain mechanism involved in the early processing of non-nociceptive repetitive stimulation might be a further consequence of those neuroplastic changes due to central sensitization associated with chronic pain.

These findings indicate that central factors are important in the processing of pain in people with fibromyalgia. The neuroimaging findings are highly consistent with studies done in pain more generally. These findings suggest that individuals with fibromyalgia have a narrow range of tolerance for pain and perhaps other sensory stimuli, before it becomes noxious.

**MANAGEMENT OF FIBROMYALGIA: RESEARCH STUDIES AND RECOMMENDATIONS**

The value of contemporary treatment can be gauged by review of outcome studies. Fibromyalgia outcome has been the subject of a number of reports, usually in small studies encompassing short periods of time. In general, results of these studies tend to suggest little change in symptoms, suggesting a limited effect of treatment. Most long-term observational studies do not show improvement in fibromyalgia symptoms and outcomes, even when patients are followed in centers with special interest and knowledge of fibromyalgia. In a recent longitudinal study, 1555 patients displayed continuous high levels of self-reported symptoms and distress despite treatment over a mean of 4 years of follow-up. Service utilization (a measure of symptom activity) does not lessen after diagnosis. Benefit of treatment is generally not sustained in long-term randomized clinical trials. These data should be kept in mind when evaluating the results of treatment clinical trials. The null hypothesis for a chronic, painful disorder should not be no short-term treatment effect, but instead no long-term treatment effect. Short-term studies should be regarded with suspicion, and most fibromyalgia studies are short term.

Compliance with treatment is an important problem in fibromyalgia, and in fibromyalgia clinical trials the dropout rate is high. Even when intention-to-treat analyses are performed, the effectiveness of treatment is overestimated. Patients who follow exercise recommendations have better outcomes than patients who do not; however, most patients in clinical practice do not or will not perform aerobic exercises. It is fair to conclude that exercise prescription is often an ineffective recommendation, rather than concluding that it is an effective treatment.

Treatment trials without a true, contemporaneous control group cannot provide meaningful estimates of efficacy because they often exaggerate efficacy. In evaluating study results, the degree of improvement must be examined and the degree of improvement must be clinically meaningful. Even when improvement is clinically meaningful, the
baseline and final outcome values such as values of pain and fatigue must be considered. If the patients are selected for trials in relative (temporary) flare conditions, they may improve “significantly” but still have high levels of the outcome variables at the conclusion of the trial.

Numerous reviews of the short-term treatment in fibromyalgia are available. Most such reviews rely on the concept of efficacy and rank evidence as a function of study quality. One review indicates that “evidence for treatment efficacy was ranked as strong (positive results from a meta-analysis or consistently positive results from more than one randomized controlled trial [RCT]), moderate (positive results from one RCT or largely positive results from multiple RCTs or consistently positive results from multiple non-RCT studies), and weak (positive results from descriptive and case studies, inconsistent results from RCTs, or both).” As noted by these authors, studies are necessary “... to determine whether the improvement is maintained over months or years.” Recent meta-analyses have included measurements of standardized mean differences (effect sizes) but still do not assess long-term benefit.

Still another problem with the interpretation of fibromyalgia studies relates to study scales. Because patients diagnosed as having fibromyalgia have problems with pain, fatigue, cognition, and anxiety and depression, to name some issues in fibromyalgia, studies may select different scales and outcomes according to the interests of the investigators. This leads to problems in comparing study results. In addition, when multiple outcomes and study instruments are selected, frequently studies can show positive results for one outcome and negative results for another. Even when an outcome such as pain is being measured, if there is more than one pain scale, positive results may be found with one pain scale and not with another. Complex scales are also difficult to interpret, as is the case with the commonly used FIQ total scale. This composite summary scale has no simple interpretation: A reader may note an improvement but not have a clear idea of what such improvement means.

From 6750 fibromyalgia patients screened in the National Data Bank for Rheumatic Diseases, the mean (standard deviation) VAS pain and fatigue scores were 6.3 (2.5) and 7.0 (2.5). As an aid in interpreting effect sizes, the following data are presented; assuming a baseline score of 7.0 on a 0 to 10 VAS scale, the following are the effect size, change score, post-treatment score, and percent improvement at the last assessment: 0.3, 0.75, 6.25, 10.7%; 0.4, 1.25, 6.0, 14.3%; 0.5, 1.25, 5.75, 17.9%; 0.6, 1.5, 5.5, 21.4%.

Finally, the main limitations of results and inferences from fibromyalgia clinical trials is that they cannot be extrapolated to patients in practice because of the artificial nature of clinical trials, issues of compliance, and absence of long-term results.

Häuser and colleagues have provided a detailed compendium of the full range of fibromyalgia therapy, citing research evidence and committee recommendations. In making recommendations for therapy, these reviewers also considered costs and adverse effects. Readers should find this review particularly helpful, although they should keep in mind the degree of observed benefit, its persistence, and other issues mentioned earlier.

Diagnosis

Diagnosis may be an important aspect of treatment. Diagnosis fibromyalgia in individuals with short-term stress-related illnesses is harmful and leads to prolonged illness and medicalization. No valid evidence supports the assertion that diagnosis of fibromyalgia in patients with long-term symptoms has a salutary effect. A study of primary care patients in the United Kingdom reported that “… patients who had been diagnosed as having [fibromyalgia] reported higher rates of illness and health care resource use for at least 10 years prior to their diagnosis, which suggests that illness behavior may play a role. … Diagnosis has a limited impact on health care resource use in the longer term, possibly because there is little effective treatment.”

At the patient level, there is no evidence that diagnosis is harmful. Using the diagnostic term in the presence of severe symptoms often makes it easier for physicians and patients to discuss the condition; when fibromyalgia is not diagnosed, patients sometimes ask directly, “Do I have fibromyalgia?” In considering making the diagnosis of fibromyalgia, the physician should consider the following comment by Barsky and Borus: “The hyperbole, litigation, compensation, and self-interested advocacy surrounding the FSS can exacerbate and perpetuate symptoms, heighten fears and concerns, prolong disability, and reinforce the sick role. Excessive medical testing and treatment expose patients to iatrogenic harm and amplify symptoms.” But if fibromyalgia is “diagnosed,” it is important to be clear to the patient that fibromyalgia is a name given to the symptoms, not a cause of the symptoms.

When a fibromyalgia diagnosis is applied to the larger community, rather than at the level of the individual patient, it has been suggested that a virulent idea and a maladaptive social construction of disease such as fibromyalgia can induce and sustain illness in susceptible persons: a psychosomatic meme, acting as a transmissible template. Direct-to-patient advertising and disease mongering by drug companies expand the definition of fibromyalgia and recruit patients to the diagnosis, offering support to this idea.

Education

Education in some reports may have a modest effect on fibromyalgia symptoms such as fatigue, anxiety, and depression but has limited to no effect on pain. What is called education is actually composed of two components—education and rapport or engagement—and it is impossible to distinguish the two components. Most education studies are derived from formal university-based treatment programs; only one study was applicable to clinical practice, and the sample size was too small to evaluate the effect of the intervention in fibromyalgia. All studies had deficiencies in the validity of the control groups; there are no long-term data on the effect of education. Although it is sensible that education should always be part of any treatment program and is part of establishing rapport, its content should depend on the patient, the duration of illness, and the diagnostic label already present. The goal of education is to help the patient understand and manage his or her
symptoms optimally, reduce dependence on the medical system, and work effectively within that system when necessary. There are no data, however, as to whether, within the clinical setting, extensive education is more or less effective than limited education. In a group of 100 consecutive enrollees in a 1.5-day multidisciplinary group outpatient fibromyalgia treatment program, after 30 days a 12.8% improvement was noted in the 78 who completed the study.140

Exercise
Aerobic exercise increases cardiovascular fitness and reduces pain and other fibromyalgia symptoms. In a short-term RCT, exercise improved aerobic performance by 16% and pain by 13%.141 A carefully done, well-powered RCT of a 12-week community-based exercise program compared with relaxation controls showed a 4% difference in FIQ scores at 1 year but nonsignificant changes in McGill pain scores and SF-36 scores.142 At the 12-month follow-up, 38% of subjects in the exercise arm and 22% in the control arm rated themselves much better or very much better. Only 53% of patients attended more than half of the intervention sessions. A follow-up report at 12 months on patients who participated in a 23-week, three-times-per-week exercise program indicated general improvement compared with baseline values.143 The degree of improvement as measured by the FIQ was 5%.

The Cochrane collaboration evaluated 34 studies that included exercise, noting that there is moderate-quality evidence that aerobic-only exercise training at recommended intensity levels has positive effects on global well-being (standard mean difference [SMD], 0.49) and physical function (SMD, 0.66) and possibly on pain. The researchers concluded that “supervised aerobic exercise training has beneficial effects on physical capacity and fibromyalgia symptoms.”144 A noncontrolled study comparing water-based exercise with land-based exercise showed an average 36% reduction in pain.145 Exclusions in this study included 67 for work schedule incompatibility and 32 for nonspecified refusals; 60 patients were randomly assigned, and 52 completed the study.

Practically, the problem with exercise prescription is that it is difficult to get fibromyalgia patients to participate. Exercise may produce “short-term increases in pain and fatigue that should abate within the first few weeks of exercising.”146 but this may be unacceptable to patients in ordinary clinical settings. Even in formal programs, adherence to exercise is poor.147,148 In a 4.5-year follow-up of a randomized trial of exercise, only 20% of patients maintained an adequate physical activity level.149 In the National Data Bank for Rheumatic Diseases from 1999 to 2010, 16% of 3115 fibromyalgia patients noted the following usage in a 6-month period: aspirin, 20.6%; NSAIDs, 55.9%; acetaminophen, 27.6%; strong opioid analgesics, 6.4%; and nonopioid analgesics, 21.5%.150,151 Tramadol use was 15%. Tramadol use remained at 15% in 2010 in the National Data Bank for Rheumatic Diseases. These data, showing the substantial use of NSAIDs, are important because it is often suggested that NSAIDs are ineffective.126

A few analgesic and NSAID treatments have been formally evaluated. Naproxen, 500 mg twice daily (n = approximately 15), which is the only NSAID that has been studied, was indistinguishable from placebo (n = approximately 15) in a controlled clinical trial of relatively young subjects (age 48 years).161 The combination of tramadol and acetaminophen reduced pain 18.5% more than did the use of placebo.162 In this trial, 48% in the active treatment group and 62% of placebo users were noncompleters in this 3-month trial.

Psychotropic Agents
Many drugs that have antidepressant and other psychotropic attributes have been used in fibromyalgia treatment. Such drugs reduce pain centrally, even in the absence of depression, and may be employed at doses that are insufficient to treat depression. Because of the many different trials and classes of drugs studied, meta-analyses have provided a useful overall overview.129,163,164 We summarize the results of Häuser and colleagues.129 In their meta-analysis of 18 RCTs (1427 participants), there was strong evidence for an association of antidepressants with reduction in pain (SMD, 0.43); fatigue (SMD, 0.13); depressed mood (SMD, 26); and sleep disturbances (SMD, 0.32).

The major classes of drugs included tricyclic and tetracyclic antidepressants (TCAs): amitriptyline and nortriptyline; selective serotonin reuptake inhibitors (SSRIs): paroxetine, fluoxetine, and citalopram; serotonin and noradrenaline reuptake inhibitors (SNRs): duloxetine, milnacipran; and monamine oxidase inhibitors (MAOIs): moclobemide and pirlindole. In subanalysis by class, effect sizes for pain reduction were large for TCAs (SMD, 1.64); medium for MAOIs (SMD, 0.54); and small for SSRIs (SMD, 0.39) and SNRs (SMD, 0.36). Similar, although slightly weaker, results are noted with cyclobenzaprine.

Compared with clinical trial results, results in longitudinal studies and clinical practice show marginal effectiveness of tricyclic antidepressants and similar treatments. A high-quality RCT found no difference in the response to amitriptyline and cyclobenzaprine.165

Other Pharmacologic Treatments
On the basis of clinical trial criterion for efficacy, there is no evidence for efficacy of NSAIDs, corticosteroids, benzodiazepine and nonbenzodiazepine hypnotics, guaifenesin,
melatonin, calcitonin, opioids, thyroid hormone, dehydroepiandrosterone and magnesium, or anti–tumor necrosis factor therapy.\textsuperscript{126}

**Nonpharmacologic Treatments**

There is some evidence for efficacy of numerous nonmainstream treatments including strength training,\textsuperscript{127,149,166} and hypnosis.\textsuperscript{167} There is weak evidence for chiropractic, manual, and massage therapy and no evidence of efficacy for tender or trigger point injections or flexibility exercise. Evidence for acupuncture is contradictory,\textsuperscript{168,169} as is evidence for the efficacy of biofeedback\textsuperscript{170–172} and balneotherapy.\textsuperscript{173–175} Local injections in muscular areas of pain are also commonly employed by rheumatologists. The authors surveyed rheumatologists regarding the use of injections and found them to be used frequently, in agreement with others.\textsuperscript{126} Rheumatologists reported that patients “like injections,” but also that the rheumatologists did not know what else to do. A comprehensive review of nonpharmacologic therapies is available.\textsuperscript{176}

**Combination Therapy**

Although most studies reported earlier concern monotherapy, in practice most fibromyalgia treatments combine multiple therapies. Ordinarily these treatment regimens use analgesics, antidepressants, education, and exercise (at least, exercise recommendations). The extent to which several or many therapies is superior to one or few therapies is not clear. But the effect seems small. So one cannot simply add effect sizes of individual therapies to gauge the multitherapy effect.

A meta-analysis of multicomponent treatment in RCTs (at least one educational or other psychological therapy with at least one exercise therapy) included nine RCTs with 1119 patients.\textsuperscript{128} The authors reported: “There was strong evidence that multicomponent treatment reduces pain (SMD, 0.37); fatigue (WMD, 0.85); depressive symptoms (SMD, 0.67); and limitations to health-related quality of life (HRQOL) (SMD, 0.59) and improves self-efficacy pain (SMD, 0.54) and physical fitness (SMD, 0.30) at post-treatment. There was no evidence of its efficacy on pain, fatigue, sleep disturbances, depressive symptoms, HRQOL, or self-efficacy pain in the long term. There was strong evidence that positive effects on physical fitness (SMD, 0.30) can be maintained in the long term (median follow-up 7 months).” Overall, these data indicated increased benefits of multicomponent treatment as defined here, compared with “other” therapies. But the benefit is still modest and cannot be clearly extrapolated to the long term.

**Practical Recommendations in the Approach to a Patient with Fibromyalgia**

The goal of fibromyalgia treatment is to improve the physical and mental health of patients and their quality of life. This goal implies helping patients manage distressing symptoms, but with decreased dependence on the medical care system. There are no studies as to how often the simple recommendations of education, exercise, and limited pharmacologic treatment provide results at an acceptable level of symptoms and functional ability. Data from the National Data Bank for Rheumatic Diseases show, however, that 61% of 3276 fibromyalgia patients observed from 1998 to 2010 were somewhat or very dissatisfied with their health compared with 35% of 24,891 patients with rheumatoid arthritis. These data indicate that contemporary treatment of fibromyalgia is generally unsatisfactory.

This high level of dissatisfaction is reflected in physician and patient interactions. An unknown but probably small proportion of rheumatology experts refuse to accept referral of fibromyalgia patients. A larger proportion is unhappy seeing such patients or is uncomfortable providing care. Patients, sensing this attitude, are equally unhappy with physicians: Patient support groups provide specific advice on finding positive, sympathetic physicians including identifying them by name. Physician behavior results from a general uncomfortableness with illnesses that are often unresponsive to treatment and have strong psychologic and psychosocial components. There is no simple resolution to this problem. Physicians who are unable to provide helpful care to patients with fibromyalgia should make that known to the patients.

Interest in fibromyalgia and drug company support has resulted in extensive studies of treatment,\textsuperscript{127} often with recommendations for treatment.\textsuperscript{127} The practical result of applying recommendations based on short-term clinical trials to an often poorly responsive chronic illness is uncertain because there is as yet no evidence of long-term effectiveness of treatment. In the face of ineffectiveness, treatment recommendations can lead to switching from one therapy to the next and increased medicalization.

In considering fibromyalgia treatment, physicians should determine what resources are available in the community, and whether the resources are effective and helpful. The educational, exercise, and cognitive behavioral therapy programs described in the research studies earlier are often not available to U.S. community physicians. Available programs may or may not be competent, appropriate, or helpful. Pain management programs sometimes mean little more than spinal blocks and “trigger point” injections, and physical therapy referral often results in treatments that are ineffective for fibromyalgia. The referring physicians should investigate the quality and outcomes of referral resources.

Although the common recommendations of education, exercise, and pharmacotherapy are often appropriate, particularly in newly diagnosed cases, patients with established fibromyalgia have often experienced these recommendations and treatments. Whether such treatments have strong evidence for effectiveness or not, as measured by clinical trials, they are often not clinically effective enough, and patients return to the physician for additional suggestions and care.

The European League Against Rheumatism (EULAR) task force points out, on the basis of limited evidence,\textsuperscript{127} and consensus recommendation, that full understanding of fibromyalgia requires comprehensive assessment of pain, function, and psychosocial context. Fibromyalgia should be recognized as a complex and heterogeneous condition where there is abnormal pain processing and other secondary features. Optimal treatment requires a multidisciplinary approach with a combination of nonpharmacologic and pharmacologic treatment modalities tailored according to
pain intensity; function; and associated features such as depression, fatigue, and sleep disturbance in discussion with the patient.

The question arises as to how to approach a resistant patient with fibromyalgia, given the knowledge that after failure with several standard treatments, success with other medications is unlikely. Should the physician simply go from one (dubious) treatment to another? Should the physician use treatments of dubious or uncertain value? The adverse effects of inappropriate or unnecessary treatments are not inconsequential and include dependence, medicalization of common symptoms, overuse of medical care, increased costs, and side effects. One point of importance is that the physician should at least measure pain using a VAS scale. A similar simple measure is available for fatigue. One really cannot know how the patient is doing without such measurements.

The physician must be friendly and interested—a resource the patient can rely on. Testing should be limited and reserved for times when it is truly necessary to investigate comorbid conditions. Comorbid conditions such as arthritis and obesity should be treated because they can contribute to increasing physical and mental symptoms. The worst problem should be identified. Sometimes identifying where the pain problem began can offer clues to appropriate treatment of the coexisting condition. If many fibromyalgia treatments have been tried and have been unsuccessful, it is generally not a good idea to try even more similar, and soon to be unsuccessful, therapies. We often ask patients, “Which treatment has been most helpful?” and suggest (assuming treatment is necessary and helped at all) that they return to that treatment.

There is no blanket rule on the use of opioids. Experience has shown that they often do not truly help and often cause problems. Strong opioids are generally not recommended. However, there are exceptions to this recommendation, however, and physicians should exercise clinical judgment and use opioids when they think such therapy is necessary, provided that appropriate guidelines are followed. “Tender points” never need injection therapy. Painful areas in muscle may respond to local injections of local anesthetics; corticosteroids are never indicated. If injections relieve pain for more than short periods of time, they may represent a reasonable therapy. In illnesses with strong psychosocial components, medically ineffective therapies can result in overall benefit to patients. The circumstances where dubious therapies might be used are limited. The physician should understand clearly why he or she is administering such therapies and what results are anticipated.

Physical Therapy and Spa Treatment

Physical therapy is not recommended because the aerobic exercise required in fibromyalgia does not usually require formal physical therapy and increases medicalization. Because medical therapy is unsatisfactory, patients find their way to alternative therapies. Some of these therapies may be helpful to individual patients such as massage, water therapy, spa treatment, and acupuncture. These therapies tend to have high cost-effectiveness ratios, and the decision to use such therapies is often best left to the patients and the reimbursement authority. That is not to say that such treatments do not help—everything helps—but they do not help often enough and importantly enough, and some decision point is required. One important goal of therapy is to reduce medicalization and increase independence.

In Europe and the Mediterranean a long-standing tradition of spa treatments exists and many U.S. patients, especially those whose parents came from Europe, fly over to be treated. It appears that fibromyalgia patients significantly improve after different spa treatments. In a controlled study in Tunisia there was a significant improvement directly after 2 weeks of treatment and after 3 months regarding general well-being, function, pain, depression, and fatigue in 58 fibromyalgia patients compared with 76 controls. Comparable results were seen in Turkey and the Dead Sea in Israel. Reviews showed good results of spa treatment and hydrotherapy regarding pain, general well-being, and tender points continuing after 14 weeks and a EULAR advisory committee concluded that treatment with hot baths with or without exercises had a good effect in fibromyalgia.

Complementary or Alternative Treatments

There is insufficient evidence on any complementary and alternative medicine or alternative treatment, taken orally or applied topically for fibromyalgia. The small number of positive studies lack replication.

A frustrated physician may not know where to turn next in a nonresponsive patient. Should the patient be referred to a pain clinic? Sometimes such a referral is inevitable. The quality of pain clinics varies, however, and the results in fibromyalgia are often not good. The decision to refer should depend on the experience with the available clinics and the results that they have produced. In some countries, reimbursement authorities limit referrals, providing a cost-effectiveness analysis that may be alien to the physician-patient relationship.

Treatment options sort themselves out over time. Decisions that are difficult resolve. In the end, the physician who provides support and interest is a strong resource and a guide for patients with fibromyalgia, even when medical therapies are limited.

Medicolegal Issues and Fibromyalgia

Frequently, fibromyalgia becomes a medicolegal issue when an individual with fibromyalgia asserts that he or she is unable to work because of fibromyalgia. Because fibromyalgia symptoms are felt only by the patient, there are no objective medical findings to help in the disability assessment. Gaining a disability award is complex, depending on the source of payment (e.g., government vs. private insurance), the physician’s belief and documentation, the availability of legal services, and the impact of the illness on the patient. Various guidelines have been suggested for evaluating disability as they apply to fibromyalgia. Determination of disability does not depend on proving the existence of fibromyalgia.

The second medicolegal issue arises when an individual claims that trauma caused him or her to develop or exacerbate fibromyalgia and that the fibromyalgia is disabling. Although it is proposed that trauma can alter the CNS
(“neural plasticity”) and cause fibromyalgia, the relationship between the severity of trauma and the report of fibromyalgia is weak. There is no way to determine scientifically if trauma causes or caused fibromyalgia. In addition, it is often difficult to establish the severity of the fibromyalgia symptoms. In reality, the relationship between trauma and disability does not require a diagnosis of fibromyalgia because symptom severity and work impairment are important, not the presence or absence of fibromyalgia.

OUTCOME OF FIBROMYALGIA

The outcome of fibromyalgia can be studied in the context of change and level of symptoms, use of services, and work disability. Many studies have addressed the issue of outcome. Some have suggested that “…knowledge of the potential reversibility of the syndrome [is] resulting in improved outcomes” and that “…outcome is good with minimal intervention.” In a prospective study of fibromyalgia patients referred to a specialty clinic, 70 of 82 were reassessed after 3 years. The returnees were generally improved (pain reduced from 6.8 to 5.4 and fatigue reduced from 6.8 to 5.7). The authors concluded that the overall outcome was favorable. In 33 of 51 patients seen 6 to 8 years after initial participation in a fibromyalgia treatment study, pain was reduced from 6.7 to 5.3 and fatigue was reduced from 7.5 to 6.5. The authors concluded that the results of these returnees suggest a benign long-term outcome in patients with fibromyalgia. A six-center, 7-year study of 538 patients noted that, “Although functional disability worsened slightly and health satisfaction improved slightly, measures of pain, global severity, fatigue, sleep disturbance, anxiety, depression, and health status were markedly abnormal at study initiation and were essentially unchanged over the study period. Half the patients were dissatisfied with their health, and 59% rate their health as fair or poor.” In one report of 45 of 70 patients who had participated in a 3-week trial 6 years earlier, symptoms of fibromyalgia persisted over 6 years. A study of prediagnosis and postdiagnosis use of services found that no changes in the high-use rates were seen over time.

In a longitudinal study of 1555 fibromyalgia patients during 7448 semiannual observations for up to 11 years, there was minimal improvement in symptoms. The SMDs (improvement effect sizes) between start and study completion were patient global, 0.03; pain, 0.22; sleep problems, 0.20; SF-36 PCS, 0.11; SF-36 Mental Component Summary, 0.03; and EuroQoL (EQ-5D), 0.10. These data suggested that the course of fibromyalgia was one of continuous high levels of self-reported symptoms and distress despite available treatments.

A study of 27 of 48 (56%) patients had a 2-year follow-up. In general, the patients showed no improvement in their symptoms over the observation period, regardless of the type of therapy they had received. General satisfaction with quality of life improved, as did satisfaction regarding health status and the family situation, although the degree of pain experienced remained unchanged. In comparison with the initial examination, there was no change in either work capacity or disability-pension status. Taken as a whole, although some patients improve, the data tend to suggest minimal improvement in most cases despite treatment. Even among the positive studies cited, the degree of improvement is small. These data, which are representative of the actual outcome of fibromyalgia patients in practice, provide a more realistic evaluation of treatment effect than the assessments based on clinical trials.

Selected References


References

25. Illich I: Limits to medicine: medical nemesis, the expropriation of health, New York, 2000, Marion Boyars Publishers Ltd.
38. Crofford LJ, Clauw DJ: Fibromyalgia: where are we a decade after the American College of Rheumatology classification criteria were developed? Arthritis Rheum 46(5):1136–1138, 2002.


Analysis of synovial fluid and synovial tissue obtained from diseased joints provides important diagnostic information in specific clinical settings, and is valuable in addressing a spectrum of research questions aimed at enhancing our understanding of the pathogenesis and mechanisms of rheumatic diseases. Many peripheral joints are readily accessible to sampling of both synovial fluid effusions and synovial tissue, although the knee is the most frequently sampled joint. The techniques used to obtain and analyze synovial fluid and tissue samples are discussed in this chapter.

SYNOVIAL FLUID ANALYSIS

Synovial Fluid in Health

Under normal conditions, a small volume of synovial fluid is present in each joint, forming a thin interface between the surfaces of the articular cartilage, and providing for friction-free movement of these surfaces. In a large joint such as the knee, the volume of synovial fluid is estimated to be less than 5 mL. Moreover, intra-articular pressure is typically subatmospheric. Compositionally, normal synovial fluid is an ultrafiltrate of plasma to which proteins and proteoglycans are added by fibroblast-like synoviocytes in the lining layer. Most of the small-molecular-weight solutes such as oxygen, carbon dioxide, lactate, urea, creatinine, and glucose diffuse freely through the fenestrated endothelium of the synovium and are normally present at levels comparable with those of plasma. Evidence for active transport of glucose has been found. The total protein concentration of normal synovial fluid is 1.3 g/dL. The concentration of individual plasma proteins is inversely proportional to the molecular size, with small proteins such as albumin present at approximately 50% of plasma levels, and large proteins such as fibrinogen, macroglobulins, and immunoglobulins present at low levels. It should be noted that in contrast to this selective entry on the basis of size, clearance of synovial fluid proteins through the synovial lymphatics is unrestricted by size. Hyaluronan is the major proteoglycan synthesized by synovial cells and secreted into synovial fluid. Hyaluronan is highly polymerized and reaches molecular weights exceeding one million daltons, giving this fluid its characteristic viscosity. The hyaluronan also acts to retain small molecules in the synovial fluid. The lubricating capacity of the synovial fluid is attributed to a glycoprotein called lubricin. This molecule has been fully characterized on the basis of the study of individuals with mutations of the PRG4 gene, which encodes for its production. These mutations result in an autosomal recessive loss-of-function disorder called the camptodactyly–arthropathy–coxa vara–pericarditis syndrome, which features a progressive, noninflammatory arthropathy characterized by severe cartilage destruction associated with proliferation of synovial lining cells. The role of lubricin in maintaining the health of the cartilage has been further demonstrated in a murine knock-out model.
Accumulation of Synovial Effusions

Synovial fluid and its contents are cleared through the synovial lymphatics through a process that is aided by joint motion. Excess fluid can accumulate in any diarthrodial joint as a result of a broad range of processes, including noninflammatory, inflammatory, and septic disorders. In addition, overt hemarthroses can result from both traumatic and nontraumatic disorders. The most important mechanism contributing to the accumulation of joint effusions is an increase in synovial microvascular permeability. This allows for an increase in the eflux of plasma proteins, particularly larger proteins, which in turn increases osmotic pressure and contributes to the effusion. Leukocytes accumulate in the fluid after transmigration through the endothelium, stimulated by chemokines produced in the synovium. The capacity of synovial lymphatics to clear proteins, cells, and debris is rapidly exceeded, which in turn contributes to their accumulation in the synovial compartment.

Arthrocentesis

Most peripheral joints are readily accessible for diagnostic arthrocentesis, and the procedure can be performed in almost any ambulatory care setting equipped for sterile procedures. Joints that are less accessible because of a deeper location, such as the hip, may require an imaging technique that uses fluoroscopy or ultrasound to guide the needle and ensure accurate placement. Details of techniques used for arthrocentesis are described in Chapter 54. Because the ease with which joint fluid is aspirated depends on the gauge of the needle that is used, it is important to attempt arthrocentesis with a needle of adequate gauge, particularly in the larger joints. Moreover, high suction gradients created by large syringes should be avoided, because they may actually reduce the capacity to successfully aspirate synovial fluid. Difficulty in aspirating synovial fluid may relate to a number of intra-articular factors, including viscosity, the presence of debris such as rice bodies, and loculation of fluid into inaccessible areas. Instillation of a small amount of sterile saline may assist in obtaining enough fluid for culture in situations where infection is highly suspected, yet direct aspiration is difficult.

Once obtained, it is important to analyze aspirated synovial fluid samples as quickly as possible to avoid spurious results. In particular, leukocyte count and differential ideally should be performed on fresh specimens. If the specimen cannot be analyzed quickly and short-term storage is needed, the specimen should be kept at 4°C, and an aliquot preferably placed in ethylenediaminetetraacetic acid (EDTA) to prevent clotting. Delays in analysis beyond 48 hours should be avoided. A simplified algorithm for analyzing synovial fluid samples is shown in Figure 53-1.

Gross Examination

First impressions regarding the nature of the synovial fluid occur as fluid enters the syringe during the arthrocentesis procedure itself. For example, the viscosity of the fluid is readily appreciated during this step. As has been mentioned, normal synovial fluid is highly viscous because of its hyaluronan content and forms a long string when a drop is expressed from the end of the needle. With increasing levels of inflammation associated with recruitment and activation of leukocytes in the synovial cavity, the hyaluronan is digested, resulting in loss of viscosity that is appreciated as a reduction in the “stringiness” of the fluid. Large pieces of debris such as rice bodies, thought to arise from detached ischemic synovial villi, may be visible as they are aspirated.

---

**Figure 53-1** A simplified algorithm for analyzing synovial fluid samples and initiating a plan of management. CPPD, calcium pyrophosphate dihydrate; NSAID, nonsteroidal anti-inflammatory drug; WBC, white blood cell.
These can cause sudden arrests in the flow of fluid in the syringe, requiring manipulation and redirection of needle placement.

Inspection of the aspirated synovial fluid can yield other important diagnostic information. For example, floridly purulent fluid will be completely opaque because of the very high number of leukocytes present, but synovial fluid that is transparent, to the point where printed text can be read through it, is seen in noninflammatory settings. Inflammatory synovial fluid, as would be aspirated from an individual with active rheumatoid arthritis (RA), appears cloudy and translucent; the degree of translucency depends on the intensity of the inflammatory response and the concentration of leukocytes in the sample. Synovial fluid from patients with ochronosis may have a speckled appearance, and particulate debris from joint prostheses may be visible on gross inspection.

During the arthrocentesis procedure, an important challenge may be to determine whether the presence of blood in the aspirated synovial fluid indicates a hemarthrosis or, alternatively, is a result of trauma from the procedure itself. In the latter case, the blood may remain unmixed with the synovial fluid, appearing as red streaks in an otherwise yellow fluid, but in the case of hemarthroses, the synovial fluid is generally homogeneously bloody and does not form a clot. The causes of frank hemarthrosis are varied and include trauma, pigmented villonodular synovitis, tumors, hemophilia and other bleeding disorders or anticoagulant therapy, Charcot joint, and occasionally intense inflammation from a chronic arthropathy such as RA or psoriatic arthritis.

### Leukocyte Count

Analysis of leukocyte counts and cytology provide important diagnostic information regarding the cause of a synovial effusion (Table 53-1). A fresh specimen should be placed in a heparinized tube for rapid analysis, and if the fluid is particularly viscous, it may need to be diluted in normal saline before counting. Normal synovial fluid contains fewer than 180 nucleated cells/mm³, most of which originate as desquamated synovial lining cells. The leukocyte count broadly classifies synovial fluids as noninflammatory (<2000 cells/mm³), inflammatory (2000 to 50,000 cells/mm³), and septic (>50,000 cells/mm³). It should be kept in mind that these definitions provide broad guidelines to help narrow the differential diagnosis rather than representing inherent biologic properties of the fluid.

The most common causes of noninflammatory synovial fluids are mechanical derangements of the joint and osteoarthritis. Other causes include endocrinopathies such as acromegaly and hyperparathyroidism; inherited disorders such as ochronosis, hemochromatosis (which can also present with hemarthrosis), Ehlers-Danlos syndrome, Wilson's disease, and Gaucher's disease; acquired disorders such as Paget's disease, avascular necrosis, and osteochondritis dissecans; and an uncommon condition called intermittent hydrarthrosis, in which joints become effused in a cyclic manner. At the other extreme, leukocyte counts of 50,000 to 300,000 cells/mm³ are most commonly associated with septic arthritis and should prompt the clinician to empirically treat the individual as such until this diagnosis is excluded with a high degree of certainty, which typically requires definitive culture results, and possibly repeat aspiration. It should be added that leukocyte counts exceeding 50,000 cells/mm³ are not uncommonly seen in acute crystal-induced arthritis, particularly gout. Inflammatory cell counts between 3000 and 50,000 cells/mm³ are seen in a wide spectrum of articular disorders, including many cases of septic arthritis. Thus most patients with acute attacks of gout and pseudogout, active RA, reactive arthritis, and psoriatic arthritis, as well as patients with gonococcal arthritis and other nonpyogenic forms of septic arthritis, will typically present with synovial fluid cell counts in this range (see Table 53-1).

### Synovial Fluid Cytology

Characterization of the cells present in synovial fluid is an important diagnostic step that can be achieved initially by performing cytology on a wet mount of the synovial fluid. To perform the wet mount analysis, a single drop of synovial fluid is placed on a clean glass slide, which then is covered by a coverslip and is examined under low- and high-power

---

**Table 53-1 Characteristics of Synovial Fluid**

<table>
<thead>
<tr>
<th></th>
<th>Appearance</th>
<th>Viscosity</th>
<th>Cells per mm³</th>
<th>% PMNs</th>
<th>Crystals</th>
<th>Culture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Transparent</td>
<td>High</td>
<td>&lt;200</td>
<td>&lt;10%</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>Transparent</td>
<td>High</td>
<td>200-2000</td>
<td>&lt;10%</td>
<td>Occasional calcium pyrophosphate and hydroxapatite crystals</td>
<td>Negative</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Translucent</td>
<td>Low</td>
<td>2000-50,000</td>
<td>Variable</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
<td>Translucent</td>
<td>Low</td>
<td>2000-50,000</td>
<td>Variable</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>Reactive arthritis</td>
<td>Translucent</td>
<td>Low</td>
<td>2000-50,000</td>
<td>Variable</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>Gout</td>
<td>Translucent to cloudy</td>
<td>Low</td>
<td>200-50,000</td>
<td>&gt;90%</td>
<td>Needle-shaped, negatively birefringent monosodium urate monohydrate crystals</td>
<td>Negative</td>
</tr>
<tr>
<td>Pseudogout</td>
<td>Translucent to cloudy</td>
<td>Low</td>
<td>200-50,000</td>
<td>&gt;90%</td>
<td>Rhomboid, positively birefringent calcium pyrophosphate crystals</td>
<td>Negative</td>
</tr>
<tr>
<td>Bacterial arthritis</td>
<td>Cloudy</td>
<td>Variable</td>
<td>2000-50,000</td>
<td>&gt;90%</td>
<td>Negative</td>
<td>Positive</td>
</tr>
<tr>
<td>PVNS</td>
<td>Hemorrhagic or brown</td>
<td>Low</td>
<td>—</td>
<td>—</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>Hemarthrosis</td>
<td>Hemorrhagic</td>
<td>Low</td>
<td>—</td>
<td>—</td>
<td>Negative</td>
<td>Negative</td>
</tr>
</tbody>
</table>

PMNs, polymorphonuclear neutrophils; PVNS, pigmented villonodular synovitis.
light microscopy. In addition to leukocytes, and in the case of traumatic taps or hemarthroses, or large numbers of erythrocytes, wet mount may reveal the presence of clumps of fibrin and crystals, cartilage and synovium fragments, and lipid droplets. These can all appear as amorphous material, and care should be taken to avoid assuming their composition without further characterization.

Characterization of synovial fluid leukocytes is best achieved by staining a dried smear of the fluid. Wright stain is most commonly used for this purpose. The phenotype and morphology of the leukocytes can then be assessed under high power using oil immersion. Septic range synovial fluid containing more than 50,000 cells/mm³ is almost always associated with a high preponderance of polymorphonuclear leukocytes, often greater than 90%. Monocytes and lymphocytes predominate in the synovial fluid of patients with viral arthritis, lupus, and other connective tissue diseases. Synovial fluid samples from patients with active RA, reactive arthritis, psoriatic arthritis, and acute attacks of crystal-induced arthritis typically demonstrate a preponderance of polymorphonuclear leukocytes, although early RA fluids may have a low leukocyte count with primarily mononuclear cells. The presence of large numbers of "ragocytes," which are granulocytes that have phagocytized immune complexes, is associated with active RA, and their presence may indicate an unfavorable prognosis in this disease. Reiter's cells represent cytophagocytic mononuclear cells that have phagocytized apoptotic polymorphonuclear leukocytes, this possibly representing a pathway by which autolysis and release of damaging mediators from the latter cells are avoided. The presence of Reiter's cells is not specific for reactive arthritis, nor indeed for spondyloarthropathies in general. Occasionally, eosinophils will be seen to predominate in the synovial fluid. This may be associated with parasitic infection, urticaria, or hypereosinophilic syndrome. It has been suggested that cytocentrifugation of synovial fluid is the optimum method for performing cytopathology, although the cost-effectiveness of this technique is questionable in most clinical settings.

**Wet Smear Analysis by Polarized Microscopy**

A search for crystals using polarized microscopy is particularly valuable in the diagnosis of acute monoaahritis or oligoarthritis, in which gout and pseudogout are often on the differential diagnosis. In such a clinical situation, if indeed the absence of pathogenic crystals in synovial fluid can be established, the likelihood of septic arthritis increases, prompting the initiation of intravenous antibiotics and potentially necessitating a hospital admission. Thus, the rapid and accurate diagnosis of a crystal-induced process can serve to prevent a costly and unnecessary sequence of events. It is helpful if the individual or the team that performs the arthrocentesis is also in a position to rapidly examine the specimen by polarized microscopy. This requires the availability of a functional polarizing microscope, as well as adequate operator experience in the identification of crystals using this technique. This is particularly important in the case of calcium pyrophosphate crystals, which are notoriously difficult to detect.

Care should be taken to make sure that the slide and the coverslip are free of dust, talc, and other particulate matter. Crystals present in the specimen rotate the light in such a way that they appear as bright objects in an otherwise dark field. It should be noted that birefringent material in the specimen appears as a bright yellow or blue color in the red field generated by the first-order compensator. As birefringent crystals are rotated relative to the axis of the first-order compensator, the color changes from yellow to blue or vice versa. Crystals that are yellow when oriented parallel to the axis of the compensator are negatively birefringent, and those that are blue are positively birefringent.

Identification of crystals in synovial fluid is greatly facilitated by detailed examination of the specimen, under both low and high power, using the approach previously described. A combination of morphology and birefringence serves to identify the crystals. Monosodium urate (MSU) crystals, as shown in Figure 53-2, are easily identifiable because the crystal load is typically high during an acute attack of gout. A good degree of concordance between laboratories in the identification of MSU crystals has been shown. These crystals appear as strongly negatively birefringent needle-shaped objects, many of which are intracellular, having been phagocytized by synovial fluid leukocytes. In contrast, calcium pyrophosphate dihydrate (CPPD) crystals seen during attacks of pseudogout tend to be smaller, rhomboid-shaped objects that are weakly positively birefringent, as shown in Figure 53-3. Because the CPPD crystal load during an attack of pseudogout tends to be relatively low, and because CPPD crystals are only weakly birefringent, it is important to examine all areas of the specimen on the microscope slide, and possibly to prepare a second wet mount to exclude or confirm this diagnosis. Concordance between laboratories in the recognition of CPPD has been shown to be substantially lower than in the case of MSU crystals. A particularly challenging situation arises when intracellular crystals cannot be identified, yet birefringent extracellular objects resembling crystals are seen scattered throughout the slide. This may be caused by powder from gloves or dirt on the slides.

As with other analyses on the synovial fluid, wet mount preparation and analysis should be performed as quickly as possible, although identification of crystals can still be successful after prolonged storage of specimens. The crystal load decreases substantially as the acute inflammatory attack

**Figure 53-2** A, Urate crystals in the tophus from a patient with gouty arthritis. Crystals are negatively birefringent and needle shaped. B, Intracellular urate crystal as seen on Wright stain. (Courtesy H. Ralph Schumacher, Jr.)
Corticosteroid crystals can be highly birefringent and mimic urate or CPPD crystals. Large amounts of lipid in the synovial fluid can be visible on gross examination. The significance of these crystals in synovial fluid is unclear, but it is unlikely that they are pathogenic in most cases.

Detection of Microorganisms by Gram Stain, Culture, and Polymerase Chain Reaction Analysis of Synovial Fluid

A wide spectrum of organisms can cause septic arthritis, although the most common pathogens are gram-positive bacteria such as staphylococci and streptococci. Because septic arthritis causes rapid destruction of the joint, and because it can spread hematogenously to other areas and is associated with significant mortality, it is imperative that a specific diagnosis be made as quickly as possible, and that empiric therapy with broad-spectrum antibiotics be instituted until this diagnosis can be confirmed or excluded.

A Gram stain performed on fresh synovial fluid will identify an organism in an estimated 50% of cases of septic arthritis, with highest sensitivity for gram-positive organisms. Moreover, the specificity of a positive Gram stain approaches 100%. Clearly this indicates that the positive predictive value for the Gram stain is very high, and that the negative predictive value is substantially lower. The gold standard for diagnosing septic arthritis remains bacteriologic culture, which has a sensitivity of 75% to 95% and a specificity of 90% in cases of nongonococcal septic arthritis. It has been shown that the use of blood culture bottles further increases the yield of positive synovial cultures. It is important to note that bacteriologic cultures declines dramatically after antibiotic therapy is instituted, it is important that the clinician perform arthrocentesis before any antibiotics are administered. Cultures...
should be performed even when uric acid or other crystals are demonstrated in the synovial fluid, because it has been shown that gout and septic arthritis can coexist. In the case of gonococcal arthritis, the sensitivity of bacteriologic culture, even if performed on a sample collected using appropriate media, is low and is estimated to be less than 10%.

Polymerase chain reaction (PCR) carries a high degree of sensitivity and specificity for the detection of microorganisms in synovial fluid and tissue, even in individuals who are culture negative. Most bacteria can be detected on the basis of amplifying specific sequences in their ribosomal RNA (16S rRNA). PCR has been shown to be the procedure of choice for making the diagnosis of gonococcal arthritis and is a highly sensitive and specific method of detecting tuberculous arthritis, although as discussed later, analysis of synovial tissue is better than analysis of synovial fluid for making this diagnosis. PCR has also been proposed as a method of verifying the successful elimination of the offending organism in cases of septic arthritis.

The sensitivity and specificity of PCR in detecting synovial microorganisms need to be balanced against the biologic significance of a positive test. Contaminants are easily detected using this method, and highly stringent conditions for sample collection are required to prevent false-positive tests. Moreover, PCR studies of synovial fluid and tissue from a spectrum of chronic forms of arthritis, including RA, osteoarthritis, reactive arthritis, and undifferentiated arthritis, have indicated the presence of microorganisms in a significant number of specimens. The biologic significance of these findings and the potential role of bacterial DNA or cell wall fragments in the pathogenesis of these arthropathies remain unclear.

Biochemical Analysis of Synovial Fluid
A number of widely available biochemical tests may add to the diagnostic impression of aspirated synovial fluid samples, although lack of specificity of these biochemical analyses tends to limit their value. Testing for synovial fluid glucose, protein, and lactate dehydrogenase (LDH) has long been included in routine practice, and values obtained should be compared with serum values. Samples from septic arthritis typically exhibit very low glucose, low pH, and high lactate levels; these levels are indicative of a switch to anaerobic metabolism. Highly inflammatory synovial fluids from RA patients exhibit a similar profile, along with high protein and LDH levels. Levels of pressure of oxygen in the blood (pO_2) are often in the hypoxic range in RA synovial fluids, and are correlated with increased lactate and levels of pressure of carbon dioxide in the blood (pCO_2). A prospective study conducted to evaluate these tests in a spectrum of inflammatory and noninflammatory disorders demonstrated considerable variability in each diagnostic category, which limits their clinical utility.

Serologic testing of synovial fluid to detect rheumatoid factor, antinuclear antibodies, and complement levels has been suggested as a method that can be used to confirm a diagnosis of RA or other connective tissue diseases. In particular, RA synovial fluids may be positive for rheumatoid factor, even when serum is not, and complement levels are typically low as a result of consumption by immune complexes. These findings are of insufficient sensitivity and specificity to be of value on a routine clinical basis.

Synovial Fluid Analysis in Arthritis Research
The ease with which synovial fluid is aspirated from effused joints has allowed a wide spectrum of research studies to be conducted on this biologic material. In research settings, cells in synovial fluid samples are typically separated by centrifugation, and cellular and noncellular components of the fluid are analyzed separately. Detailed analysis of the phenotype and functional properties of synovial fluid leukocytes has been particularly informative in RA and reactive arthritis research, where immunophenotyping of lymphocyte subpopulations has provided important clues to the pathogenesis of these diseases. In the case of reactive arthritis, in which triggering organisms are often identified, the proliferative and cytokine responses of synovial fluid lymphocytes to antigens derived from Chlamydia, Yersinia, and other pathogens have been elucidated. It has generally been shown that synovial fluid T cells from reactive arthritis patients are biased toward production of T helper (Th)2 cytokines such as interleukin (IL)-10 and IL-4, whereas synovial fluid T cells from RA patients are Th1 biased and exhibit defects in Th2 differentiation.

Analysis of the noncellular portion of synovial fluid has provided important information regarding a spectrum of soluble molecules, including cytokines and growth factors, extracellular matrix proteins, autoantibodies, and therapeutic drug levels. Moreover, broad-based proteomic studies of synovial fluid using fractionation techniques and mass spectrometry are beginning to provide novel approaches to understanding pathogenesis and prognosis in arthropathies such as RA.

SYNOVIAL BIOPSY

Sampling of synovial tissue is a direct approach to defining the pathologic processes that cause joints to be swollen and painful. In clinical settings, it can be particularly valuable in evaluating an undiagnosed persistent monoartthritis when other investigations, including synovial fluid analysis, have failed to provide a specific diagnosis. In research settings, analysis of synovial tissue samples has dramatically improved our understanding of the pathogenetic mechanisms underlying RA, spondyloarthropathies, and other chronic articular disorders. More recently, synovial biopsy has been explored as a method for defining the target tissue response to therapeutic agents, particularly targeted biologic therapies.

Blind Percutaneous Synovial Biopsy
Percutaneous needle biopsy is most commonly performed according to the method originally described by Parker and Pearson, utilizing a biopsy needle that now carries their name. Percutaneous synovial biopsy is most often performed on the knee joint, although the technique can readily be adapted for use in other joints such as wrist, elbow, ankle, or shoulder. A modification of the original Parker-Pearson needle has facilitated synovial biopsy of small hand joints such as metacarpophalangeal and proximal interphalangeal joints. The technique for Parker-Pearson synovial biopsy
uses a 14-gauge needle with a lateral aperture just proximal to the inserted end of the needle. This lateral opening features a sharp cutting edge for severing trapped synovial tissue that is captured by applying suction with a 3- to 5-mL syringe. With this approach, multiple 1- to 3-mm samples are obtained by angling the trocar in several directions. This also serves to minimize the sampling error involved. Synovial samples are typically pink and are easily removed with a slight twisting motion. Because of the blind nature of the procedure, samples of fat, muscle, or fibrous tissue may be obtained and need to be separated from true synovial samples.

Percutaneous synovial biopsy is easily performed in most ambulatory care settings with the use of relatively inexpensive equipment. The overall morbidity of the procedure is low and is comparable with that of arthrocentesis, with perhaps a slightly higher rate of hemarthrosis, which can be minimized if the patient does not bear weight for a few hours after the procedure. The main disadvantage of the procedure is its blind nature. In comparison with visually guided arthroscopy, samples derived from the interface between synovium and adjacent cartilage are under-represented when the blind procedure is used. 38,39 As is discussed later, this drawback is particularly relevant to a number of research questions.

Arthroscopically Guided Synovial Biopsy

Arthroscopy is widely used by orthopedic specialists for the diagnosis and treatment of a variety of articular disorders, particularly mechanical derangements of intra-articular structures such as cruciate ligaments and menisci. Over the past two decades, the arthroscopic procedure has been adapted for acquiring diagnostic synovial biopsies in settings that do not require a fully equipped operating theater and general anesthetic. In most cases, intra-articular local anesthesia suffices for the procedure, although conscious sedation may be required in some individuals. The procedure is well tolerated and is associated with low morbidity, although the risks of hemarthrosis and infection after the procedure are slightly higher than that of percutaneous needle biopsy. The patient should be instructed to minimize weight bearing for 24 to 48 hours after the procedure.

The primary advantage of arthroscopy is its ability to visually guide the biopsy procedure. This permits macroscopic evaluation of the synovium and sampling of areas that appear to be particularly severely affected by the pathologic process, and it allows for sampling of the interface between inflamed synovium and adjacent cartilage, this being an area of particular interest for understanding the pathogenesis of destructive arthropathies such as RA. 38 As with samples obtained by percutaneous synovial biopsy, individual samples are allocated for specific laboratory studies depending on the clinical or research question being addressed.

Processing Synovial Tissue Samples

In all cases, an adequate number of individual synovial specimens need to be allocated for routine light microscopy with the use of formalin fixation and paraffin embedding. This provides the highest-quality sections for hematoxylin and eosin (H&E) histologic analysis, and it allows the most accurate delineation of pathologic processes within tissue. Although formalin-fixed sections can be used in some cases for immunohistology, formalin fixation alters the conformation of many protein antigens, making them inaccessible for specific identification by immunohistology. Many of the molecular markers used to analyze diseased synovium, including cell surface markers, cytokines, adhesion molecules, and proteases, require that tissue samples are snap frozen in a suitable mounting medium such as optimal cutting temperature compound, and then are sectioned with the use of a cryostat. The sections can be processed using antigen-specific monoclonal or polyclonal antibodies and color development achieved by one of several available immunofluorescence or immunoperoxidase methods. Typically, a nuclear counterstain is also used to assist in orientation of the tissue—hematoxylin in the case of immunoperoxidase studies. If only formalin-fixed, paraffin-embedded tissue is available, an alternative method for detecting antigens that are sensitive to formalin fixation is antigen retrieval.

Several antigen retrieval methods are available, including enzymatic and thermal methods, 40 which have been used to successfully retrieve a spectrum of antigens from archival synovial tissue samples for immunohistologic studies, although the quality of the tissue sections often deteriorates after antigen retrieval. A number of double-staining immunohistology techniques have been developed for simultaneous evaluation of the expression of two markers in the same tissue section, although these techniques are labor intensive and often require considerable experimentation to generate good stains. 41 It should be noted that formalin fixation dissolves crystals, and if this is a diagnostic consideration, the specimen should be fixed in ethanol.

The sensitivity and specificity of molecular DNA and RNA techniques provide unprecedented opportunities for exploring the pathogenesis of synovial disorders. Although these studies can be carried out on very small quantities of tissue, great care needs to be taken in handling and processing tissue samples to prevent degradation of the nucleic acids, particularly in the case of RNA when RNAase enzymes are ubiquitous and can rapidly degrade the small quantity of RNA present in a tissue sample. As is discussed later, the search for microbial DNA and RNA has been of particular interest in attempts to understand the cause and pathogenesis of reactive arthritis, rheumatoid arthritis, and other forms of chronic synovitis of unknown cause. Techniques used to analyze human gene expression in small tissue samples have been rapidly improved. This has enabled the detection and quantitation of multiple mRNA transcripts in very small quantities of biopsy material, in many cases without the need for amplification. 42,43

SYNOVIAL PATHOLOGY

Synovial Membrane in Health

A detailed description of the composition of normal synovium is provided in Chapter 2. Histologically, the normal synovial lining layer is one to three cells thick and is composed of closely associated macrophage-like (type A) and fibroblast-like synoviocytes (type B) that are
Synovial Histopathology in the Evaluation of Monoarthritis

Pathologic analysis of synovial tissue samples can be of considerable value in certain clinical settings. Having said this, it should be kept in mind that the histopathologic interpretation of synovial biopsy specimens is often nondiagnostic and lacking in specificity.\(^4\) Pathologic analysis of synovial samples from patients with undiagnosed monoarthritis may be of particular value. The presence of large numbers of neutrophils in the synovial tissue stroma is highly suggestive of septic arthritis, and in such cases Gram stain may reveal bacteria in the tissue. Because septic arthritis is usually acute in onset, synovial biopsy is rarely required, and the diagnosis can be made by analyzing synovial fluid as described previously. Gonococcal arthritis may require synovial biopsy for diagnosis (Figure 53-7). A mononuclear cell infiltrate, on the other hand, is more consistent with a chronic inflammatory process and has a wide differential diagnosis, as has been described. The presence of granulomas supports a diagnosis of tuberculous arthritis or sarcoidosis, both of which cause chronic monoarthritis. The synovial granulomas of tuberculosis (TB) may be caseating or noncaseating, and staining of the tissue for acid-fast bacilli, culture, and molecular probing can yield a definitive diagnosis in an estimated 50% of cases. Similarly, a spectrum of fungal infections can be diagnosed using similar approaches, but special stains such as Gomori may be required. The diagnosis of sarcoid arthropathy is suspected in synovial specimens with noncaseating granulomas in cases where mycobacterial or fungal infection has been excluded.

Pigmented villonodular synovitis is an important consideration in individuals with chronic monoarthritis of a large joint such as the knee or hip. This disorder has a characteristic magnetic resonance imaging (MRI) appearance caused by hemosiderin deposits in the synovium and large cystic lesions in adjacent bone. Histopathologic analysis of the synovium can confirm this diagnosis and demonstrates a diffusely hypervascular proliferative lesion with mononuclear cells of the monocyte/macrophage lineage, foamy multinucleated cells resembling osteoclasts, and hemosiderin deposits (see Figure 53-7). Synovial sarcomas are rare tumors that must be diagnosed on the basis of synovial pathology.

Synovial Histopathology in the Evaluation of Polyarthritis

In current clinical practice, the availability of well-validated diagnostic criteria and specific serologic tests, combined with a relative lack of specificity in synovial histopathologic features, limits the clinical utility of synovial pathology in the differential diagnosis of oligoarthritis and polyarthritis. On
the other hand, analysis of synovial tissue samples obtained in the context of research studies from patients with RA and various spondyloarthropathies has dramatically enhanced our understanding of the cellular and molecular mechanisms of these disorders. This is reflected in a large body of literature published over the past three decades.38,48 RA synovium has been the most extensively studied histopathologically, and a detailed discussion of RA synovitis can be found in Chapter 69. The two characteristic features seen in RA synovitis are hyperplasia of the lining layer and infiltration of the sublining stroma with mononuclear cells (Figure 53-8). The surface of the lining layer

Figure 53-7  A, Synovial pathology of gonococcal arthritis. A marked infiltrate with polymorphonuclear leukocytes and vascular congestion is present. B, Synovial pathology of scleroderma shows loss of the lining layer with surface fibrin deposition, and mononuclear inflammation in sublining areas. C, Pigmented nodular synovitis with hemosiderin deposits and foamy cells. D, Amyloidosis with deposits on the synovial surface, Congo red stain. (A-D, Courtesy H. Ralph Schumacher, Jr.)

Figure 53-8  Histopathology of rheumatoid arthritis synovitis. A, Lymphoid aggregate. B, Diffuse lymphocytic infiltrate. C, Hyperplasia of the lining layer. D, Fibrin cap replacing a denuded lining layer.
is often covered with fibrin deposits generated from activation of the fibrinolytic system in inflammatory synovial fluid. Occasionally the synovial lining layer is completely denuded and is replaced by a dense fibrin cap. In highly inflamed tissues, fibrin deposits extend deeply into the sublining stroma, which may be edematous owing to the marked increase in vascular permeability. The earliest synovial changes in RA appear to feature microvascular abnormalities, and mononuclear cell infiltrates have been detected in asymptomatic joints of RA patients. These features are nonspecific and are seen in the synovium of acutely inflamed joints from a spectrum of disorders, including reactive arthritis and psoriatic arthritis.

In RA, the mononuclear cell infiltrate in the sublining stroma can be diffuse but more commonly is arranged in perivascular aggregates resembling lymphoid follicles (see Figure 53-8). Although the presence of lymphoid aggregates in the synovial membrane is typical of RA, this histopathologic feature is by no means unique to RA synovitis. Lymphoid follicles are typically located near vessels with tall endothelium, which are termed high endothelial venules; these vessels specialize in the recruitment of lymphocytes (Figure 53-9). Multinucleated giant cells are occasionally seen in RA synovium (Figure 53-10), and some tissues demonstrate granuloma formation. Finally, it should be noted that synovial tissue obtained at the time of joint arthroplasty often exhibits extensive fibrosis and may be indistinguishable from arthroplasty samples obtained from patients with osteoarthritis.

The synovial histopathology of psoriatic arthritis, ankylosing spondylitis, and reactive arthritis has been compared with that of RA. In all cases, a similar spectrum of inflammatory cell populations has been identified, but several subtle and potentially important differences have been observed. Overall, synovial histologic and immunohistologic features of psoriatic arthritis, both oligo- and polyarticular, resemble those of other spondyloarthropathies to a greater extent than RA (see later under “Synovial Immunohistology”). Comparative studies have suggested that synovial lesions in psoriatic arthritis are more vascular than those of RA, with more tortuosity of the synovial microvasculature. This is evident both macroscopically and microscopically. Moreover, lymphoid aggregates of various sizes were identified in 25 of 27 synovial tissue samples from patients with psoriatic arthritis, and 13 of 27 had large organized aggregates with all of the features of ectopic lymphoid neogenesis that have been associated with RA synovitis. Studies of synovium from the peripheral joints of ankylosing spondylitis patients have revealed intense infiltrates of lymphocytes, plasma cells, and lymphocytic aggregates. It should also be noted that synovium from patients with osteoarthritis often features the presence of lymphocyte aggregates, although these tend to be small and less well developed than those seen in RA.

The synovium of lupus patients showed synovial hyperplasia, inflammatory infiltrates, vascular proliferation, edema and congestion, fibrinoid necrosis and intimal fibrous hyperplasia of blood vessels, and superficial fibrin deposits, although these changes were quantitatively modest compared with those of RA. In early scleroderma, the lining layer was seen to be thin with deposits of fibrin and stromal lymphocytes and plasma cells, and similar changes were seen in patients with dermatomyositis and polymyositis (see Figure 53-7). A recent study comparing the immunopathologic features of early untreated Behçet’s disease versus those of psoriatic arthritis (PsA) noted that although a similar degree of inflammation was seen in the two disorders, Behçet’s synovitis demonstrated higher numbers of neutrophils and T cells than were seen in psoriatic synovitis.

In patients with chronic crystal arthropathies, large deposits of birefringent material can be detected in the synovium. Amyloid arthropathy can be diagnosed by demonstrating amyloid deposits in the synovium using Congo red staining (see Figure 53-7). The synovium in ochronosis contains brownish shards of cartilage. Multicentric reticulohistiocytosis can be diagnosed pathologically by the presence of large foamy cells and multinucleated cells in the synovium. In arthritis of hemochromatosis, the synovium
exhibits brown hemosiderin deposits in the lining cells, and CPPD crystals can also be observed.\(^7^0\)

**SYNOVIAL IMMUNOHISTOLOGY**

**Considerations Regarding Sampling and Quantitative Analysis**

Immunohistology utilizes specific monoclonal or polyclonal antibodies with well-defined molecular targets and is an effective tool for analyzing the cellular and molecular features of the synovium. As the field has progressed over the past two decades, it has become clear that algorithms for generating reproducible quantitative data from immunohistologically stained sections are required. Moreover, approaches are needed for minimizing the sampling bias that is inherent in biopsy-based studies.\(^7^1\) Studies have suggested that if six or more individual specimens from different parts of the joint are examined, variance is reduced to less than 10% for T cell and activation markers.\(^7^2\) Furthermore, it has been shown that synovial inflammatory features are similar in areas adjacent to and distant from the pannus cartilage junction, with the possible exception of macrophage numbers, which tend to be higher in adjacent areas.\(^7^3,7^4\)

Various methods have been proposed by which quantitative data for immunohistologically stained synovial tissue sections can be generated.\(^7^5,7^6\) The easiest and least costly method is to generate semiquantitative scores of staining intensity (e.g., on a 0 to 3 scale) from multiple areas of the tissue, and on the basis of these to obtain an average score for the entire tissue. The reliability and reproducibility of this method are increased if two observers score the tissue sections independently and a final average of the scores is generated. Computer-assisted image analysis involves capturing images from multiple areas of the tissue samples to which color-specific quantitative software algorithms are then applied. This method generates the greatest quantity of reproducible data but requires expensive equipment and certain levels of operator skill. Furthermore, differences in the background staining intensity of individual sections can make this type of analysis technically difficult.

**Synovial Lining Cell Layer**

Compared with normal synovium, the lining layer in RA is often hyperplastic, resulting from an increase in both type A and type B cells, as indicated by an increase in CD68 and CD55 staining, respectively (Figure 53-11). It is assumed that macrophage-like synoviocytes are recruited from the blood and then migrate through the synovial stroma and ultimately are retained in the lining layer in close association with fibroblast-like synoviocytes. It is currently proposed that the increase in fibroblast-like synoviocytes may be related more to defects in apoptosis than to recruitment or local proliferation. Expression of several families of adhesion molecules by both types of lining cells results in their close association and modulates their activation status. These include \(\beta_1\) and \(\beta_2\) integrins and their respective immunoglobulin supergene family ligands, particularly intercellular adhesion molecule (ICAM)-1 and VCAM-1.\(^7^7,7^8\) Cadherin-11 expressed by fibroblast-like cells likely plays a key role in the adhesive interactions that sustain the lining layer hyperplasia.\(^7^9\) This adhesion molecule is widely expressed in the lining layer of normal cells, as is shown in Figure 53-12. The relationship between fibroblast-like synoviocytes in the lining layer and other populations of

---

**Figure 53-11** A-D, Immunoperoxidase staining of normal synovium and rheumatoid arthritis (RA) synovium for CD55 (fibroblast-like synoviocytes) and CD68 (macrophage-like synoviocytes). Both subsets of lining cells are increased in the hyperplastic lining layer of RA synovium.
mesenchymal cells in the sublining stroma remains uncertain. Immunohistology indicates that expression of CD55, VCAM-1, and cadherin-11 is primarily seen in the lining cell layer with minimal evidence of expression in sublining fibroblast populations. Similarly, our understanding of the relationship between lining layer macrophage-like cells and sublining macrophages is incomplete, and both express widely used macrophage markers such as CD68 and CD14. Seminal work from Edwards and associates has suggested that macrophage-like lining cells preferentially express FcRγIIIa receptors, which may serve to localize immune complexes to the synovium.

Functionally, the lining cell layer in RA is highly activated. Human leukocyte antigen (HLA)-DR is highly expressed, particularly by macrophage-like cells, which may suggest a role for these cells in antigen presentation. Several studies have indicated that cells in the RA lining layer are the principal source of cartilage-degrading proteases, particularly matrix metalloproteinase (MMP)-1 and MMP-3, (Figure 53-13). The lining layer is generally less hyperplastic in spondyloarthropathies such as PsA and reactive arthritis compared with RA. Less is known about the functional state of the lining cells in these disorders, although it is likely that differences compared with RA are quantitative rather than qualitative.

**Synovial Lymphocytes and Plasma Cells**

In the synovial tissues of RA and spondyloarthropathies patients is a predominance of CD3+ T cells, and the CD4/CD8 ratio is 4:1 or greater in the lymphocytic aggregates but is lower in more diffuse infiltrates. Moreover, the CD4
cells in the aggregates also express CD27, which facilitates B cell help. Considerable attention has been paid to whether the infiltrating T cells in RA and other arthropathies are primarily Th1 (interferon [IFN]-γ producing) or Th2 (IL-4 producing) biased, but the data in this area have been inconsistent. Until recently, it was suggested that T cells in RA synovium are more Th1 biased compared with those in synovium from spondyloarthropathies, with a higher Th1/Th2 cytokine ratio. Identification of a third subset of T helper cells expressing IL-17 and playing a central role in chronic inflammatory disorders has necessitated a revision in the role that T cells play in synovitis. The presence of IL-17, IL-1β, and tumor necrosis factor (TNF) in RA synovium was found to be predictive of progressive damage. Furthermore, a subset of CD4 T cells expressing CD25 and the gene FoxP3, so-called regulatory T cells (Tregs), are now known to play a regulatory role in antigen-specific T cell expansion. Although Tregs are readily detected in the joints of patients with RA and other inflammatory arthropathies, their suppressor function appears to be defective in this microenvironment. It has been suggested that CD8 T cells are needed to maintain the structure of ectopic lymphoid-like structures in RA synovium, even though the numbers of these T cells typically are substantially lower than the numbers of CD4+ cells.

B cells are identified by expression of CD19 and CD20 and are particularly abundant in tissues exhibiting large lymphoid aggregates with germinal centers. B cells typically are found in close association with CD4-positive T cells in these aggregates (see Figure 53-9). In severe combined immunodeficiency (SCID) mice suggest that B cells may be critical for maintaining the microarchitecture of synovial lymphoid follicles and for T cell activation. Memory B cells are efficient antigen-presenting cells, and rheumatoid factor–producing B cells are well suited for capturing a wide spectrum of antigens in immune complexes. In RA synovium in particular, areas surrounding the lymphoid aggregates are often densely infiltrated with sheets of CD38+ plasma cells. Analysis of V gene variants and rearrangements in B cells and plasma cells in RA and reactive arthritis synovium indicates that plasma cells from a particular aggregate are clonally related, suggesting that their terminal differentiation occurred in the synovial microenvironment. Synovial plasma cells actively synthesize immunoglobulin, some of which has been shown to result in the production of autoantibodies such as anti-citrulline antibodies, which recognize local citrullinylated proteins. As has been stated, plasma cell infiltrates are also seen in PsA, ankylosing spondylitis, and reactive arthritis synovium, although a systematic analysis of synovial samples from patients with early arthritis has suggested that their presence is most suggestive of RA. It was found in one study that intracellular citrullinated proteins were detected in RA but not in spondylarthropathy synovium. In contrast, another study found that the presence of citrullinated proteins was not specific for RA synovitis. The areas immediately adjacent to the dense lymphoid aggregates, which comprise primarily CD4+ T cells and B cells, have been called transitional zones (see Figure 53-9). These areas feature a lower CD4/CD8 ratio and appear to be particularly active immunologically.

Transitional areas are rich in macrophages and interdigitating dendritic cells, both of which are highly efficient antigen-presenting cells. Lymphoblasts, in particular CD8+ T cells, are seen to be present in close proximity to antigen-presenting cells.

Natural killer cells can be identified by cell surface markers, expression of granzymes, and functional assays. Several studies have suggested an expansion of subsets of natural killer cells in RA synovial tissue and synovial fluids. Mast cells are abundant in RA synovium and co-localize with inflammatory mediators and proteases in the synovial microenvironment. Synovial Sublining Macrophages and Dendritic Cells

Macrophages are present in the sublining areas of healthy and diseased synovium but are particularly abundant in the sublining stroma of RA synovium. Indeed, when markers such as CD68 and CD14 are used to study highly inflamed tissues, no clear distinction can be made between the sublining macrophage population and the macrophage-like synoviocytes present in the hyperplastic lining layer, although expression of complement receptor for C3b and iC3b was shown to be unique for lining macrophages in RA, osteoarthritis, and normal synovium. Studies using various macrophage markers suggest that recently migrated macrophages in perivascular areas express CD163 brightly, in addition to expressing CD68 and CD14, whereas macrophages in large lymphocytic aggregates and in the lining layer are less likely to express CD163. CD163+ macrophages, which have recently been called M2 macrophages, were found to be more abundant in spondylarthropathy than in RA synovium. The functional correlates of these phenotypic differences remain unclear.

M1 macrophages, which produce TNF and IL-1β, are more abundant in RA and are under-represented in PsA and other spondyloarthropathies in which M2 macrophages are more abundant. Furthermore, it has been shown that the number of macrophages in RA synovium, primarily of the M1 subset, correlates well with the destructive potential of the synovitis, as evidenced by erosive radiographic damage. This may reflect the highly activated status of these cells, which serve as the principal source of synovial TNF and IL-1β. A body of evidence has suggested that populations of synovial macrophages serve as osteoclast precursors that mature in the synovial microenvironment and then directly mediate erosive damage to adjacent bone.

Mature dendritic cells are the most efficient and potent of the antigen-presenting cells, and are found abundantly in RA synovium in close contact with T lymphocytes. Two major subsets of dendritic cells have been described: myeloid dendritic cells (mDCs) and plasmacytoid dendritic cells (pDCs). They can be identified by immunohistology as stellate cells with dendrites expressing high levels of HLA-DR and co-stimulatory molecules such as CD80, CD83, and CD86. mDCs express CD11c and CD1c, and pDCs express CD304. One study suggested that compared with psoriatic synovitis, the RA synovium is particularly enriched in pDCs. Detailed studies that have examined the expression of chemokines involved in dendritic cell
migration and recruitment suggest that a substantial proportion of dendritic cells in the synovium arrive in an immature state and subsequently undergo maturation within the synovial microenvironment in a T cell–rich area.\textsuperscript{117,118} Follicular dendritic cells in the germinal centers of large lymphocytic aggregates express the markers CD16, FDC, and VCAM-1.

**Synovial Microvasculature, Endothelium, and Stromal Mesenchymal Cells**

The stromal elements in RA are often expanded in parallel with the inflammatory cell infiltration. The microvasculature appears to be markedly increased, particularly in the deep sublining areas, and this expansion is presumed to relate to local stimulation of angiogenesis (see Figure 53-13). Morphometric studies have suggested that the number of vessels immediately adjacent to the lining layer is actually reduced compared with normal.\textsuperscript{120} This situation, combined with the metabolic demands of this tissue, may actually produce a relatively ischemic and hypoxic environment, which is reflected in the biochemical properties of RA synovial fluid.\textsuperscript{121} Immunohistologic studies have indicated that the molecular consequences of hypoxia, particularly expression of hypoxia-inducible factor-1α (HIF-1α), a key regulator of the cellular hypoxic response, are increased in RA synovitis.\textsuperscript{122,123} Studies that have directly measured synovial tissue pO\textsubscript{2} using arthroscopic probes have confirmed the hypoxic nature of RA synovitis.\textsuperscript{124-126} The synovial endothelium in RA and other inflammatory arthropathies is activated by proinflammatory mediators in the microenvironment to express adhesion molecules such as E-selectin, ICAM-1, and VCAM-1, which are involved in the recruitment of inflammatory cells.\textsuperscript{127}

**Synovial-Cartilage-Bone Interface**

The interface between inflamed synovium and adjacent cartilage and bone in RA and other chronic arthropathies is a site of particular interest because much of the articular damage occurs in these areas. In RA, this destructive synovial tissue is called pannus, which may spread to cover most of the surface of the cartilage and invade the bone in bare areas at the joint margin (Figure 53-14). Pannus has been pathologically characterized primarily from samples obtained at the time of joint arthroplasty, although arthroscopic studies at earlier stages of disease have attempted to characterize synovial samples adjacent to this area. Immunohistology suggests that synovial macrophages and fibroblasts are abundant at the pannus-cartilage interface, and that high levels of proteases are expressed by these cells. At the interface between pannus and bone, substantial numbers of multinucleated osteoclasts can be identified morphologically and by specific markers such as calcitonin receptors, cathepsin K, and staining for tartrate-resistant acid phosphatase\textsuperscript{128} (see Figure 53-14). Moreover, expression of receptor activator of nuclear factor κB (NFκB) ligand (RANKL), a key cytokine in osteoclastogenesis, was prominent in these areas.\textsuperscript{129}

**Synovial Biopsy and Pathology as Tools for Predicting and Assessing Response to Therapy in Inflammatory Arthritis**

Numerous academic rheumatology centers have focused on using serial arthroscopic biopsy and quantitative immunohistology as tools to assess the impact of therapeutic interventions on synovial lesions in RA. These studies have been particularly valuable in evaluating the effects of targeted biologic therapies, where the molecular target and the biologic basis of the mechanism of action are well defined.\textsuperscript{130-141} It has been proposed that synovial biopsy–based studies, which are relatively small and inexpensive to undertake, offer a unique opportunity to assess the impact of novel therapeutic agents on the target tissue at an early stage of pharmaceutical drug development. On the basis of these studies, it may be possible to make important decisions regarding the future development of a particular agent.

This appealing proposition is currently hindered by several important considerations. First, the arthroscopic equipment, expertise, and infrastructure needed to undertake these studies remain limited to a small number of centers. Second, considerable concern has arisen regarding the issue of sampling bias in these studies, particularly because serial biopsies are compared in the same individual. As has been discussed, various approaches are used to minimize this bias, including systematic sampling of the same areas of the joint, computerized image analysis of multiple representative tissue samples, quantification of an adequate number of microscopic fields, and utilization of quantitative

**Figure 53-14** Interface between pannus tissue and bone in a patient with rheumatoid arthritis. **A**, The synovial lesion is invading adjacent bone. **B**, Staining for tartrate-resistant acid phosphatase in the circled area demonstrates the presence of osteoclasts.
PCR and proteomic techniques to assess overall levels of specific molecules. Finally, and most important, has been the lack of a synovial biomarker(s) with which to reproducibly evaluate outcomes across various studies. The number of macrophages in the tissue has been proposed as a good candidate biomarker, although this remains to be systematically tested.132,142

**SUMMARY**

Analysis of synovial fluid and tissue samples provides valuable diagnostic information in specific clinical settings. In cases where septic or crystal-induced arthritis is suspected, as in acute monoarthritis, synovial fluid analysis is critical for making the diagnosis. In cases of undiagnosed chronic monoarthritis, synovial biopsy may provide definitive evidence of conditions such as tuberculosis, sarcoidosis, and pigmented villonodular synovitis.

Systematic analysis of synovial tissue in RA and other forms of inflammatory arthritis, particularly with the use of immunohistology, has provided a wealth of information concerning the cellular and molecular mechanisms that sustain synovial lesions. Research protocols are currently exploring the utility of synovial biopsy in predicting response to antirheumatic therapies.

**Selected References**


samples selected at arthroscopy compared with samples obtained blindly by needle biopsy, Arthritis Rheum 41:663–669, 1998.


Full references for this chapter can be found on www.expertconsult.com.


### KEY POINTS

<table>
<thead>
<tr>
<th>Point</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthrocentesis with joint injection</td>
<td>A simple, low-risk, office-based procedure that can be extremely useful diagnostically and therapeutically.</td>
</tr>
<tr>
<td>Diagnostic arthrocentesis</td>
<td>Indicated in patients with effusion without a known diagnosis or if a new diagnosis is suspected, and it can be definitive for infection- or crystal-induced disease.</td>
</tr>
<tr>
<td>Therapeutic corticosteroid injection</td>
<td>Potentially provide clinical benefit for many painful joints or periarticular structures.</td>
</tr>
<tr>
<td>Most forms of noninfectious inflammatory arthritis respond to local injection.</td>
<td>For example, bursitis of the shoulder.</td>
</tr>
<tr>
<td>Clinical trials of injection in osteoarthritis of the knee show benefit compared with placebo, but duration varies.</td>
<td>Evidence for efficacy of corticosteroid injection for nonarticular conditions is best for painful shoulders, but many other conditions have been shown to respond in small trials or anecdotal reports.</td>
</tr>
</tbody>
</table>

**Arthrocentesis and injection of joints**

Arthrocentesis and injection of joints are safe and simple procedures that can be performed routinely at an outpatient visit. With analysis of synovial fluid, few procedures in medical practice have the potential to be as diagnostically definitive as arthrocentesis and few modalities can be as effective in achieving symptomatic relief of painful or swollen articular structures as the injection of corticosteroids. For these reasons, one out of five visits to a rheumatology practice includes aspiration or injection of a joint or periarticular structure. However, a majority of internists finishing their residency training believe they need more training in these important, safe, and effective procedures, and most injections performed in primary care settings are done by a small percentage of practitioners with experience and comfort with the procedure. Recent efforts to formalize educational processes in arthrocentesis and joint injection have the potential to increase the number of primary care physicians who use these procedures more regularly.

Paracelsus is credited with the first descriptions, in the early sixteenth century, of the viscous fluid present within synovial cavities, but aspiration of synovial fluid for analysis and aid in diagnosis did not become a topic of increasing interest until the first half of the twentieth century. Numerous studies of synovial fluid components and techniques used to obtain this fluid appear in a 1935 textbook by Pemberton. An early description of arthrocentesis technique can be found in the classic book by Ropes and Bauer on synovial fluid analysis, published in 1953, which used data collected over a 20-year period. During this era, swollen, distended joints were often aspirated for relief of discomfort. In most instances, the prompt reaccumulation of fluid and concerns about infection from repeated aspirations limited the usefulness of aspiration for the relief of arthritis symptoms. A wide variety of substances were injected into joints throughout the early twentieth century including formalin and glycerin, ethibonded oil (Lipiodol), lactic acid, petroleum jelly, and liquefied oil prepared from the patient’s own subcutaneous fat. Most of these therapies were apparently abandoned, and the most therapeutic injections discussed in arthritis textbooks from the 1940s related to temporary relief via injections of procaine for osteoarthritic knees and bursitis of the shoulder.

In 1951, after observations on the efficacy of topical cortisone for ocular inflammation, Hollander first reported a minimal, transient improvement in 25 knees of patients with rheumatoid arthritis when injected with cortisone. In subsequent years, Hollander injected hydrocortisone acetate with a much better response, providing further evidence that this was the active anti-inflammatory metabolite of cortisone. Further reports of benefit from injectable corticosteroids appeared in the 1950s. During this period, studies showed that more stable and less soluble compounds in the form of esterified crystaline hydroxyocortisone and its analogues were even more effective and had longer duration of anti-inflammatory effects. By the early 1960s, Hollander had reported a series of more than 100,000 injections of joints, bursae, and tendon sheaths in 4000 patients with a variety of conditions. In rheumatology practice since then, aspiration and therapeutic injection of joints and periarticular tissues have become common and essential procedures.

### INDICATIONS AND CLINICAL EVIDENCE

#### Arthrocentesis

Aspiration of synovial fluid may be indicated in any joint with detectable effusion or may be attempted in joints without detectable effusions when diagnosis is in doubt (Table 54-1). In patients in whom a diagnosis is uncertain, synovial fluid analysis usually provides important information regarding the inflammatory or noninflammatory nature of the process within the affected joint and may be definitive in patients with crystal-induced or infectious arthritis. In patients with recently diagnosed bacterial arthritis, repeated aspiration of accumulated fluid is often an important adjunct to antibiotic therapy. Joints with detectable effusions may be aspirated for relief of discomfort, with or without a subsequent injection of corticosteroid. In some
patients, aspiration alone, without steroid injection, may be particularly effective for noninflammatory effusions or self-limited conditions.

**THERAPEUTIC INJECTION**

**Inflammatory Arthritis**

Therapeutic injection of corticosteroids is generally believed to be most effective in joints affected by inflammatory arthritis (Table 54-2). Most of the experience in this area has been with common conditions such as rheumatoid arthritis, juvenile rheumatoid arthritis, crystal-induced arthritis, psoriatic arthritis, and reactive arthritis. Anecdotal experience has been reported in less common conditions such as systemic lupus erythematosus and sarcoidosis. In self-limited conditions such as gout, injections generally lead to more prompt resolution of exacerbations.

In rheumatoid arthritis, injections are used frequently to suppress inflammation in individual joints. Such injections are generally considered to be adjunctive to disease-modifying drug therapy and are not believed to affect overall outcomes. The efficacy of individual joint injections in rheumatoid arthritis is supported by many large, uncontrolled case series dating back to the 1950s. Most of Hollander’s early reports suggested long-lasting relief in most joints injected, with improvement lasting several months in most patients. In this series, patients received about two injections during the first year of treatment and averaged 0.6 injection per patient-year for the next 15 years.

More recently, multiple intra-articular injections in inflamed joints have been shown to be a useful part of an overall regimen of disease-modifying therapy in rheumatoid arthritis, resulting in improvement superior to similar doses of systemic steroids, and helpful in obtaining clinical remission and reducing radiographic progression of disease. In particular, intra-articular injections of steroids added to systemic methotrexate therapy have been shown to improve periarthritis osteopenia in swollen hand joints over a 3-month period in patients with rheumatoid arthritis. Removal of fluid before aspiration increases the efficacy of steroid injections in most patients with inflammatory arthritis. One study of 191 knee injections in patients with rheumatoid arthritis showed that aspiration of fluid reduced the rate of relapse from 47% to 23% within a 6-month period after injection compared with joints not aspirated.

Corticosteroid injection is considered to be a safe and effective option for prompt relief of acute crystal-induced

---

**Table 54-1**  **Indications for Arthrocentesis**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Indication for Arthrocentesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undiagnosed Arthritis with Effusion</td>
<td>Characterize type of arthritis</td>
</tr>
<tr>
<td></td>
<td>Noninflammatory (WBC &lt; 2000/mm³)</td>
</tr>
<tr>
<td></td>
<td>Inflammatory (WBC &gt; 2000/mm³)</td>
</tr>
<tr>
<td></td>
<td>Septic (WBC &gt; 50,000/mm³)</td>
</tr>
<tr>
<td>Definitive diagnosis</td>
<td>Gout (urate crystals)</td>
</tr>
<tr>
<td></td>
<td>Pseudogout (calcium pyrophosphate dihydrate crystals)</td>
</tr>
<tr>
<td></td>
<td>Septic arthritis (Gram stain [rare] or culture)</td>
</tr>
<tr>
<td>Undiagnosed Arthritis without Effusion</td>
<td>May be definitive in gout (knee, first metatarsophalangeal joint)</td>
</tr>
<tr>
<td>Patient with Known Diagnosis</td>
<td>Septic arthritis (repeated taps for adequate drainage)</td>
</tr>
<tr>
<td></td>
<td>Other types of arthritis for symptomatic relief (with or without injection)*</td>
</tr>
</tbody>
</table>

*Most studies show improved effect if fluid aspirated before injection.

---

**Table 54-2**  **Indications for Therapeutic Injection**

<table>
<thead>
<tr>
<th>Arthritis</th>
<th>Indication for Therapeutic Injection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflammatory Arthritis</td>
<td>Rheumatoid arthritis: almost always effective; duration varies; should be used as an adjunct to an overall regimen of disease-modifying therapy, efficacy equivalent to or superior to systemic steroids</td>
</tr>
<tr>
<td></td>
<td>Crystal-induced arthritis: few published studies; effective in 24-48 hr</td>
</tr>
<tr>
<td></td>
<td>Septic arthritis: complete response in 2 wk in 57%; predictor of good outcome</td>
</tr>
<tr>
<td></td>
<td>Spondyloarthropathies: peripheral joints respond as in rheumatoid arthritis</td>
</tr>
<tr>
<td></td>
<td>Sacroiliac joint injections under imaging guidance</td>
</tr>
<tr>
<td></td>
<td>Juvenile rheumatoid arthritis: particularly useful in oligoarticular form; may be “disease modifying” and safer than nonsteroidal anti-inflammatory drugs</td>
</tr>
<tr>
<td></td>
<td>Miscellaneous diseases (anecdotal): Sarcoidosis</td>
</tr>
<tr>
<td></td>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>Noninflammatory Arthritis</td>
<td>Osteoarthritis: Knees: 60%-80% response in 1-6 wk versus placebo; no difference at 12 wk, global improvement over 2 years with every 3-month strategy but no lasting effect</td>
</tr>
<tr>
<td></td>
<td>Hips: anecdotal reports; require fluoroscopy; usually avoided</td>
</tr>
<tr>
<td></td>
<td>Hyaluronic acid derivatives (Hyalgan, Synvisc) weekly for 31–35 wk: moderately better than placebo</td>
</tr>
<tr>
<td></td>
<td>Hemophilic arthropathy: reported, but rarely used</td>
</tr>
<tr>
<td>Nonarticular Conditions (Tendinitis, Bursitis, Myofascial Pain)</td>
<td>Trigger point injections: done frequently; not supported by studies</td>
</tr>
<tr>
<td></td>
<td>Painful shoulder (rotator cuff tendinitis, frozen shoulder): efficacy compared with placebo; lasting 4-6 mo</td>
</tr>
<tr>
<td></td>
<td>Lateral epicondylitis (tennis elbow): efficacy for 1-2 mo versus placebo; less in some studies, possible increased recurrences later</td>
</tr>
<tr>
<td></td>
<td>Carpal tunnel syndrome: 90% short-term response; variable at 6-12 mo; good response to injection may be predictor of surgical response</td>
</tr>
<tr>
<td></td>
<td>de Quervain’s tenosynovitis: 70%-90% improved with 1-2 injections, relapse in 30% at 1 yr</td>
</tr>
<tr>
<td></td>
<td>Trochanteric pain of hip (bursitis, tendinitis): 60%-70% at 6 mo (uncontrolled)</td>
</tr>
<tr>
<td></td>
<td>Knee pain syndromes: Anserine bursitis</td>
</tr>
<tr>
<td></td>
<td>Patellofemoral pain syndromes</td>
</tr>
<tr>
<td></td>
<td>Synovial plica</td>
</tr>
<tr>
<td></td>
<td>Popliteal cyst (usually treated with intra-articular knee injection)</td>
</tr>
<tr>
<td></td>
<td>Planter fasciitis: variable; probably better for 1-3 mo</td>
</tr>
<tr>
<td></td>
<td>Morton’s neuroma: response often prolonged (no controls)</td>
</tr>
<tr>
<td></td>
<td>Tarsal tunnel syndrome: rarely reported; usually only temporary</td>
</tr>
<tr>
<td></td>
<td>Achilles tendinitis, bursitis: usually avoided</td>
</tr>
<tr>
<td></td>
<td>Cervical girdle, lumbar areas, posterior hip: uncertain what structure injected (without fluoroscopic facet block); efficacy not proved</td>
</tr>
</tbody>
</table>
arthritides in gout and pseudogout. Steroid injections are so widely accepted as an effective treatment that few reports have attempted to address the degree or duration of efficacy for injections in these conditions. Many of the patients described in early reports of steroid injections were being treated for acute crystal-induced arthritis, with prompt relief being almost uniform. In a more recent report, small doses of intra-articular steroids were successful in relieving pain and swelling completely in all patients within 48 hours, and no relapses were noted in 20 patients over a 3-month period.14

In patients with recent onset of inflammatory oligoarthritis, without definitive diagnosis, corticosteroids can be used to relieve symptoms of swelling in individual joints, and the response to these injections can be used as a prognostic marker. In a series of 51 patients with recent-onset inflammatory arthritis involving five or fewer joints, a strategy of injecting all joints with clinical synovitis resulted in improvement in all patients and a complete resolution of synovitis after 2 weeks in 57% of patients. The response at 2 weeks was the best predictor of continued improvement persisting for 26 weeks and 52 weeks.15 A trial of steroid injection may be used for symptomatic relief in such patients, and a complete response at 2 weeks can be used as a prognostic indicator for a better outcome. Local injections are effective in 41% and 23% of joints at 3 and 12 months, respectively, in patients with localized flares of psoriatic arthritis, with the response more prolonged in patients taking effective long-acting systemic therapy.16

Patients with refractory sacroiliac joint pain related to ankylosing spondylitis or other spondyloarthropathies may benefit from injection of the sacroiliac joints.17 Because of the anatomy of this joint, such injections often require radiographic confirmation of needle placement in the joint space, and use of fluoroscopy-guided, computed tomography (CT)–guided, and magnetic resonance imaging (MRI)–guided injections has been reported. In uncontrolled studies, a good response has been reported in about 80% of injections, with an average time of improvement of 6 to 9 months. At least one controlled study in a small group of patients showed slight benefit of steroid compared with placebo injection. The degree of improvement seen after sacroiliac injections has not been consistent among studies to date, however, probably because of a lack of uniformity of patient selection and outcomes assessed.

Joint injection has been used with increasing frequency in recent years in patients with juvenile rheumatoid arthritis,18 particularly in the pauciarticular variant of the disease, in which only a few joints are involved, and potentially toxic systemic therapy can be avoided. Complete remission lasting more than 6 months has been reported in approximately 65% to 80% of joints injected in this condition, most commonly in the knees. Benefit has also been shown in smaller numbers of ankles, wrists, shoulders, elbows, and temporomandibular joints, with most children being able to stop oral medications, and correction of joint contracture being noted in most as well.19 A median duration of improvement of approximately 74 weeks has been documented in another large study. Joint lavage before injection may be useful in prolonging response in patients with a poor response to previous injection.20 One study showed a significant decrease in leg-length discrepancy in children treated with repeated injections (average of 3.25 injections per child over 42 months) compared with children in another center who were not injected.21 A recent decision analysis model suggests that intra-articular injection is superior to a strategy of initial nonsteroidal anti-inflammatory (NSAID) use in patients with monoarthritis of the knee.22

A more recent study specifically addressed the efficacy and safety of steroid injections in the hip in juvenile rheumatoid arthritis.23 In this prospective study of 67 hip injections, 58% of hips remained in remission for 2 years after a single injection; another 18% required a second injection to maintain remission. Only two cases of avascular necrosis were seen in this group, and both of these were in patients receiving systemic steroids, suggesting no role for local steroid injections in the development of avascular necrosis in this population.

Finally, local steroid injection may be a useful adjunct in managing patients with hemophilic arthropathy.24 In an open trial of 19 injections, 79% of joints improved within 24 hours; this improvement persisted for 8 weeks in 58% of joints. A decrease in need for clotting factor was shown in this small group.

Noninflammatory Arthritis

Corticosteroid injection is used frequently in common non-inflammatory articular conditions such as osteoarthritis, internal derangements, and post-traumatic arthritis. Clinical studies that support efficacy are less convincing and suggest a less predictable and smaller degree of response in these conditions than is seen in inflammatory arthritis. Most studies of steroid injections in osteoarthritis have studied patients undergoing knee injections.25 Early uncontrolled studies suggested improvement in approximately 60% to 80% of patients.4 In controlled studies, most benefit, compared with placebo, seems to last 1 to 6 weeks, with return to the same pain levels seen in placebo groups by around 12 weeks after injection.26,29

Factors associated with a better response to steroid injections have included less severe radiographic changes, the presence of effusion at the time of injection, and successful aspiration of fluid at the time of injection. The theoretic concern about the potential for negative effects of injected steroids on cartilage (discussed in the following section) is often cited as a reason to limit injections in osteoarthritis and other forms of noninflammatory arthritis. A more recent trial in 68 patients comparing corticosteroid injections every 3 months with saline injections showed no worsening of radiographic changes, however, after 2 years of repeated steroid injection, along with significant improvement in symptoms during the period of study.30

Injections in patients with osteoarthritis of the hip are less likely to be helpful and are more technically difficult. Some relief can be obtained, however, in patients with less severe disease or in a rare patient with more severe involvement. A prospective, open study of intra-articular steroid in 45 patients with hip arthritis, 27 of whom had osteoarthritis, found a significant reduction in pain at 2 weeks and 12 weeks, although the effect was lost by 26 weeks.31 In a report of 510 patients treated with a single injection done under fluoroscopic guidance, pain relief that persisted.
8 weeks was seen in 90% of patients with mild disease, 58% of patients with moderate disease, and 9% of patients with severe hip osteoarthritis; improved range of motion was shown in most of the patients responding. Controlled trials have shown that steroid injection provides modest benefit compared with placebo for 2 to 12 weeks, and the benefit was no longer apparent at 3 months.33 Steroid injection followed by non-weight bearing was not helpful in reducing the need for hip replacement in a retrospective study of patients with rapidly progressive osteoarthritis of the hip.35 In patients with osteoarthritis of the thumb, local steroid injection may be helpful for 1 year in a few patients (=20%), but most patients are improved for 1 to 3 months at the most.36

Injectable hyaluronic acid derivatives have been studied extensively and are frequently used for injection into osteoarthritic knees and occasionally in other joints. A series of one to five weekly injections has been shown to provide more pain relief than placebo in most studies. The duration and rate of improvement in these studies have varied, however, and the optimal role for hyaluronic acid injections in the management of arthritis has yet to be determined (see later discussion and Chapter 100).35,38

Nonarticular Conditions

Patients with various forms of tendinitis, bursitis, myofascial pain, and nerve entrapment syndromes are frequently treated with local injections of corticosteroids.39 In many of these conditions, uncontrolled clinical experience suggests a high response rate, and in many others, controlled trials show variable levels of benefit, often depending on whether short-term or long-term outcomes are considered. The injection of trigger points for pain relief has been used by many practitioners over the past several decades, but few controlled studies to support efficacy have been published.

Steroid injections are frequently used in the management of rotator cuff tendinitis, frozen shoulder, and other causes of shoulder pain. Most controlled studies have shown significant short-term improvement from steroid injection compared with placebo injection, usually lasting 4 to 6 months.40 In most such trials, short-term treatment success of 75% to 80% is usually reported in steroid-treated groups compared with 40% to 50% in placebo groups. Most, but not all, controlled studies have shown that steroid injections are superior to physical therapy without injection, and that combining steroid injection with physical therapy has an additive benefit.41-43 In a subset of patients with painful shoulder related to calcific tendinitis, local injection of ethylenediamine tetra-acetic acid (EDTA) may result in pain relief and radiographic resolution of calcification.44 Local injections may also be useful for shoulder pain related to the acromioclavicular joint.

Lateral epicondylitis is also commonly treated by local injections.40 Pain may worsen for 1 or 2 days after injection but usually improves after 4 to 5 days.41 Longer controlled studies typically document improvement of 90% compared with 50% in placebo treatment in the first 1 to 2 months after injection, but outcomes at 6 to 12 months are usually not affected and one study has shown that recurrences of pain after 6 weeks were more common in patients who had steroid injections.42-44 Local injection of botulinum toxin has been shown to be beneficial compared with placebo for epicondylitis in small trials.45 In femoral trochanteric pain syndromes (i.e., bursitis), the response rate to locally injected steroids has been reported to be 60% to 100%, but no placebo-controlled studies have been done. A prospective study reported significant improvement after a single injection in 77% of patients at 1 week; this number decreased to 69% at 6 weeks and 61% at 26 weeks.46 Patients receiving larger amounts of locally injected steroid (24 mg of betamethasone) were more likely to have sustained improvement. Around the knee, anecdotal and retrospective studies have shown that most patients with anserine bursitis respond to local steroid injection.

Local steroid injection may be a useful nonsurgical therapy for carpal tunnel syndrome. Most studies report 90% short-term relief of symptoms from a single injection; longer-term relief is 20% to 90%, and surgery is eventually required in about half of patients treated with injection. Controlled trials comparing local steroid injection with surgical decompression have shown variable results, suggesting that injection may provide better relief within the first 3 to 6 months, but more patients benefit from surgery when followed for 6 to 12 months.47-49 A good response to local injection is sometimes useful as a diagnostic test and is a predictor of good surgical response. As noted later, care should be taken to avoid injection into the body of the median nerve.50

Most patients with de Quervain’s tenosynovitis involving the tendons at the base of the thumb respond to local steroid injection. In three prospective studies, 60% to 76% of patients with this condition had their symptoms adequately controlled with a single injection, and another 10% to 33% required a second injection.51,52 About 30% had exacerbations an average of 1 year later, but overall, only 10% to 17% of patients were not controlled and required surgical release. Another small controlled study showed that injection was much better than splinting.53 In patients with flexor tenosynovitis, steroids are effective in 88% compared with 36% of patients receiving saline injections, with improvement lasting for up to a year in most.54 Similar success rates have been reported in prospective studies of patients with ganglion cysts.55

In the ankle and foot area, injection therapy has been used to treat plantar fasciitis, tarsal tunnel syndrome, Achilles tendinopathy or bursitis, and interdigital neuroma (Morton’s neuroma). Most of the data about efficacy for these conditions are anecdotal and uncontrolled. Generally, the response in tarsal tunnel syndrome is temporary, whereas the response in Morton’s neuroma is more often prolonged. Reported response rates for plantar fasciitis vary, but one controlled trial showed a significant improvement at 1 month compared with placebo, whereas results were no different from placebo at 3 months.56,57 Studies of the efficacy of local corticosteroid injections for Achilles tendinopathy are inconclusive, and some have demonstrated an increased risk for Achilles tendon rupture.58

Local injections for neck pain and low back pain have been used for many years, with anecdotal reports of improvement, but controlled or prospective studies have shown variable results, depending on patient selection and methodology. Few controlled studies have assessed local trigger point or other soft tissue injections in the paracervical or
paralumbar areas. Most studies of radiographically assisted facet joint injection of steroids in the lumbar or cervical areas show no difference compared with placebo, facet block, or local paraspinal injections.\textsuperscript{69, 70} Injections of the sacroiliac joint in patients with noninflammatory pain have shown a slight benefit from steroids compared with lidocaine alone.

**PREPARATIONS**

**Corticosteroids**

All hydroxyxocorticosteroid preparations are effective for intra-articular and periarticular injections (Table 54-3). The originally injected hydrocortisone acetate is still available, widely used, and inexpensive. Triamcinolone hexacetonide is one of the least soluble agents with the presumed most prolonged effect. Not all preparations are equivalent in efficacy or duration of effect, but few studies have been done to compare the efficacy of the various preparations. More recent reports have shown that methylprednisolone is superior to triamcinolone acetonide, that triamcinolone hexacetonide has a more prolonged response compared with triamcinolone acetonide, and both of these are more effective than hydrocortisone.\textsuperscript{54, 65} Most clinicians have become familiar with certain preparations and have continued to use these with efficacy for years. Some clinicians prefer to inject combinations of short-acting and long-acting preparations. Steroid preparations are often mixed with local anesthetics, particularly for injecting small joints, tendon sheaths, and periarticular structures. Mixing with a local anesthetic reduces the local discomfort of injection into a confined space and dilutes the concentration of the locally injected steroid and reduces the risk of soft tissue atrophy. Guidelines for the dosage of steroid injected into given joints are based roughly on the size of the joint injected. Although no consensus exists regarding these amounts, most experts suggest injecting 1 mL of steroid preparation into large joints, with smaller amounts into smaller joints.

**Table 54-3 Injectable Preparations for Intra-articular Injection**

<table>
<thead>
<tr>
<th>Corticosteroids</th>
<th>Prednisone Equivalent (mg/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Betamethasone sodium phosphate (6 mg/mL)</td>
<td>50</td>
</tr>
<tr>
<td>Dexamethasone sodium (4 mg/mL)</td>
<td>40</td>
</tr>
<tr>
<td>Dexamethasone acetate (8 mg/mL)</td>
<td>80</td>
</tr>
<tr>
<td>Hydrocortisone acetate (24 mg/mL)</td>
<td>5</td>
</tr>
<tr>
<td>Methylprednisolone acetate (40 mg/mL)</td>
<td>50</td>
</tr>
<tr>
<td>Prednisolone (20 mg/mL)</td>
<td>20</td>
</tr>
<tr>
<td>Triamcinolone acetonide (40 mg/mL)</td>
<td>50</td>
</tr>
<tr>
<td>Triamcinolone hexacetonide (20 mg/mL)</td>
<td>25</td>
</tr>
</tbody>
</table>

**Hyaluronic Acid (Indicated in Osteoarthritis of Knee)**

**Hyaluronate Derivatives**

- Euflexxa: inject 20 mg (2 mL) once weekly for 3 wk
- Hylan: inject 20 mg (2 mL) once weekly for 5 wk;
  - some may benefit with a total of 3 injections
- Orthovisc: inject 30 mg (2 mL) once weekly for 3-4 wk
- Supartz: inject 25 mg (2.5 mL) once weekly for 5 wk

**Hylan Polymers**

- Synvisc: inject 16 mg (2 mL) once weekly for 3 wk
  - (total of 3 injections)
- Synvisc-One: inject 48 mg (6 mL) one time

**Other Injectable Products**

Over the years, many other agents have been injected into joints including salicylates, phenylbutazone, gold, orgotein, progesterone, glycosaminoglycan polysulfate, and various antibiotics, but most have been abandoned because of lack of efficacy or local reactions. Various cytotoxic agents have been used sporadically or in small numbers of patients for intrasynovial tumors and refractory proliferative synovitis including nitrogen mustard, osmic acid, and methotrexate. Radioactive preparations (yttrium-90, colloidal 12\textsuperscript{3}P chronic phosphate, dysprosium-165-ferric hydroxide, rhenium-186) have been used in both inflammatory and noninflammatory arthritis in occasional reports, but the evidence for efficacy and safety of these substances in patients with arthritis is limited.\textsuperscript{56}

Intra-articular hyaluronic acid preparations have been in use for many years in Europe and more recently in Canada and the United States. Several preparations of hyaluronic acid are approved for the treatment of osteoarthritis of the knee and seem to be superior to placebo injections in most, but not all, clinical trials, although no evidence for long-term efficacy or disease modification has been reported.\textsuperscript{33, 66, 69} These preparations are usually given weekly in a series of three to five injections, and more recent studies have used single injections or higher molecular weight preparations.\textsuperscript{70} In direct comparisons with corticosteroids, hyaluronate is generally less effective in the first 4 weeks but more effective between 8 and 26 weeks after injections.\textsuperscript{71} Studies of hyaluronic acid preparations in osteoarthritic knees, hips, and shoulders have shown inconclusive or minimal levels of improvement.\textsuperscript{73, 72} Hyaluronic injections may also be more effective than placebo in knees in rheumatoid arthritis and chronic painful rotator cuff disease, although these uses have not been as extensively studied.

Other injectable substances recently studied in small series for various conditions include botulinum toxin (for tennis elbow, painful shoulders, postoperative knee pain, and small joints in rheumatoid arthritis), EDTA for calcific tendinitis,\textsuperscript{73} and collagenase for Dupuytren’s contractures.\textsuperscript{74} In addition, platelet-rich plasma has been used for soft tissue injuries (e.g., Achilles tendinopathy, tennis elbow) with variable results in controlled studies.\textsuperscript{75}

**CONTRAINDICATIONS**

There are few contraindications to diagnostic arthrocentesis (Table 54-4). Established infection such as cellulitis in periarticular structures is generally considered to be an absolute contraindication to inserting a needle into a joint. If inflammation in an underlying joint or bursa is thought to be the cause of the appearance of infection, however, aspiration of the joint or bursa should be attempted. Septicemia carries the theoretic risk of introducing blood-borne bacteria into a joint, but such complications are not well documented and joints suspected of being infected should be aspirated regardless of the presence of septicemia. Arthrocentesis
through an area of irregular or disrupted skin, as seen in psoriasis, should be avoided because of the increased numbers of colonizing bacteria in these areas. Caution should be exercised in patients with bleeding disorders or patients taking anticoagulants, owing to the theoretic risk of inducing hemarthrosis. The risk of significant hemarthrosis after arthrocentesis is low, however, even in patients on regular warfarin therapy with international normalized ratios of 4.5.76

### COMPLICATIONS

Iatrogenic infection is the most serious, but least common, complication of arthrocentesis and joint injection (Table 54-5). In Holland’s large series, an incidence of infection of 0.005% was reported in a series of 400,000 injections. Gray and colleagues77 reported an incidence of 0.001% several years later. An infection rate of 1:2,000 to 1:10,000 (0.01% to 0.05%) has been noted in patients with rheumatoid arthritis, occurring exclusively in debilitated patients on immunosuppressive therapy.78 A recent national database review in Iceland has estimated an infection rate of 0.037% from arthrocentesis.79 Few other prospective or systematic studies of infection after arthrocentesis have been published, but most reported anecdotal experience has noted a similar low incidence of this serious complication.80

An arthroscopic study showed that a small fragment of skin stained with a surgical marking pen could be identified within the joint space after most percutaneous insertions of a needle into the joint space, with identifications of bacterial nucleic acid by polymerase chain reaction in about one patient.81 Considering the rarity of joint infection after arthrocentesis, these findings suggest that bacteria introduced at the time of arthrocentesis are either not viable or quickly cleared in almost all cases.

The most common complications of local steroid injections are related to local irritation of synovial and subcutaneous tissues and atrophy of soft tissues. Postinjection “flare” may develop in 1% to 6% of patients a few hours after injection and may last 48 hours, sometimes mimicking iatrogenic infection.82 These flares are reportedly more common with needle-shaped crystals and are believed to be similar to the acute arthritis related to other crystals phagocytosed by leukocytes, but they may also be caused by preservatives in some steroid suspensions.

Weakening of tendons and tendon rupture have also been reported as a result of locally injected steroids, emphasizing the importance of avoiding direct injection of steroids into the body of tendons. Most reports of tendon rupture have been anecdotal and described in patients involved in athletic activities or with rheumatoid arthritis. The risk of tendon rupture has not been adequately determined, but it seems to be quite low in the hands and wrists, where no ruptures were seen in a series of more than 200 injections84 and only 2 were seen in another series of 956 injections.85 Areas believed to be at highest risk for rupture include the Achilles tendon, bicipital tendon, and plantar fascia, where the risk of rupture has been estimated to be 10%.85

Systemic absorption occurs with locally injected depot corticosteroids. Since the earliest intra-articular injections of steroids, an anti-inflammatory effect has been shown not only in the injected joint but also in other joints in the same patient.86 Subsequent studies have documented decrease in plasma cortisol and suppression of the hypothalamic-pituitary axis lasting 2 to 7 days after a single injection. The degree and duration of adrenal suppression from a single intra-articular dose of depot steroid is less pronounced than that seen from an equal intramuscular dose.87

### Table 54-4  Contraindications to Arthrocentesis and Joint Injection

<table>
<thead>
<tr>
<th>Contraindication</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Established infection in nearby structures (e.g., cellulitis, septic bursitis)</td>
<td>Sometimes gout mimics cellulitis, creating a confusing picture</td>
</tr>
<tr>
<td>Septicemia (theoretic risk of introducing organism into joint)</td>
<td>Need to tap suspected septic joints in septic patients</td>
</tr>
<tr>
<td>Disrupted skin barrier (e.g., psoriasis)</td>
<td>Do not tap through lesions</td>
</tr>
<tr>
<td>Bleeding disorder (not absolute, but use more care)</td>
<td>Risk of bleeding very low, even in patients taking warfarin</td>
</tr>
<tr>
<td>Septic joint</td>
<td>Steroid injection contraindicated</td>
</tr>
<tr>
<td>Prior lack of response</td>
<td>Relative contraindication</td>
</tr>
<tr>
<td>Difficult-to-access joint</td>
<td>Relative contraindication without imaging aid</td>
</tr>
</tbody>
</table>

### Table 54-5  Potential Complications of Arthrocentesis and Joint Injection

<table>
<thead>
<tr>
<th>Complication</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iatrogenic infection</td>
<td>0.01%-0.05%; may be higher in RA patients</td>
</tr>
<tr>
<td>Postinjection “flare”</td>
<td>1%-6%; lasting 48 hr; may be related to preparation</td>
</tr>
<tr>
<td>Local soft tissue</td>
<td>May occur 1-6 mo later; pigment change</td>
</tr>
<tr>
<td>Local nerve damage</td>
<td>In structures near prominent nerves (e.g., carpal tunnel syndrome)</td>
</tr>
<tr>
<td>Tendon rupture or weakening</td>
<td>Case reports; animal studies show highest risk in Achilles tendon and plantar fascia</td>
</tr>
<tr>
<td>Systemic steroid absorption</td>
<td>Inevitable; usually subclinical</td>
</tr>
<tr>
<td>Avascular necrosis of bone</td>
<td>Hypothalamic-pituitary suppression 2-7 days; changes in bone formation 14 days</td>
</tr>
<tr>
<td>Negative effects on cartilage</td>
<td>Flushing; facial warmth; diaphoresis</td>
</tr>
<tr>
<td>Transient elevation of blood glucose; lymphopenia; eosinopenia</td>
<td></td>
</tr>
<tr>
<td>Controversial; reported but usually explained by underlying disease or systemic steroids in same patient (ischemic necrosis)</td>
<td></td>
</tr>
<tr>
<td>Found in animal models of normal cartilage but not in primates</td>
<td></td>
</tr>
<tr>
<td>Case reports in humans receiving multiple injections</td>
<td></td>
</tr>
<tr>
<td>Some animal models of arthritis are better with steroid injections</td>
<td></td>
</tr>
<tr>
<td>Large human observational studies have not documented more problems than expected (osteoarthritis, RA, juvenile RA)</td>
<td></td>
</tr>
</tbody>
</table>

RA, rheumatoid arthritis.
markers of bone turnover, a single injection of triamcinolone in knees of rheumatoid arthritis patients resulted in no change in bone resorption markers but yielded a drastic reduction in markers of bone formation within 1 day, which returned to normal levels in 14 days.\(^{87}\) A transient and variable effect on blood glucose levels after local steroid injection has also been observed in small studies, and unsurprisingly, this appears to be more pronounced in patients with diabetes.\(^{88}\)

Some patients experience prominent erythema, warmth, and diaphoresis of the face and torso within minutes to hours after steroid injections.\(^{93}\) This reaction is most likely related to systemic absorption, but idiosyncratic reaction to preservatives in steroid preparations has also been implicated. Similarly, some patients may experience other typical metabolic effects of systemic steroids such as transient increases in blood glucose or decreases in peripheral blood eosinophil or lymphocyte counts.

Avascular necrosis of bone (ischemic necrosis) has long been considered a potential complication of intra-articular steroids, with a reported prevalence of this complication in injected joints ranging from less than 0.1% to 3%.\(^{6,18,21}\) Most studies have suggested, however, that the occurrence of this complication is related more to the severity of the associated disease or systemic steroid therapy and is unrelated to local injections.

The potential for negative effects of locally injected corticosteroids on cartilage metabolism has been a controversial area of study for several decades. Anecdotal reports of Charcot-like arthropathy attributed to intra-articular steroids first appeared in the late 1950s and 1960s, often occurring in patients having more than 10 (and sometimes hundreds) joint injections over many months or years. Several studies in the 1960s and 1970s showed that locally injected steroids caused destructive changes, catabolic effects, or both in normal animal cartilage.\(^{89,90}\) These included findings of decreased protein and matrix synthesis with degenerative cellular changes in chondrocytes, as well as fissures and decreased proteoglycan content in cartilage matrix. Similar studies done in primate joints failed to show any negative effects from intra-articular corticosteroids, however.\(^{91}\)

Studies done in subsequent years have shown protective effects on cartilage lesions and reduction in osteophyte development in animal models of experimentally induced osteoarthritis, as well as associated reduction in metalloproteinase levels in cartilage and an increase in lubricating synovial surfactant.\(^{92,93}\) In humans with osteoarthritis, intra-articular steroids have been shown to decrease macrophage infiltration of the synovial lining, but no change was noted in metalloproteinase levels.\(^{94}\) Observations in humans treated with frequent corticosteroid injections have yielded conflicting information regarding changes in articular cartilage. More recent observations in patients with oligoarticular juvenile rheumatoid arthritis, mentioned previously, suggest that frequent steroid injections have the potential to help protect cartilage from the destructive process of the underlying disease process and are not associated with negative effects on articular cartilage.\(^{19,21,24}\) In addition, a study of patients with rheumatoid arthritis has shown no increase in the need for subsequent joint replacement surgery in the joints receiving four or more injections in a 1-year period.\(^{95}\)

### GENERAL ARTHROCENTESIS TECHNIQUES

#### Materials

Most practitioners find that an arthrocentesis tray containing needed items allows more flexibility in preparing for aspirating or injecting joints or other articular structures. Syringes greater than 20 mL are not necessary for most procedures, but a swollen knee may occasionally contain 60 mL or more of fluid and it is reasonable to have at least one syringe of this size in a tray, with others available for rare patients with effusions greater than 100 to 200 mL in volume. Heparinized or citrated tubes to prevent coagulation of inflammatory fluids for accurate cell counts and crystal analysis, plain tubes for chemistry evaluations, and sterile tubes for transporting fluid to a microbiology laboratory for culture should be included. In joints being aspirated for the presence of bacteria and crystals, small amounts of fluid or debris may be present in the bore of the needle, even when no obvious fluid is obtained. In such situations, it is best to have clean microscope slides and coverslips available at the bedside for microscopic examination for cellularity, Gram stain, and crystals.

A hemostat is helpful for changing syringes after aspiration to inject corticosteroid. The size and length of needles used depend on the anticipated amount of fluid to be obtained from a joint and the size of the involved joint. A 20- to 22-gauge needle is usually sufficient to aspirate most detectable effusions, but large effusions with large amounts of debris, as seen in septic joints, may require larger-bore needles. In small joints, a 23- to 27-gauge needle may be used, particularly when no fluid appears to be present, and only therapeutic injection is being considered. A needle 1½ inches in length is adequate for almost all procedures, but a 3-inch spinal needle may occasionally be necessary for a large knee or hip.

#### Site Preparation and Technique

Sterile technique designed to avoid the introduction of skin bacteria into the joint should be observed in all procedures, although precautions taken to avoid infection in clinical practice vary widely.\(^{90}\) After careful examination and identification of the specific point of aspiration, this point may be marked by the end of a ballpoint pen with the writing point retracted. The area should be carefully cleaned with one or two layers of iodine followed by alcohol. These precautions are sufficient to minimize the risk of infection, although a single “swipe” of isopropyl alcohol has been shown to provide a similar level of protection.\(^{96}\)

A physician experienced in arthrocentesis may elect to not use topical anesthesia in many patients because the amount of pain is often no more than that experienced from phlebotomy. In an anxious patient or when small joints or joints with minimal fluid are being aspirated, topical anesthesia may be attained by the use of spray coolant (ethyl chloride) or an intradermal wheal and subcutaneous infiltration of lidocaine. Spray coolant may be applied after sterile preparation and has been shown to not contaminate the field.\(^{97}\) In pediatric patients, particularly when multiple joints are injected, sedation or general anesthesia may be required for safe and accurate injections.\(^{98}\) Nonsterile gloves

should be worn by the operator to avoid contamination with the patient's synovial fluid or blood. Drapes and sterile gloves are unnecessary, but the gloved hand should not touch the prepared site. A joint is usually entered at a 90-degree angle to the skin, slowly and evenly, and negative pressure should be applied to the syringe when the needle has been advanced \( \frac{1}{2} \) to 1 inch (in a large joint). If the needle's course is obstructed by bone, the needle should be withdrawn slightly and redirected at a slightly different angle. If no fluid is obtained, the needle should be slowly advanced and negative pressure continued. If fluid flows initially and then stops, the needle may be advanced or retracted slightly or rotated in case it is blocked by an intra-articular structure or synovial tissue. After an adequate amount of fluid is obtained, a hemostat may be used to secure the needle, the syringe may be removed, and a new syringe with injectable steroid may be attached if injection is indicated. Overall accuracy of arthrocentesis has been estimated to be between 80% and 100% using anatomic landmarks in most joints. In “dry joints” a “backflow technique,” in which saline is injected and then aspirated back, may be used to help confirm correct needle positioning. A three-way stopcock is preferred by some practitioners when aspiration and injection are performed in the same setting.

After injection, the needle and syringe should be removed and pressure applied over the site until a bandage is applied. When synovial fluid is to be examined for crystals, care should be taken to not replace the needle used for steroid injection on the syringe with synovial fluid because the contamination of fluid with steroid crystals would make accurate identification of urate and calcium pyrophosphate crystals more difficult. In some situations, positioning the barrel of the syringe and simultaneous aspiration may result in difficulty controlling needle position, and a newly available one-handed reciprocating syringe may allow for better operator control of the syringe in more difficult aspirations.

Postprocedure Instructions and Care

After the procedure, patients should be reminded about the risk for postinjection “flare” within 24 to 48 hours after local steroid injection. If pain, redness, or swelling progresses afterward, are particularly severe, or are accompanied by fever, the patient should be instructed to call and be re-evaluated for the remote possibility of iatrogenic infection. Patients should also be reminded about the soft tissue atrophy and hypopigmentation that may occur weeks to months after the procedure, particularly when structures close to the skin are injected. For patients in whom tendon sheaths are injected, heavy activities using the involved tendons should be avoided for several days.

After injection of joints, particularly weight-bearing joints, activities should be limited owing to evidence that activity restriction prolongs the effect of the injected steroid. In a large series reported in 1995, McCarty and colleagues used a regimen that emphasized 3 weeks of splinting for upper extremity joints and 6 weeks of crutch walking for lower extremity joints, suggesting that rest after the injection was important in prolonging the effect of injections. Several other retrospective studies and anecdotal observations have suggested a role for strict rest or non-weight bearing after injection to improve duration of efficacy. One small controlled study showed that rest provided no advantage over regular activities in regard to short-term or long-term outcomes. In a larger prospective controlled trial of knees in patients with rheumatoid arthritis, a 24-hour period of strict bed rest after injection resulted in a more prolonged improvement, however, compared with patients who were not restricted. A similar study of elbow injections showed no benefit from sling immobilization. Most practitioners advise restricted activities after steroid injections, particularly in weight-bearing joints, but opinions on the relative importance of rest after injections still vary and no specific regimen of rest would be considered standard practice among physicians.

Specific Regional Arthrocentesis Techniques

Cervical Spine Area

Most injections in the region of the cervical spine, trapezius, and scapular areas are best considered to be myofascial or trigger point injections. The point of injection is usually determined by palpating for the areas of most tenderness, and few reliable landmarks are available for localization of anatomic structures. Using a 22- to 25-gauge needle, a combination of 0.5 mL of steroid and 0.5 to 2 mL of lidocaine can be injected into areas in the paracervical muscles or other areas where tenderness can be elicited. Some of the injections done in these areas are likely to be close enough to the posterior cervical facet joints to reduce inflammation in the joints themselves, whereas others probably reduce inflammation in ligaments, tendons, or bursal structures. More precise injection of posterior cervical facet joints may be accomplished under fluoroscopic visualization.

Anterior Chest Wall

The anterior chest area occasionally may be the site of inflammatory disease of the sternoclavicular joints. It is usually difficult to obtain much fluid for diagnostic purposes from this joint, but a few drops for microscopic analysis and culture may be available in some patients. The point of aspiration should be dictated by the point of maximal swelling near the surface. Aspiration should be attempted with caution, using as short and small a needle as possible to avoid damage to nearby vascular structures, lung, or airway. A small amount of steroid (0.1 to 0.5 mL) may be injected in this joint, as long as the suspicion of infection is low. In Tietze’s syndrome, one or more of the anterior costochondral junctions may be swollen, but such areas do not contain fluid for analysis and should be approached with caution for injection of small amounts of steroid and lidocaine. Inflammation of the manubriosternal joint in patients with spondyloarthropathy may be improved by fluoroscopically guided injection.

Temporomandibular Joint

The temporomandibular joint may be involved in patients with rheumatoid arthritis, spondyloarthropathies, or
osteoarthritis, and internal derangement of this joint may be a source of discomfort usually treated by oral surgeons.

This joint is palpated as a depression just below the zygomatic arch about 1 to 2 cm anterior to the tragus. The depression is usually more easily palpated by having the patient open and close the mouth. After a mark is made over the area, a 22- to 25-gauge needle is inserted perpendicular to the skin and directed slightly posteriorly and superiorly; 0.1 to 0.25 mg of steroid preparation can be injected. This joint seldom has enough fluid for diagnostic aspiration.

**Shoulder**

Numerous structures in and around the shoulder may be involved in systemic processes, injury, or overuse syndromes, and each has the potential to benefit from local steroid injection. Ideally, injections should be directed toward specific anatomic sites on the basis of clinical findings. For practical purposes, injections into the glenohumeral joint or subacromial bursal space are often beneficial for pain related to other nearby structures such as the rotator cuff tendons or bicipital tendon.

**Glenohumeral Joint**

The glenohumeral joint may be entered from an anterior or posterior approach. For an anterior approach, the patient should be in a sitting position with the shoulder externally rotated (Figure 54-1). A mark is made just medial to the head of the humerus and slightly inferiorly and laterally to the coracoid process. A 20- to 22-gauge, 1½-inch needle is directed posteriorly and slightly upward and laterally. For a posterior approach, the upper arm should be against the lateral chest and forearm across the chest. A mark should be made about 2 inches inferior to the acromion. One should be able to feel the needle enter the joint space, but if bone is hit, the needle should be pulled back and redirected at a slightly different angle. When the joint is entered, fluid should be aspirated, if present, and the joint should be injected with 1 mL of steroid preparation, with or without 1 to 3 mL of lidocaine.

**Acromioclavicular Joint**

The acromioclavicular joint, similar to the sternoclavicular joint, is composed of fibrocartilage and rarely contains fluid. The joint is palpated as a groove at the lateral end of the clavicle, just medial to the tip of the acromion, and may display some degree of soft tissue swelling or bony prominence, depending on the underlying disease process. To enter the joint with a needle, a mark should be made over the groove, and a 22- to 25-gauge needle should be introduced 1 inch or less. After an attempt to obtain fluid, 0.25 to 0.5 mL of steroid preparation may be injected.

**Rotator Cuff Tendon and Subacromial Bursa**

The rotator cuff tendon and subacromial bursa area may be entered using a 22- to 25-gauge needle, usually 1½ inches in length (Figure 54-2). Over the lateral or posterolateral aspect of the shoulder, the groove between the acromion (marked in black) and humerus is palpated and marked (spot). The needle is inserted at this point and advanced in a horizontal plane medially.

**Bicipital Tendon**

Bicipital tendinitis may be treated by injecting the shoulder joint or the tendon sheath itself. If the tendon is to be injected, it can be palpated over the anterior aspect of the shoulder in the bicipital groove of the shoulder; it is usually tender and can be rolled under the examiner’s finger. A 22-gauge, 1½-inch needle should be inserted in the sheath, and portions of the steroid and lidocaine
weeks or months after the injection. The medial epicondyle is injected in a similar fashion, with more care required to avoid inadvertent injection of steroids into the area of the ulnar groove just behind the bony prominence of the epicondyle.

Olecranon Bursa and Nodules

The area of the olecranon bursa and nodules is located just under the skin over the tip of the olecranon process at the posterior aspect of the elbow. Swelling in this area is detected easily as a localized collection of fluid and can be easily aspirated and injected if fluid is present. A smaller-gauge needle (22 to 23 gauge) may be used for noninflammatory processes, but a larger gauge (20-gauge) is often necessary for bursal effusions related to rheumatoid arthritis or gout. After preparation, the needle should be inserted under the skin into the easily palpable area of fluid, and as much fluid as possible should be aspirated. For noninfectious processes, 0.5 to 1 mL of steroid can be injected into the space. Subcutaneous nodules in this area, or at other locations around the body, may be aspirated for diagnostic purposes, usually to differentiate rheumatoid nodules from tophi. For this type of aspiration, an 18- to 20-gauge needle is inserted into the nodule and rotated, retracted to near the surface, and then reinserted and rotated, with negative pressure applied on the syringe. After removal, the contents of the syringe should be expelled onto a microscope slide and may be examined for cellular content and crystals.

Wrist and Hand

Radiocarpal Joint

The wrist joint is complex, but most of the intercarpal spaces communicate with the radiocarpal joint, which may be entered from a dorsal approach. A mark should be made just distal to the radius and just ulnar to the “anatomic snuffbox” (Figure 54-5). A 22- to 25-gauge needle, \( \frac{1}{2} \) to 1

Medial and Lateral Epicondyle

The areas of the medial and lateral epicondyle are commonly affected by overuse syndromes involving the origins of muscle groups of the forearm, particularly the lateral epicondyle, which is the area of inflammation in tennis elbow (Figure 54-4). With the elbow flexed, the area of tenderness over the anterolateral surface of the external condyle of the humerus should be marked. After preparation, a 22- to 25-gauge needle 1 to 1\( \frac{1}{2} \) inches long should be inserted about 1 to 2 cm distal to the mark; 0.5 mL of steroid preparation mixed with 1 to 3 mL of lidocaine is administered in several small doses after partially withdrawing and redirecting the needle and reinjecting in two to three passes of the needle in the area. Injections in this area often deliver steroids to subcutaneous tissues close to the skin, and patients should be advised about the likelihood of subcutaneous atrophy and pigment changes that may occur.
inch long, is usually adequate. Occasionally, 3 to 5 mL of fluid may be obtained from the wrist by aspiration, and, if indicated, 0.5 mL of steroid may be injected into the space.

**Dorsal Wrist Tendons**

The extensor tendon sheaths over the dorsal wrist may become inflamed and swollen secondary to numerous inflammatory processes, most commonly rheumatoid arthritis, but occasionally crystal-induced arthritis or infectious processes. The areas of swelling are well defined and close to the surface, and they are entered easily with a direct aspiration, usually at a 30- to 45-degree angle, with the needle directed along the course of the swollen tendon. Fluid is often easily obtained, but in some patients, in particular, patients with rheumatoid arthritis, proliferative synovial tissue limits the amount of fluid that can be aspirated. After aspiration, the area can be injected with 0.5 mL of corticosteroid mixed with 0.5 to 1 mL of lidocaine, if indicated.

**de Quervain’s Tenosynovitis**

de Quervain’s tenosynovitis, a common overuse syndrome involving the tendons at the radial aspect of the anatomic snuffbox, is often helped by local injection of the tendon sheath. After examination, the area of greatest tenderness along the course of the tendon should be marked, and the needle should be inserted almost parallel to the skin, either proximally or distally (Figure 54-6). As the needle is advanced, 0.5 mL of steroid with 0.5 to 2 mL of lidocaine can be injected along the sheath of the tendon, and a palpable bulge is usually felt along the tendon. Care should be taken to avoid injection of steroid into the body of the tendon.

**Carpal Tunnel Syndrome**

Inflammation with swelling in the many flexor tendons in the carpal tunnel area may result in median nerve compression, and injection in this area has the potential to relieve symptoms by reducing this inflammation (Figure 54-7). This area should be injected by making a mark on the volar aspect of the wrist along the flexor carpi radialis tendon (on the radial side of the long palmar tendon), approximately 1 to 2 cm proximal to the distal wrist crease. A 22- to 26-gauge needle may be introduced perpendicular to the skin or, alternatively, at a 30- to 45-degree angle, directing the needle proximally or distally along the course of the tendon. The needle should be introduced about ½ to 1 inch, and the area is injected with 0.5 mL of
steroid with 0.5 to 1 mL of lidocaine. If the needle meets obstruction or if the patient experiences paresthesias, the needle should be withdrawn and redirected slightly to avoid injecting into the body of a tendon or into the median nerve itself.

**Ganglia**

Small, often hard, nodular structures known as ganglia are frequently present around the hands and wrists, and they may occur in many other areas near joints or tendons. These structures usually contain a thick, gelatinous substance that is difficult to aspirate. In cases in which pain, tendon dysfunction, or nerve entrapment symptoms are bothersome to the patient, aspiration may be attempted, usually with an 18- to 20-gauge needle. Even if no fluid is obtained, the process of puncture occasionally causes the structure to dissipate its contents and symptoms are relieved. A small amount (0.2 to 0.5 mL) of steroid with lidocaine may be injected in an attempt to prevent reaccumulation of fluid.

**Thumb Carpometacarpal Joint**

Aspiration of fluid from this joint is seldom possible and rarely indicated. This joint is commonly involved in osteoarthritis, however, and may be a source of localized pain amenable to local injection (Figure 54-8). The joint space is narrowed and often surrounded by osteophytes, but it may be entered accurately by flexing the thumb across the palm and making a mark at the base of the thumb metacarpal away from the border of the snuffbox. A 22- to 25-gauge needle should be inserted $\frac{1}{2}$ to 1 inch at this mark and directed away from the radial artery; 0.2 to 0.5 mL of steroid may be injected.

**Metacarpophalangeal and Interphalangeal Joints**

Inflammation in the small joints of the hands usually causes the synovium to bulge dorsally. Occasionally, these small joints may have enough swelling for a drop or two of fluid to be obtained for crystal analysis, culture, or both. In most cases, arthrocentesis is performed for symptom relief; 0.1 to 0.2 mL of corticosteroid is injected, with or without local anesthetic. A 24- to 27-gauge, $\frac{1}{2}$-to 1-inch needle can be inserted on either side of the joint at a mark made at the joint line, just under the extensor tendon mechanism. Some physicians prefer to have the joint slightly flexed to improve the chances for entry into the joint space itself.

**Flexor Tenosynovitis (Trigger Fingers)**

The pathologic process in this condition usually involves the tendon at the level of the metacarpophalangeal joint in the palm. Usually, a localized swelling that moves with the tendon sheath may be palpated in this area. After a mark has been made at this area, a 22- to 27-gauge needle may be introduced at a 30- to 45-degree angle, directing the needle proximally or distally along the course of the tendon. The needle should be introduced about $\frac{1}{2}$ inch, and the area may be injected with 0.5 mL of steroid with 0.5 mL of lidocaine. Lack of resistance during injection indicates proper needle placement, as is the case with other tendon sheath injections.

**Lumbosacral Spine Area**

Back pain is difficult to explain anatomically in most patients, but many patients with back pain have areas tender to deep palpation, particularly in the presacral area and paraspinal muscles. As is the case in the cervical spine area, most injections in this region are best considered to be myofascial or trigger point injections. The point of injection is usually determined by palpating for the areas of most tenderness, with few reliable landmarks available for localization of anatomic structures. Using a 22- to 25-gauge needle, a combination of 0.5 to 1 mL of steroid and 0.5 to 2 mL of lidocaine can be injected into areas of the paraspinal muscles or other tender areas. Some of the injections performed in these areas are likely to be close enough to the posterior lumbar facet or upper sacroiliac joints to reduce inflammation in the joints themselves, whereas others probably reduce inflammation in ligaments, tendons, or bursal structures. More precise injection of posterior lumbar facet or sacroiliac joints requires radiographic guidance, with fluoroscopy, CT, or MRI.

**Pelvic Girdle**

**Ischiogluteal Bursitis**

The bursa over the ischial tuberosity is located by direct palpation in the buttock while the patient lies on the opposite side with knees fully flexed. The prominence is more easily palpated as the gluteus muscles are displaced from the area. After marking the area of tenderness over the prominence, a 3-inch needle should be inserted horizontally until bone is hit and 1 mL of steroid with 1 to 2 mL of lidocaine is instilled. Care should be taken to avoid the sciatic notch medially.
**Trochanteric Pain Syndrome (Bursitis)**

Trochanteric pain syndrome is diagnosed by physical findings of normal hip joint motion and a reproducible tender area in the region of the greater trochanter where the gluteal muscles insert (Figure 54-9). This area can be injected easily after marking the area of tenderness with the patient lying on the opposite side. A 1- to 3-inch needle should be inserted perpendicular to the skin of the bony prominence, and 1 mL of steroid with 2 to 4 mL of local anesthetic are injected into the area.

**Hip (Acetabular) Joint**

The hip is a difficult joint to aspirate and inject, and synovial fluid is seldom obtained from the hip in clinical practice. Two approaches can be attempted—either an anterior or a lateral approach—but accuracy of each varies. A cadaver study found the rate of correct needle placement was only 60% with the anterior approach and 80% with the lateral approach, and the anterior approach frequently resulted in needle placement in the vicinity of the femoral nerve. Accuracy using a lateral approach may improve with experience.

If aspiration without fluoroscopic guidance is attempted, a 20-gauge, 3-inch needle should be used. For the anterior approach, the patient should be supine with the hip fully extended and externally rotated. A mark should be made 2 to 3 cm below the anterior superior iliac spine and 2 to 3 cm lateral to the femoral pulse. The needle is inserted at a 60-degree angle and directed posteriorly and medially until bone is hit. The needle is withdrawn slightly, and an attempt should be made to aspirate fluid. Injection of 1 mL of steroid with lidocaine may follow if indicated. For a lateral approach (Figure 54-10), the patient should be supine and the hips rotated internally with knees apart and toes touching. A mark should be made just anterior to the greater trochanter, and the needle should be inserted and directed medially and slightly cephalad toward a point slightly below the middle of the inguinal ligament. Often, the clinician can feel the tip of the needle slide into the joint, and aspiration can be attempted.

**Knee**

**Knee Joint**

The knee is the easiest joint to enter with certainty by arthrocentesis and is the joint most frequently aspirated for synovial fluid analysis in clinical practice. The knee may be aspirated with the patient in the supine or sitting position and from medial, lateral, or anterior aspects. Aspiration is usually considered to be easiest with the patient in the supine position with the knee almost fully extended. A mark should be made just posterior to the medial or lateral aspect of the patella in the recess behind the patella, where a bulge or “fluid wave” can be detected on physical examination if fluid is present (Figures 54-11 and 54-12). An 18- to 22-gauge needle should be directed posteriorly and slightly inferiorly, and fluid can be aspirated after advancing the needle 1/2 to 1 1/2 inches into the joint space.

A lateral approach is preferred by some clinicians because more fluid can be removed from this side in many patients. In patients with rheumatoid arthritis or septic arthritis, synovial debris or proliferative synovial tissue may occlude the needle and it may be necessary to rotate the needle to facilitate aspiration. In some patients, the knee can be aspirated and injected with the patient in a sitting position with the knee flexed. A mark can be made just below the distal border of the patella in the recess on either side of the infrapatellar tendon. One more recent study showed that a lateral midpatellar approach had a higher accuracy rate than the anteromedial or anterolateral approaches done in a
Injection of corticosteroid into the knee joint usually results in the passage of medication into the cyst with a therapeutic effect.

**Periarticular Knee Pain Syndromes**

Pain related to the anserine bursa, located over the medial aspect of the tibia just below the joint line, can be treated with a local injection into this area. In addition, some patients may have an area of localized soft tissue tenderness above the joint line over the medial and lateral condyles, often believed to be related to an irritated iliotibial band (lateral) or infrapatellar plica (medial). In each of these conditions, the area of tenderness should be marked and a 22-gauge needle should be introduced to the bone and withdrawn slightly. The area should be injected with 0.5 mL of steroid with 1 to 3 mL of local anesthetic. Prepatellar bursitis may result in swelling in the soft tissues anterior to the patella and should be distinguishable from a knee effusion. Similar to olecranon bursitis at the elbow, this area may be aspirated directly at the point of maximal swelling. The area may be injected with 0.5 mL of corticosteroid preparation if indicated.

**Ankle and Foot**

**Tibiotalar Joint**

The tibiotalar joint is best aspirated with the patient in a supine position with the leg-foot angle at 90 degrees (Figure 54-13). A mark is made just medial to the tibialis anterior tendon and lateral to the medial malleolus. A 20-to 22-gauge, 1½-inch needle is directed posteriorly and should enter the joint space without striking bone. If resistance is felt and no fluid is obtained, the needle should be withdrawn to close to the surface and redirected slightly while aspirating for fluid. After aspiration, the area may be injected with 0.5 mL of corticosteroid, with or without lidocaine.
Plantar Fascia
Tenderness is usually elicited along the course of the plantar fascia and at its insertion at the calcaneus. This area may be injected by inserting a 22- to 25-gauge, 1½-inch needle from the lateral or medial aspect of the heel and directing it through the tissues of the heel pad toward the area of tenderness. About 0.5 mL of corticosteroid with 0.5 to 1 mL of lidocaine can be injected. Repeated injections in this area should be avoided because of the risk of plantar fascia rupture.

Metatarsophalangeal Joints
The small joints of the toes are aspirated and injected using techniques similar to those of the small joints of the hands. The metatarsophalangeal joint of the great toe is an area of particular interest because this joint is involved so often in gout. This joint may be aspirated in patients with a history suggestive of gout, even between attacks, and may yield enough fluid, sometimes to allow a diagnosis of gout to be confirmed by microscopic analysis. A small mark can be made at the joint line over the medial aspect of this joint, a 22- to 25-gauge needle can be inserted, and the area can be aspirated. Care should be taken to express the tiny amount of fluid often found in the needle hub onto a microscope slide. These joints can be injected for therapeutic benefit, usually with 0.1 to 0.25 mL of steroid.

Interdigital Neuroma
Interdigital neuroma causes pain between the metatarsal heads in the foot, usually between the second and third or third and fourth toes (Figure 54-15). Injection into this space may reduce inflammation and relieve symptoms of nerve compression between the metatarsal heads. The area is injected most easily from the dorsal aspect. The area of tenderness should be marked, a 22- to 27-gauge needle may be inserted ½ to 1 inch, and 0.25 to 0.5 mL of steroid with equal volume of anesthetic should be injected.

Subtalar Joint
Swelling from the subtalar joint is usually detected by swelling beneath the lateral malleolus. A mark is made just inferior to the tip of the lateral malleolus, usually over the area of swelling. A 20- to 22-gauge needle should be directed perpendicular to the skin, and the area should be aspirated as the needle is advanced (Figure 54-14). The needle may be withdrawn partially and advanced again if no fluid is obtained with the first pass, and 0.5 mL of corticosteroid may be injected after fluid is aspirated, if indicated.

Achilles Tendon Area
Generally, the area around the insertion of the Achilles tendon on the calcaneus should not be approached with a needle. In some patients, an area of swelling may be detected, however, in the subcutaneous Achilles bursa between the skin and tendon or in the retrocalcaneal bursa between the tendon and calcaneus. In either case, the area may be aspirated for fluid analysis, using a lateral or medial approach for the deeper area, to avoid inserting the needle through the Achilles tendon. In rare patients, these areas or the sheath of the Achilles tendon itself may be injected with 0.25 to 0.5 mL of steroid preparation with lidocaine. As noted previously, any injection in this area should be undertaken with extreme care to avoid injection within the body of the Achilles tendon, owing to the risk and potential consequences of rupture.

Tarsal Tunnel Syndrome
Tarsal tunnel syndrome, an uncommon condition, is sometimes amenable to local steroid injection, although the optimal approach to therapy in this condition is uncertain. The skin should be marked just inferior and posterior to the medial malleolus. The tendon sheaths in this area may be injected with a 22- to 25-gauge needle with 0.5 mL of steroid and 0.5 to 1 mL of lidocaine. Care should be taken not to inject the nerve.
CURRENT AND FUTURE TRENDS IN ARTHROCENTESIS AND JOINT INJECTION

Ultrasound-Guided Arthrocentesis and Injection

In recent years, ultrasound guidance has been the subject of numerous studies as a means to increase accuracy for needle placement for arthrocentesis and therapeutic steroid injection in joints and tendon sheaths, particularly in Europe, and is being used more often in the United States as well. One comparative study showed that ultrasound guidance increased the ability to obtain synovial fluid from joints to 97% of patients compared with 32% when using conventional techniques without ultrasound. Ultrasound has been shown to improve response to injection in many studies of painful sacroiliac joints, shoulders, fingers, and ankles compared with injections using traditional anatomic landmarks. However, some studies have shown no improvement in clinical effect with ultrasound guidance, possibly because many patients may benefit from periarticular steroid injection using anatomic landmarks without ultrasound guidance.

Ultrasound can be employed by having the area to be aspirated marked by the ultrasonographer or by having concurrent ultrasound monitoring while the needle is inserted into the joint or tendon sheath area. The latter approach has the potential to be particularly helpful in areas difficult to assess, but it is more cumbersome and requires the use of sterile components for the ultrasound machine and sterile ultrasound gel. In current practice, musculoskeletal ultrasound has the potential to improve outcomes in individual patients with pain associated with difficult-to-access joints and periarticular structures. Further studies to better define cost and clinical outcomes will be necessary to justify the more routine use of ultrasound to assist arthrocentesis and injection in clinical rheumatology practice.

Joint Irrigation

The irrigation of joints in osteoarthritis with large volumes of saline, known as tidal irrigation, was controversial from 1992 to 2002. On the basis of observations that some patients undergoing arthroscopy seemed to improve from the process of lavage that accompanies the procedure, subsequent studies suggested that irrigation of the joint was superior to medical management and comparable with local steroid injection. More recent studies using “sham” irrigation as control and recent meta-analyses and reviews of multiple controlled trials have shown no benefit from lavage compared with placebo or added to corticosteroid in knee osteoarthritis.

Intra-articular Biologic Therapy

Advances in understanding of the underlying biologic processes in rheumatoid arthritis and osteoarthritis in recent years have raised hopes that intraarticular therapy with biologic agents might have a role in the treatment of various forms of arthritis. Potential intraarticular therapies might include agents that suppress inflammation and bone destruction (e.g., interleukin-1, interleukin-4, or tumor necrosis factor inhibitors, adrenomedullin, bradykinin) or agents that promote cartilage growth (e.g., dehydroepiandrososterone, insulin-like growth factor, transforming growth factor-β). To date, clinical trials of intra-articular biologic therapies have been disappointing. Intra-articular tumor necrosis factor inhibitors have shown improvement that has been either inconsistent or inferior to local corticosteroid injections. A study of an interleukin-1 inhibitor has shown no difference compared with placebo injection in osteoarthritis.

The ability to transfer genes to synoviocytes in vivo and ex vivo has led to speculation that the delivery of genes directly to the joints might lead to reduced inflammation and cartilage degradation in rheumatoid arthritis and osteoarthritis.

Numerous strategies have been outlined, and animal studies and early human trials of intraarticular gene therapy using viral vectors for rheumatoid arthritis, osteoarthritis, and other diseases of the joints are ongoing. The role of such therapies in the future will depend on the ability to find safe and effective vectors that can deliver genetic material that persists in synovial tissues for long periods.

Selected References


33. Joshi R: Intraarticular corticosteroid injection for first carpmetacar


Full references for this chapter can be found on www.expertconsult.com.


88. Young L, Katrib A, Cuello C, et al: Effects of intraarticular glucocorticoids on macrophage infiltration and mediators of joint damage in...
Antinuclear antibodies (ANAs) include an ever-growing diversity of autoantibodies directed against multiple intracellular antigens, classically consisting of nuclear specificities such as deoxyribonucleic acid (DNA) or small nuclear ribonucleoproteins (snRNPs).\(^1\,^2\) ANA diseases (Table 55-1) include syndromes whose patients are characterized by an unusually high prevalence of ANAs, often screened for by the fluorescent antinuclear antibody (FANA) test: systemic lupus erythematosus (SLE), systemic sclerosis (SSc), and mixed connective tissue disease (MCTD). The prevalence of ANAs in polymyositis (PM), dermatomyositis (DM), and Sjögren’s syndrome (SS) has been reported to be somewhat lower than in other ANA diseases, but they are often grouped together because they share similar target antigens and therefore presumably similar causes. For decades, ANA testing has remained an important diagnostic and prognostic tool for these connective tissue diseases and has provided routine assays in the evaluation of patients with suspected rheumatic disease. However, these autoantibodies arise in a variety of infectious, inflammatory, and neoplastic diseases, as well as in normal individuals; therefore some knowledge regarding their intricacies and limitations of the assays is required for appropriate clinical application. This chapter summarizes the more common ANA specificities, outlining their history, implicated relevance to pathogenesis, methods of detection, and clinical disease associations, as well as the biology of their target autoantigens, in an effort to help guide the clinical utility of testing for these specificities.

**HISTORY**

In the first formal description of an ANA-related phenomenon in 1948, bone marrow specimens from a patient with SLE were found to contain lupus erythematosus (“LE”) cells, which were subsequently found useful in the diagnosis of SLE and drug-induced lupus, as well as SS and rheumatoid arthritis (RA).\(^1\) LE cells were soon discovered to be the result of a plasma factor, autoantibody against deoxyribonucleoprotein, which opsonized and/or induced apoptosis in free-cell nuclei, resulting in antibody sensitization and phagocytosis by polymorphonuclear neutrophils. In 1957, indirect immunofluorescence allowed the development of the FANA as a more sensitive assay for SLE and related diseases.\(^3\) Finer distinction of autoantibody reactivities...
Because of the characteristic presence of these autoantibodies among the ANA diseases, ANAs have long been speculated to play a role in disease pathogenesis. Anti-DNA antibodies, for instance, have been suspected to promote inflammation in SLE nephritis via immune complex deposition, direct binding to cross-reactive glomerular antigens, and/or intracellular penetration and induction of cellular toxicity. Similarly, ribonucleoprotein antibodies such as anti-Ro/SSA, anti-La/SSB, and anti-Sm have been implicated in the pathogenesis of cutaneous or cardiac manifestations by penetrating live cells and/or binding to exposed antigens in the skin and/or the heart. Sera containing anti-Scl-70 (topoisomerase I) activity can induce high levels of interferon (IFN)-α, correlating with diffuse cutaneous scleroderma and lung fibrosis; also, anti-Jo-1- or anti-Ro/SSA-positive sera from myositis patients have been demonstrated to induce type I IFN and/or intercellular adhesion molecule (ICAM)-1 on endothelial cells. However, autoantibodies alone appear insufficient to account for disease pathogenesis. Induction of type I IFN activity by anti-Ro/SSA-containing sera, for instance, appears restricted to patients with SLE or SS, not asymptomatic individuals, and surface binding of topoisomerase I may be required for a pathogenic effect of anti-Scl-70 antibodies. This may reflect additional biologic issues among or effects of the autoantigens themselves, such as novel conformations or epitopes: for instance, a proteolytically sensitive conformation of histidyl-transfer RNA synthetase (HisRS), the target of the pulmonary fibrosis–related Jo-1–specific antibody, has been described in the lung, and an apoptope (epitope expressed on apoptotic cells) of Ro/SSA may be specific to SLE, suggesting a unique role of apoptosis in disease pathogenesis. The autoantigens themselves may have unique biologic functions: 60-kD Ro/SSA, for instance, may serve as a receptor for the antiphospholipid–related β2-glycoprotein I, and this dynamic may account for differences in Ro antibody pathogenicity. However, many autoantigens likely have intrinsic proinflammatory properties, such as stimulation of innate inflammation by DNA and RNA via Toll-like receptors (TLRs) or induction of smooth muscle responses by the centromere protein CENP-B via CCR3. Apparent remission of SLE in a patient has been correlated with loss of TLR responsiveness, antibody deficiency, and disappearance of anti-DNA, supporting such concepts. Thus the pathogenesis of the connective tissue diseases appears to reflect a complex interplay between direct inflammatory or other biologic effects of the autoantigens and consequences of autoantibody responses.

### METHODS OF DETECTION

**KEY POINTS**

The gold standard screening test for ANAs is the fluorescent ANA test. Many antibody tests, including ANA screening in some laboratories, are performed via enzyme-linked immunosorbent assay (ELISA) because this method affords higher throughput testing, but this technique often results in lower specificity.

Optimal clinical interpretation of ANA tests requires knowledge of the technique(s) used in each specific case.

---

**Table 55-1 Antinuclear Antibody (ANA)-Associated Diseases and Related Conditions**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Patients with ANAs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diseases for Which ANA Testing Is Helpful for Diagnosis</strong></td>
<td></td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>99-100</td>
</tr>
<tr>
<td>Systemic sclerosis</td>
<td>97</td>
</tr>
<tr>
<td>Polymyositis/Dermatomyositis</td>
<td>40-80</td>
</tr>
<tr>
<td>Sjögren’s syndrome</td>
<td>48-96</td>
</tr>
<tr>
<td><strong>Diseases in Which ANA Is Required for Diagnosis</strong></td>
<td></td>
</tr>
<tr>
<td>Drug-induced lupus</td>
<td>100</td>
</tr>
<tr>
<td>Mixed connective tissue disease</td>
<td>100</td>
</tr>
<tr>
<td>Autoimmune hepatitis</td>
<td>100</td>
</tr>
<tr>
<td><strong>Diseases in Which ANA May Be Useful for Prognosis</strong></td>
<td></td>
</tr>
<tr>
<td>Juvenile idiopathic arthritis</td>
<td>20-50</td>
</tr>
<tr>
<td>Antiphospholipid antibody syndrome</td>
<td>40-50</td>
</tr>
<tr>
<td>Raynaud’s phenomenon</td>
<td>20-60</td>
</tr>
<tr>
<td><strong>Some Diseases for Which ANA Typically Is Not Useful</strong></td>
<td></td>
</tr>
<tr>
<td>Discoid lupus erythematosus</td>
<td>5-25</td>
</tr>
<tr>
<td>Fibromyalgia</td>
<td>15-25</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>30-50</td>
</tr>
<tr>
<td>Relatives of patients with autoimmune disease</td>
<td>5-25</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>25</td>
</tr>
<tr>
<td>Idiopathic thrombocytopenic purpura</td>
<td>10-30</td>
</tr>
<tr>
<td>Thyroid disease</td>
<td>30-50</td>
</tr>
<tr>
<td>Patients with silicone breast implants</td>
<td>15-25</td>
</tr>
<tr>
<td>Infectious disease</td>
<td>Varied widely</td>
</tr>
<tr>
<td>Malignancies</td>
<td>Varied widely</td>
</tr>
</tbody>
</table>

**Healthy ("Normal") Individuals**

<table>
<thead>
<tr>
<th>ANA titer</th>
<th>Patients with ANAs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥1:40</td>
<td>20-30</td>
</tr>
<tr>
<td>≥1:80</td>
<td>10-12</td>
</tr>
<tr>
<td>≥1:160</td>
<td>5</td>
</tr>
<tr>
<td>≥1:320</td>
<td>3</td>
</tr>
</tbody>
</table>


detected by FANA testing led to the description of Smith (Sm), nuclear ribonucleoprotein (nRNP), Ro/Sjögren’s syndrome (SSA), and La/SSB specificities, which later gained further biologic prominence with the demonstration that their autoantigens play prominent roles in cellular homeostasis (e.g., snRNPs, targets of anti-Sm and anti-nRNP) and in regulation of premessenger RNA splicing. Subsequent investigations have revealed an ever growing array of autoantigens (Table 55-2), many of which remain largely uncharacterized. Thus, ANAs not only serve as diagnostic markers in autoimmunity but to this day have greatly aided studies on cellular biochemistry.

### RELEVANCE OF ANTINUCLEAR ANTIBODIES TO DISEASE PATHOGENESIS

**KEY POINT**

ANAs and their respective autoantigens have been implicated in disease pathogenesis as producing directly toxic or other proinflammatory effects.
# Table 55-2 Diagnostic Characteristics of Antinuclear Antibodies (ANAs)

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Target Autoantigen (Function)</th>
<th>ANA Pattern(s)</th>
<th>Other Tests</th>
<th>Primary Rheumatic Disease Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nuclear</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chromatin-Associated Antigens</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>dsDNA, ssDNA</td>
<td>Rim, homogeneous</td>
<td>RIA, ELISA, CIF</td>
<td>SLE</td>
</tr>
<tr>
<td>Histone</td>
<td>(Nucleosome structure) H1, H2A/B, H3, H4</td>
<td>Rim, homogeneous</td>
<td>RIA, ELISA, CIF</td>
<td>SLE, DIL, RA</td>
</tr>
<tr>
<td>Kinetochore (centromere)</td>
<td>H3</td>
<td>Large speckles</td>
<td>IF, ELISA</td>
<td>SLE, UCTD, SSc, SLE, SS</td>
</tr>
<tr>
<td>Ku</td>
<td>CENP-A, -B, -C, and/or -D (mitotic spindle apparatus)</td>
<td>Diffuse-speckled nuclear or nucleolar*</td>
<td>ID, IPP, IB</td>
<td>SLE, PM/SSc overlap</td>
</tr>
<tr>
<td>PCNA/Ga/LE-4</td>
<td>PCNA (DNA scaffold)</td>
<td>Nuclear/nucleolar speckles*</td>
<td>ELISA, ID, IB, IPP</td>
<td>SLE</td>
</tr>
<tr>
<td><strong>Spliceosome Components</strong> (Splicing of pre-mRNA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sm</td>
<td>Sm core B’/B, D, E, F, and G</td>
<td>Speckled or negative†</td>
<td>ID, ELISA, IB, IPP</td>
<td>SS, SCLE, NLE, SLE, PBC, SSc, SSc, SLE, SLE</td>
</tr>
<tr>
<td>RNP, nRNP</td>
<td>U1 snRNP 70K, A, and C</td>
<td>Speckled</td>
<td>ID, ELISA, IB, IPP</td>
<td>SLE</td>
</tr>
<tr>
<td>LINR snRNP</td>
<td>U4/U6 snRNP</td>
<td>Nucleolar</td>
<td>IPP, IB</td>
<td>SLE, SSc, DM</td>
</tr>
<tr>
<td>RNA helicase A</td>
<td>RNA helicase A</td>
<td>Nuclear/nucleolar*</td>
<td>IPP</td>
<td>SLE</td>
</tr>
<tr>
<td>TIA-1, TIAR</td>
<td>RNA helicase A</td>
<td>Nucleolar</td>
<td>IPP, IB</td>
<td>SLE</td>
</tr>
<tr>
<td>Mi-2</td>
<td>NuRD complex (transcription regulation)</td>
<td>Spodges*</td>
<td>IPP</td>
<td>SS</td>
</tr>
<tr>
<td>p80-collin</td>
<td>Coiled bodies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA-I</td>
<td>Mitotic apparatus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other Ribonucleoproteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ro/SS-A</td>
<td>Ro (ribosomal RNA processing)</td>
<td>Speckled</td>
<td>ID, ELISA, IB, IPP</td>
<td>SS</td>
</tr>
<tr>
<td>La/SS-B/Ha</td>
<td>La (ribosomal RNA processing)</td>
<td>Speckled</td>
<td>ID, ELISA, IB, IPP</td>
<td>SLE</td>
</tr>
<tr>
<td>RNA helicase A</td>
<td>RNA helicase A</td>
<td>Spodges*</td>
<td>IPP</td>
<td>SLE</td>
</tr>
<tr>
<td>TIA-1, TIAR</td>
<td>RNA helicase A</td>
<td>Homogeneous</td>
<td>IPP</td>
<td>SLE</td>
</tr>
<tr>
<td>Mi-2</td>
<td>RNA helicase A</td>
<td>Spodges*</td>
<td>IPP</td>
<td>SLE</td>
</tr>
<tr>
<td>p80-collin</td>
<td>RNA helicase A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA-I</td>
<td>RNA helicase A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nucleolar</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA polymerases (RNAP)</td>
<td>(RNA transcription)</td>
<td>Punctate</td>
<td>IPP, IB</td>
<td>SSc</td>
</tr>
<tr>
<td>RNP I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNP II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNP III</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribosomal RNP</td>
<td>Ribosomal RNPs (protein translation)</td>
<td>Nucleolar, cytoplasmic</td>
<td>ID, IB, IPP, ELISA</td>
<td>SSc</td>
</tr>
<tr>
<td>Topoisomerase I (Scl-70)</td>
<td>Topoisomerase I (DNA gyrase)</td>
<td>Diffuse, grainy, nuclear or nucleolar*</td>
<td>ID, IB, IPP</td>
<td>SSc</td>
</tr>
<tr>
<td>Topoisomerase II U3 snoRNP (fibrillarin)</td>
<td>Topoisomerase II (DNA gyrase)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U3 snoRNP</td>
<td>U3 snoRNP (ribosomal RNA processing)</td>
<td>Diffuse with sparse nuclear</td>
<td>IPP</td>
<td>SSc</td>
</tr>
<tr>
<td>Th snoRNP</td>
<td>RNase MRP (mitochondrial RNA processing)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOR 90 (hUBF)</td>
<td>hUBF (ribosomal RNA transcription)</td>
<td>Diffuse or nuclear* 10-20 discrete spots</td>
<td>IPP, IB</td>
<td>SSc</td>
</tr>
<tr>
<td>PM-Scl (PM-1)</td>
<td>Exosome (RNA processing degradation)</td>
<td>Homogeneous nuclear or nucleolar</td>
<td>IPP, IB</td>
<td>PM, DM, SSc, overlap</td>
</tr>
<tr>
<td>Nucleobindin-2 (Wa)</td>
<td>Nucleobindin-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cytoplasmic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA Synthetases</td>
<td>(Translational machinery)</td>
<td>Diffuse</td>
<td>IPP, IB</td>
<td>PM, DM</td>
</tr>
<tr>
<td>J0-1</td>
<td>tRNA&lt;sup&gt;α&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL-7</td>
<td>tRNA&lt;sup&gt;β&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL-12</td>
<td>tRNA&lt;sup&gt;ε&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EJ</td>
<td>tRNA&lt;sup&gt;γ&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OJ</td>
<td>tRNA&lt;sup&gt;ε&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KS</td>
<td>tRNA&lt;sup&gt;ε&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mas</td>
<td>tRNA&lt;sup&gt;ε&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fodrin</td>
<td>α- and/or β-Fodrin (cytoskeletal component)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 55-2  Diagnostic Characteristics of Antinuclear Antibodies (ANAs)—cont’d

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Target Autoantigen (Function)</th>
<th>ANA Pattern(s)</th>
<th>Other Tests</th>
<th>Primary Rheumatic Disease Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal recognition particle</td>
<td>Signal recognition particle (transmembrane protein handling)</td>
<td>?</td>
<td>IPP, IB</td>
<td>PM</td>
</tr>
<tr>
<td>KJ</td>
<td>Translational apparatus</td>
<td>?</td>
<td>ID, IB</td>
<td>Myositis</td>
</tr>
<tr>
<td>Elongation factor 1α (Fer)</td>
<td>Elongation factor 1α (protein translation)</td>
<td>?</td>
<td>IPP</td>
<td>Myositis</td>
</tr>
<tr>
<td>CADM-140</td>
<td>140 kD</td>
<td>?</td>
<td>IPP, IB</td>
<td>Amyopathic DM</td>
</tr>
<tr>
<td>p140</td>
<td>Nuclear matrix protein NXP-2</td>
<td>?</td>
<td>IPP</td>
<td>Juvenile DM</td>
</tr>
<tr>
<td>SUMO-1</td>
<td>Small ubiquitin-like modifier 1 activating enzyme α- and β-subunits</td>
<td>?</td>
<td>IPP</td>
<td>DM</td>
</tr>
<tr>
<td>p155</td>
<td>?</td>
<td>?</td>
<td>IPP</td>
<td>DM and cancer-associated DM</td>
</tr>
</tbody>
</table>

*Cell cycle dependent.
†In cell studies, Ro RNP associates with cytoplasmic fractions.110
AAI, aminoacylation inhibition; CIE, counterimmunoelectrophoresis; CIF, Crithidia luciliae immunofluorescence; DIL, drug-induced lupus erythematosus; DM, dermatomyositis; dsDNA, double-stranded DNA; ELISA, enzyme-linked immunosorbent assay; Farr, Farr radioimmunoassay; hUBF, human upstream binding factor; IB, immunoblot; ID, immunodiffusion; IPP, immunoprecipitation; MCTD, mixed connective tissue disease; NLE, neonatal lupus erythematosus; NOR, nuclear organizer region; nRNP, nuclear ribonucleoprotein; NuRD, nucleosome remodeling–deacetylase; overlap, overlap syndromes; PBC, primary biliary cirrhosis; PCNA, proliferating cell nuclear antigen; PM, polymyositis; RA, rheumatoid arthritis; RIA, radioimmunoassay; RNAP, RNA polymerase; RNP, ribonucleoprotein; SCLE, subacute cutaneous lupus erythematosus; SLE, systemic lupus erythematosus; snoRNP, small nucleolar ribonucleoprotein; SS, Sjögren’s syndrome; SSC, systemic sclerosis; ssDNA, single-stranded DNA; TIA-1, T cell intracytoplasmic antigen 1; TIA-1–related protein; tRNA, transfer RNA; UCTD, undifferentiated connective tissue disease.

Immunofluorescence

The FANA provides a rapid yet highly sensitive screening method for ANA detection and remains the gold standard for clinical testing.21,22 Here, test sera at varying dilutions (typically serially increasing by twofold) are incubated with substrate cells, and bound antibodies are detected by fluorescein-conjugated anti–human immunoglobulin G (IgG), followed by visualization via a fluorescence microscope. Results typically are reported by two parameters—pattern and titer—with any pattern of reactivity at a titer of 1:40 or greater considered positive.23 The former includes one or more morphologic descriptors that typically reflect localization of the respective autoantigen(s) (see Table 55-2; Figures 55-1 and 55-2). Titer is generally reported as the last dilution at which an ANA pattern is detectable, but such an assessment has been considered somewhat imprecise and subjective, and interlaboratory standardization has not been widely instituted: Attempts to standardize the protocol have included computer-based fluorescent image quantification, subjective optical scales, and the use of standardized sera to define international units (IU/mL), although this varies by laboratory.

FANA results must always be interpreted in light of the particular substrate used by individual clinical laboratories. Many laboratories continue to use heterogeneous substrates such as rodent liver or kidney tissues; although such sections possess the advantage of eliminating interference from blood-group antibodies, heterophile antibodies, or passenger viruses, cultured cell lines such as HEp-2 cells have remained a gold standard substrate owing to their higher concentrations of nuclear and cytoplasmic antigens and standardization of use.21,24 To minimize the influence of other variables, such as quality of reagents, including

produce some false-positive results, and confirmation may require further testing. Therefore recognition of the local technique used for detection of ANAs is critically important for their clinical application in diagnosis and/or prognosis.

**Anti-DNA Antibody Tests**

Anti-DNA antibodies warrant special consideration owing to their wide range of autoantigenic epitopes and their assay difficulties. Antibodies that recognize denatured single-stranded (ss)DNA, which are less specific for rheumatic disease, bind free purine and pyrimidine base sequences; SLE-specific antibodies that recognize native, double-stranded (ds)DNA bind the deoxyribose phosphate backbone or the rarer, conformation-dependent left-handed helical Z-form. Two methods to ensure the use of native dsDNA in anti-DNA tests include digestion with S1 nuclease, which removes overhanging ssDNA ends, and chromatography on a hydroxyapatite column, which separates single-stranded segments from dsDNA. Unfortunately, despite such efforts, native DNA may spontaneously denature, especially when bound to plastic ELISA plates; this may account for the results of several reports that revealed a relative lack of specificity of anti-dsDNA antibodies for SLE. Therefore reliable assays must ensure the integrity of dsDNA.

The Farr radioimmunoassay remains the gold standard for DNA antibody testing; it involves the binding of autoantibodies to radiolabeled dsDNA in solution. Precipitation of antibody-DNA complexes by ammonium sulfate allows quantification of the percentage of incorporated (antibody-bound) radioactive dsDNA. Normal sera typically bind a small fraction of added DNA (usually less than 20%), whereas SLE sera often bind nearly 100% of added DNA. The specificity of this assay, however, still depends on the quality of dsDNA and the removal of contaminating ssDNA. Also, because of the involvement of radioactivity, this assay is not routinely used in clinical laboratories.

In contrast, the *Crithidia* test provides an inherently reliable dsDNA substrate that is generally clinically available. In this assay, the hemoflagellate *C. luciliae* serves as a substrate for indirect immunofluorescence. Its kinetoplast, a
modified giant mitochondrion, contains a concentrated focus of stable, circularized dsDNA, without contaminating RNA or nuclear proteins, providing a sensitive and specific immunofluorescence substrate by which to establish anti-dsDNA activity. Thus together, ELISAs, *C. luciliae* immunofluorescence, and possibly Farr radioimmunonassay tests provide effective, complementary mechanisms by which anti-ssDNA and anti-dsDNA can be distinguished.

**Other Assays**

Several additional assays for the determination of ANA specificity have been employed in clinical and/or basic science studies. Such techniques include immunodiffusion and counterimmunoelectrophoresis techniques, two relatively insensitive assays used in many clinical studies associating ANA specificities (especially ENAs) with disease manifestations and outcome; immunoprecipitation and immunoblot, two sensitive and specific assays predominantly confined to research settings; and enzyme inhibition assays (e.g., inhibition of topoisomerase I by anti-Scl-70, inhibition of RNA splicing by anti-snRNP), which include highly specialized techniques to characterize ANAs functionally. Such assays have not achieved widespread use in diagnostic laboratories because of their cumbersome and/or highly specialized natures.

**INTERPRETATION OF THE FLUORESCENT ANTINUCLEAR ANTIBODY (FANA) TEST**

**KEY POINT**

Although the FANA pattern and titer may provide some insight into the specific autoantigen(s) targeted, as well as the potential likelihood of connective tissue disease, such correlations should only guide, not absolutely determine, clinical decisions.

**Pattern**

Patterns of staining by FANA are often reported as homogeneous, speckled, or rim/peripheral when nuclear staining is present, but they may also be reported as cytoplasmic, centromere, or nucleolar, often reflecting intracellular localization of the target antigen(s) (see Table 55-2 and Figures 55-1 and 55-3). The presence of unusual patterns may be particularly helpful in appropriate clinical settings, such as the presence of a centromere pattern in a patient with features of systemic sclerosis, suggesting anticentromere antibodies, or a cytoplasmic pattern in a patient with features of myositis, suggesting anti-transfer RNA (tRNA)

---

**Figure 55-3** Algorithm for the use of antinuclear antibodies in the diagnosis of connective tissue disorders. See text for details. ANA, antinuclear antibody; FANA, fluorescent antinuclear antibody; MCTD, mixed connective tissue disease; SLE, systemic lupus erythematosus.
synthetic antibodies. However, the presence or absence of patterns is not always highly accurate in predicting specificity, and non-nuclear patterns may not be reported at all by some laboratories, including rare patterns such as nuclear dot, Golgi, or antimitochondrial antibodies. Furthermore, the role of the FANA pattern in predicting target autoantigen specificities has been largely supplanted by widely clinically available autoantigen-specific ELISAs. As a result, the presence of any such patterns serves as evidence in an appropriate clinical setting of non–organ-specific autoimmunity, which may warrant further evaluation; however, speculation regarding the significance of a specific pattern may be worthwhile in only a select number of cases.

**Titer**

Although the widely accepted cutoff for FANA positivity has remained 1:40, greater clinical significance generally has been thought to correlate with higher titers. Normal individuals, usually older and female persons, and relatives of persons with connective tissue diseases produce positive FANAs at a frequency sometimes exceeding 30% (see Table 55-1). Although these patients often possess titers of less than 1:320 with homogeneous staining patterns, many subjects possess higher titers yet remain clinically asymptomatic for years, and occasional SLE patients may demonstrate negative FANAs; this is perhaps a more frequent observation if they possess isolated anti-Ro or anti ssDNA antibodies and/or if the laboratory uses rat or mouse tissues. As a result, high- versus low-titer FANA results may not be of sufficient clinical significance to warrant differential subsequent evaluation. Rather, a positive screening FANA, of any titer, requires clinical correlation.

### DISEASES ASSOCIATED WITH ANTINUCLEAR ANTIBODIES

**Systemic Lupus Erythematosus**

ANAs remain a hallmark of SLE; although past studies have reported FANA frequencies as low as 90%, the test is positive in more than 99% of patients with the use of current methods. SLE often evokes autoantibodies against a seemingly endless range of antigens in many cellular locations, but most SLE autoantigens reside in the nucleus and may be broadly categorized into chromatin-associated versus ribonucleoprotein antigens (Table 55-3; also see Table 55-2), which may facilitate disease subclassification and prognosis.

#### Table 55-3 Antinuclear Antibodies in Systemic Lupus Erythematosus*

<table>
<thead>
<tr>
<th>Antibody Specificity</th>
<th>Prevalence (%)</th>
<th>SLE Specific?</th>
<th>Major Disease Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chromatin-Associated Antigens</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromatin</td>
<td>80-90</td>
<td>In high titer</td>
<td>Renal LE, overall disease activity</td>
</tr>
<tr>
<td>dsDNA</td>
<td>70-80</td>
<td>In high titer</td>
<td>Drug-induced lupus, anti-DNA</td>
</tr>
<tr>
<td>Histon</td>
<td>50-70</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Ku</td>
<td>20-40</td>
<td>No</td>
<td>Overlap</td>
</tr>
<tr>
<td>RNA polymerase II</td>
<td>9-14</td>
<td>No</td>
<td>Relatively (SLE and overlap)</td>
</tr>
<tr>
<td>Kinetochore</td>
<td>6</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>PCNA</td>
<td>3-6</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td><strong>Ribonucleoprotein Component</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>snRNPs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sm core</td>
<td>20-30</td>
<td>Yes</td>
<td>Cutaneous LE</td>
</tr>
<tr>
<td>U1 snRNP</td>
<td>30-40</td>
<td>No</td>
<td>Neonatal LE and CHB</td>
</tr>
<tr>
<td>U2 snRNP</td>
<td>15</td>
<td>No</td>
<td>Neonatal LE</td>
</tr>
<tr>
<td>U5 snRNP</td>
<td>?</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>U7 snRNP</td>
<td>?</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Ro/SS-A</td>
<td>40</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>La/SS-B</td>
<td>10-15</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Ribosomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P0, P1, P2 protein</td>
<td>10-20</td>
<td>Yes</td>
<td>Neuropsychiatric LE</td>
</tr>
<tr>
<td>28S rRNA</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S10 protein</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L5 protein</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L12 protein</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR proteins</td>
<td>50-52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteosome</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNF TRs</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA helicase A</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ki-67</td>
<td>?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Shown are major antinuclear antibody specificities described in SLE, along with estimated prevalences and disease associations (bold indicates data supported by multiple studies). See text for details.

CHB, congenital heart block; dsDNA, double-stranded DNA; LE, lupus erythematosus; PCNA, proliferating cell nuclear antigen; rRNA, ribosomal RNA; SLE, systemic lupus erythematosus; smRNP, small nuclear ribonucleoprotein; TNF TRs, TNF translational regulators, including T cell intracytoplasmic antigen 1 (TIA-1) and TIA-1–related protein (TIAR).
Chromatin-Associated Antigens

Anti-DNA. Although antibodies against DNA remain one of the most widely recognized specificities in SLE, antibodies against its more physiologic forms, such as nucleosomes or chromatin, are more prevalent in and probably relevant to pathogenesis. Nonetheless, most of the clinical literature remains linked to classic anti-dsDNA antibodies (see “Anti-DNA Antibody Tests,” earlier): Many diseases exhibit anti-ssDNA activity, but only SLE sera characteristically possess high-titer anti–ds- and/or anti–Z-DNA activity, as characterized by positive Farr or Crichtidia assays, seen in approximately 73% of patients, in contrast to low titers seen often in SS, RA, other disorders, and normal individuals. In SLE, anti-DNA antibodies strongly correlate with nephritis and disease activity, in contrast to other ANA specificities. In some settings, drug-induced anti-DNA antibodies are observed, as during therapy with some tumor necrosis factor (TNF) inhibitors, although they do not necessarily correlate with clinical manifestations of connective tissue disease.

Some anti-DNA antibodies may cross-react with other autoantigens, explaining correlation with other end-organ manifestations, such as the neuronal N-methyl-D-aspartate (NMDA) receptor or ribosomal P antigens for central nervous system disease. Such findings suggest that the immunologically relevant antigen for anti-DNA antibodies in fact may not be DNA. As a result, the presence of anti-DNA activity should always prompt consideration of renal disease, but the presence of anti-DNA activities does not always indicate lupus nephritis, and vice versa.

Antihistone (Nucleosome). Antihistone antibodies target the protein components of nucleosomes, the DNA-protein complexes that form the substructure of transcriptionally inactive chromatin. They are common in SLE, associate with anti-dsDNA, and are particularly characteristic of and sensitive for drug-induced lupus, where they associate with anti-ssDNA. However, they are commonly seen in other rheumatic diseases, including myositis and SSC, as well as in chronic infections, such as Epstein-Barr virus, so that clinical correlations for antihistone antibodies have not been consistent.

Other Chromatin-Associated Autoantigens. Other chromatin-associated autoantigens in SLE include several specificities also observed in other rheumatic and nonrheumatic diseases with still somewhat undefined clinical significance. For instance, autoantibodies against Ku, the catalytic subunit of the DNA-dependent protein kinase implicated in DNA repair and the V(D)J recombination, have been associated with RNA polymerase II antibodies, and one study has associated them with Raynaud’s phenomenon, arthralgia, skin thickening, and esophageal reflux; however, other studies suggest that Ku subunit specificity may be more relevant, with anti-p70 antibodies correlating with features of SSC-polymyositis overlap, and anti-p80 antibodies with features of SSC or SLE. Other chromatin-associated autoantigens include proliferating cell nuclear antigen (PCNA), which participates in a scaffold to facilitate DNA replication, recombination, and repair; and RNA polymerase II, which transcribes some small nuclear RNA genes and all protein-encoding genes, both of which remain of uncertain clinical correlation.

Ribonucleoproteins

Anti–Small Nuclear Ribonucleoproteins. In SLE, the best-described snRNP autoantibodies include the Sm and U1 RNP specificities, which target the RNAs or proteins of the spliceosome, a complex of RNP particles involved in premessenger RNA splicing. These particles include the U1, U2, U4/U6, U5, U7, U11, and U12 snRNPs, each of which consists of its respective uridine-rich (thus U) small nuclear RNA (snRNA) and a set of polypeptides, including a common core of “Sm” polypeptides (B/B', D1, D2, D3, E, F, and G), as well as particle-specific polypeptides. Anti-Sm antibodies, which target proteins of the Sm core, B/B', and one of the D polypeptides, as well as the Sm-like LSm4, appear in only 20% to 30% of SLE patients, but are considered specific for the diagnosis; however, their presence has only inconsistently been associated with specific disease activity and/or prognosis. In contrast, anti-U1 snRNP (nRNP, nuclear RNP, or U1 RNP) autoantibodies, which target the 70 K, A, or C polypeptides specific to U1 snRNP, occur in 30% to 40% of SLE patients but are not specific for SLE and likewise have been only variably associated with disease activity, myositis, esophageal hypomotility, Raynaud’s phenomenon, leukopenia, lack of nephritis, arthralgias or arthritis, sclerodactyly, interstitial changes on chest radiographs, and central nervous system manifestations. Several other snRNP antibodies have been described in SLE, often in overlap syndromes (e.g., U2-, U5-, or U7-snRNP–specific), although their clinical importance remains uncertain.

Anti-Ro/SSA and La/SSB. These two ribonucleoprotein particles, which are part of a macromolecular complex that predominantly processes RNA polymerase III transcripts, have been often associated with SS and the neonatal lupus syndrome, as well as with ANA-negative SLE (especially anti-Ro; see “Interpretation of the FANA,” earlier). Some but not all studies have indicated that anti-Ro may segregate among rheumatic diseases based on subunit specificities, with Ro52 without Ro60 specificities correlating with SS, while Ro60, perhaps specifically a Ro60 apoptope, with or without Ro52 specificities correlating with other connective tissue diseases (CTDs), including SLE. In SLE, anti-Ro associates with several manifestations, especially skin disease (cutaneous lupus, chilblains, photosensitivity), sicca symptoms, and the neonatal lupus syndrome, including congenital heart block, anti-La, rheumatoid factor, pulmonary disease, complement (especially C4) deficiencies, thrombocytopenia, lymphopenia, and cardiac fibroelastosis; in comparison, anti-La correlates with late-onset SLE, secondary SS, the neonatal lupus syndrome, and protection from anti-Ro–associated nephritis.

Antiribosomes

The best-studied antiribosome antibodies in SLE, antiribosomal P protein (anti-P), target the P0, P1, and P2 proteins of the large 60S ribosome subunit. Although they occur in only a minority of patients, they are considered highly specific for SLE and particularly specific for neuropsychiatric lupus, classically psychosis; correlations with active disease, renal disease, liver and hematologic disease, alopecia, anti-Sm, anti-DNA, and anticardiolipin antibodies have also
been reported. Other, less prevalent antiribosomal antibodies target ribosomal RNA (rRNA), such as the 28S rRNA, or other ribosomal proteins, such as the S10, L5, and L12 subunit proteins, although their clinical significance remains unclear.

Other Antinuclear Antibodies in Systemic Lupus Erythematosus

Many other SLE ANA specificities have been described; some are apparently quite prevalent, such as SR splicing factors, proteasome, the TNF translational regulator, and the RNA helicase A specificities. Many such specificities continue to lack clear clinical context, although preliminary analyses indicate some correlation, such as Ki-67 with sicca, or RNA with overlap syndromes. Others remain of interest because of their connection with other diseases, such as perinuclear-antineutrophil cytoplasmic antibodies, topoisomerase I, and kinetochore specificities.

Systemic Sclerosis (Scleroderma)

Antinuclear antibodies against nucleolar antigens characterize the autoantibody response in SSc. Positive FANAs, sometimes speckled in appearance, appear in as much as 97% of sera, although percentages vary depending on the substrate used for detection. Unlike SLE sera, however, systemic sclerosis sera typically contain monospecific autoantibody specificities, targeting such structures as the kinetochore, topoisomerase I, or RNA polymerases (Table 55-4; also see Table 55-2).†

Antikinetochore (Centromere) and Anti–Topoisomerase I

These specificities constitute major diagnostic tools in the subclassification of SSc. Originally named anticentromere, antikinetochore targets at least four centromere (kinetochore) antigens (CENPs) of the mitotic spindle apparatus that promotes chromosome separation during mitosis: CENP-B (the predominant kinetochore autoantigen), -A, -C, and -D. As such, these specificities require mitotically active cells for robust detection, accounting for some ANA-negative SSc findings (see “Interpretation of the FANA,” earlier). Their clinical significance has been extensively studied and heavily associated with Raynaud’s phenomenon and CREST, in which up to 98% of patients have antikinetochore antibodies. Other associations include limited skin involvement, mat-like telangiectasias, pulmonary or vascular disease, increased malignancy risk, and possibly reduced type I IFN responses. In contrast, anti–topoisomerase I (Scl-70) autoantibodies, which predominantly target the catalytic region of DNA helicase topoisomerase I, generally predict diffuse cutaneous disease with proximal skin involvement, pulmonary fibrosis, longer disease duration, association with cancer, or both digital pitting scars and cardiac involvement, as well as more rapid disease activity. However, approximately 40% of all SSc patients lack either antibody, and a minority (≤1%) possess both antibodies, such that although these antibodies are clinically useful for disease classification and prognosis, they may not be used for definitive diagnosis.

Anti-RNA Polymerases

Anti-RNA polymerase (RNAP) antibodies, which target the eukaryotic RNA polymerases (see section on SLE: anti-RNA polymerase), associate with diffuse cutaneous involvement. Although anti–RNAP II antibodies appear in other diseases, such as SLE or overlap syndrome, and may be associated with other autoantibody specificities against Ku or ribonucleoproteins, anti–RNAP I and III antibodies appear specific for SSc, in which they may be useful for the prediction of renal crisis. RNAP III antibodies in particular may predict diffuse cutaneous SSc, including higher skin score, tendon friction rubs, and renal crisis.

Table 55-4  Antinuclear Antibodies in Systemic Sclerosis*

<table>
<thead>
<tr>
<th>Antibody Specificity</th>
<th>Prevalence (%)</th>
<th>Specificity for SSc</th>
<th>Mutually Exclusive?†</th>
<th>Major Disease Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinetochore (centromere)</td>
<td>22-36</td>
<td>Relatively</td>
<td>Yes</td>
<td>CREST</td>
</tr>
<tr>
<td>Topoisomerase I</td>
<td>22-40</td>
<td>Relatively</td>
<td>Yes</td>
<td>Diffuse cutaneous disease</td>
</tr>
<tr>
<td>Topoisomerase II</td>
<td>22</td>
<td>Relatively</td>
<td>Yes</td>
<td>Pulmonary fibrosis</td>
</tr>
<tr>
<td>RNA polymerases</td>
<td>4-23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA polymerase I</td>
<td></td>
<td>Relatively</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>RNA polymerase II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA polymerase III</td>
<td></td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNF TRs</td>
<td>42</td>
<td>Relatively</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>B23 nuclear phosphoprotein</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U3 snORNP (fibrillarin)</td>
<td>6-8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Th snORNP (RNase MRP, 7-2 RNA)</td>
<td>4-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U11/U12 snORNP</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM-Scl</td>
<td>2-5</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sp1</td>
<td>7</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOR 90 (HUBF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Shown are major antinuclear antibody specificities described in SSc, along with estimated prevalences and disease associations (bold indicates data supported by multiple studies). Antinuclear antibody specificities whose incidences are thought to be “mutually exclusive” of each other in SSc are indicated. See text for details.

†Antibody specificity is often present exclusive of other SSc-related antibodies.

HUBF, human upstream binding factor; NOR, nucleolar organizer region; snORNP, small nucleolar ribonucleoprotein; snRNP, small nuclear ribonucleoprotein; SSc, systemic sclerosis; TNF TRs, TNF translational regulators, including T cell intracytoplasmic antigen 1 (TIA-1) and TIA-1-related protein (TIAR).
Anti–Polymyositis-Sclerodermat (PM-Scl)

Anti–PM-Scl antibodies target the PM-Scl-75 and PM-Scl-100 components of the exosome, an exoribonuclease complex that regulates ribosomal RNA. Responses against PM-Scl-75 alone appear more common among patients with diffuse SSC, and overlap syndromes are typically associated with responses against both components. Its presence associates with myositis-SSc overlap without SLE features: 50% of anti–PM-Scl antibody–positive patients have the overlap, and 25% of overlap patients have the antibody. Also, anti–PM-Scl appears associated with arthritis and skin lesions of dermatomyositis, calcinosis, mechanic’s hands, and eczema, and it appears to increase the incidence of muscle, tendon, and renal disease.

Other Systemic Sclerosis–Related Antinuclear Antibodies

Several other specificities have been described in SSC with possible prognostic implications, including antifibrillarlin, which targets a component of the U3 small nuclear ribonucleoprotein and may associate with diffuse disease, including internal organ or skeletal muscle involvement, or pulmonary hypertension; anti–topoisomerase II, which appears to associate with pulmonary hypertension, as well as localized scleroderma; anti-Th (Th snRNP, mitochondrial RNA processing RNase MRP), which may predict pulmonary hypertension, limited cutaneous disease, puffy fingers, small-bowel involvement, hypothyroidism, pericarditis or interstitial lung disease, or reduced arthritis or arthralgias; antibodies against the nuclear phosphoprotein B23, which appear to associate with pulmonary hypertension and antifibrillarlin antibodies; antibodies against the RNAP II transcription activator Sp1, which seem to correlate with Raynaud’s phenomenon and other signs of undifferentiated connective tissue disease; TNF translational regulator specificities, which may associate with lung involvement; anti-U11/U12 RNP antibodies, which appear to correlate with pulmonary fibrosis; and anti–nucleolar organizer region (NOR) 90 (human upstream binding factor). Other ANAs described in SSC include several specificities characteristically observed in other CTDs, such as histone, Ku, Ro, tRNA, snRNP, and ANCA, although their clinical relevance in SSC remains unclear.

Inflammatory Muscle Diseases

Inflammatory muscle diseases make up a diverse group of illnesses often characterized by autoantibody responses against cytoplasmic antigens. Although between 40% and 80% of PM/DM patients have positive ANA, as many as 90% of patients with all types of inflammatory muscle diseases have autoantibodies to cellular antigens. Myositis autoantibodies are generally categorized into myositis-specific autoantibodies (MSAs), which are considered exclusive to inflammatory myositis, and those associated with overlap syndromes that include myositis (Table 55-5; also see Table 55-2).

Myositis-Specific Autoantibodies

The best-characterized MSAs include the antisynthetases, which target different aminocyl-tRNA synthetases. Some of these antibodies target the tRNA anticodon loop, enabling them to inhibit enzymatic activity. The individual prevalences of these antibodies vary, but their clinical associations remain similar. Although PM appears more common with anti-Jo-1, and DM is more common with the other synthetases, these specificities together correlate with the “antisynthetase syndrome,” which includes interstitial lung disease, arthritis, Raynaud’s phenomenon, mechanic’s hands, hyperkeratotic lines, sclerodactyly, facial telangiectasia, calcinosis, and sicca, generally with a relatively poor prognosis. Recent studies have suggested that anti-Jo-1 antibody levels may correlate with disease activity, as well as with the IFN-γ-inducible chemokines CXCL9 and CXCL10. Nonetheless, other disease associations have been reported, distinct from the antisynthetase syndrome. For example, one study has associated anti–threonyl-tRNA synthetase antibodies with fetal loss and severe relapsing myositis; a relatively rare activity, anti-KS, has been described against asparaginyl-tRNA synthetase in a few patients with interstitial lung disease and inflammatory arthritis or undifferentiated connective tissue disease; and PL-7 antibodies may correlate with milder muscle disease.

Myositis Overlap Autoantibodies

The best-described myositis overlap–associated ANAs include snRNP and PM-Scl specificities. In myositis, anti-snRNP antibodies typically target U1 snRNP, although a few anti-Sm and anti–U2 snRNP specificities have been described (see section on SLE: anti-snRNP). The former tend to associate with features of MCTD, including SLE–myositis overlap, myositis-SSc overlap, and undifferentiated features (Raynaud’s phenomenon, puffy fingers, arthritis) later progressing to myositis, possibly responding to corticosteroids; anti-U2 snRNP associates with myositis and sclerodactyly, sometimes with SLE and usually without interstitial lung disease. Anti–PM-Scl antibodies associate with myositis-SSc overlap without SLE features (50% of
Table 55-5  Antinuclear Antibodies in Inflammatory Muscle Diseases*

<table>
<thead>
<tr>
<th>Antibody Specificity</th>
<th>Prevalence (%)</th>
<th>Disease Specificity</th>
<th>Major Disease Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-tRNA synthetases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histidyl (Jo-1)</td>
<td>20-30</td>
<td>Myositis</td>
<td>Antisynthetase syndrome</td>
</tr>
<tr>
<td>Threonyl (PL-7)</td>
<td>1-5</td>
<td>Myositis</td>
<td></td>
</tr>
<tr>
<td>Alanyl (PL-12)</td>
<td>1-5</td>
<td>Myositis</td>
<td></td>
</tr>
<tr>
<td>Glycyl (EJ)</td>
<td>1-5</td>
<td>Myositis</td>
<td></td>
</tr>
<tr>
<td>Isoleucyl (OJ)</td>
<td>1-5</td>
<td>Myositis</td>
<td></td>
</tr>
<tr>
<td>Asparaginyl (KS)</td>
<td>?</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>Selenocysteinyll (Mas)</td>
<td>1-2</td>
<td>Myositis*</td>
<td>Dermatologic involvement</td>
</tr>
<tr>
<td>Mi-2</td>
<td>8 (15%-20% of DM)</td>
<td>Myositis*</td>
<td></td>
</tr>
<tr>
<td>Signal recognition particle</td>
<td>4</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>KJ</td>
<td>&lt;1</td>
<td>Myositis*</td>
<td></td>
</tr>
<tr>
<td>Proteasome</td>
<td>62</td>
<td>No</td>
<td>MCTD features</td>
</tr>
<tr>
<td>Histone</td>
<td>17</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>RNPs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1 snRNP</td>
<td>12</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>U2 snRNP</td>
<td>3</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>Ro</td>
<td>10</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>La</td>
<td>?</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>PM-5cd</td>
<td>8</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>Elongation factor 1α (Fer)</td>
<td>1</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Histone</td>
<td>?</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Ku</td>
<td>?</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>U3 snoRNP</td>
<td>?</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>CADM-140</td>
<td>?</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>MDAS</td>
<td>?</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>p140</td>
<td>?</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>p155</td>
<td>?</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>SUMO-1</td>
<td>?</td>
<td>Overlap</td>
<td></td>
</tr>
<tr>
<td>TIF1-γ</td>
<td>?</td>
<td>Overlap</td>
<td></td>
</tr>
</tbody>
</table>

*Shown are major antinuclear antibody specificities described in inflammatory myositis, along with estimated prevalences and disease associations (bold indicates data supported by multiple studies). See text for details.
†Considered a myositis-specific autoantibody (MSA) despite recent findings in autoimmune hepatitis.
‡Often referred to as MSAs.

DM, dermatomyositis; MDAS, melanoma differentiation–associated gene 5; RNP, ribonucleoprotein; snoRNP, small nucleolar ribonucleoprotein; snRNP, small nuclear ribonucleoprotein; SUMO-1, small ubiquitin-like modifier activating enzyme subunits A and B; TIF1-γ, transcriptional intermediary factor 1-γ; tRNA, transfer RNA.

anti–PM-Scl antibody–positive patients have overlap, and 25% of overlap patients have the antibody), and also with arthritis, DM skin lesions, calcinosis, mechanic’s hands, and eczema.3,91 Other antibodies associated with myositis in overlap syndromes include several specificities found in other diseases, such as Ku, which is found more commonly in SLE and SSc; and U3 small nucleolar RNP (fibrillarin), which is associated with myositis in SSc, especially of the diffuse type. Several other specificities have been described in inflammatory myositis, although their clinical significance remains largely undefined (see Table 55-5).26

**Sjögren’s Syndrome**

In SS, reported incidences of positive FANAs range widely, reflecting differences in study populations and disease criteria, and depend heavily on the inclusion or exclusion of secondary, CTD-related disease, which increases the likelihood and amplitude of positive tests.24 Thus, although as little as 40% ANA positivity has been reported, many studies report frequencies of 90% to 96%49 with a diverse range of autoantibodies, including both ubiquitous (Table 55-6; see also Table 55-2) and tissue-specific reactivities such as antithyroid, gastric parietal cell, and muscarinic receptors. Such issues hinder interpretation of disease associations and disease specificities of ANAs in SS.

Of the SS ANA specificities, the best-characterized remain Ro (SSA) and La (SSB)—two nuclear RNPs involved in RNA metabolism (see section on SLE: anti-Ro/SSA and anti-La/SSB)—as well as fodrin (nonerythroid spectrin), a cytoskeletal heterodimer composed of α- and β-subunits structurally and functionally similar to erythroid spectrin. Anti-Ro antibodies appear in approximately 40% to 95% of SS patients, are associated with various extraglandular manifestations, including serologic association with anti-La and rheumatoid factor, and may result from genetically linked alternative mRNA processing of Ro.95 Similarly, anti-La antibodies appear in as many as 87% of SS patients and are also associated with extraglandular manifestations and serologic association with anti-Ro and rheumatoid factor. Antibodies against α-fodrin have been detected in 64% to 67% of patients,96 but they are relatively uncommon in other CTDs such as SLE. Preliminary analyses suggest some extraglandular and serologic associations. In contrast, antibodies against β-fodrin have been described in as many as 70% of patients, but clinical associations have not been reported.97 Several other specificities have been reported in SS in a significant proportion (>3%) of patients, including antibodies against MA-I, a 200-kD protein localized to the mitotic apparatus in dividing cells (which may be identical to NuMA); p80-Coli, an 80-kD protein associated with nuclear coiled bodies (although this antibody...
might not be CTD specific); and specificities characteristically found in other rheumatic diseases such as kinetochore and perinuclear antineutrophil cytoplasmic antibodies, although their clinical importance remains uncertain (see Table 55-6). 96,99

**Mixed Connective Tissue Disease and Overlap Syndromes**

Although overlap syndromes of connective tissue disease remain matters of nosologic debate, virtually all investigators concur that ANAs are universal in these conditions. 102 Indeed, since the initial formal description of MCTD in 1972, the presence of anti-U1 snRNP antibodies has been consistently required for classification and/or diagnosis, with associated ANA titers typically exceeding 1:1000 and often 1:10,000. 101 However, several investigators have noted that many such patients develop manifestations, both clinical and serologic, that allow the diagnosis of a defined connective tissue disease, such as SLE, RA, SSc, or PM/DM, and therefore the accuracy of specific clinical associations of specific ANAs in “overlap” settings is hindered by issues of disease classification. In such instances, it seems most reasonable to base the importance of individual autoantibody specificities upon their primary disease association, despite the lack of definitive, well-coded evidence (e.g., anti–topoisomerase I as predictive of eventual diffuse SSc-like skin disease or pulmonary fibrosis, or anti-dsDNA for lupus-like glomerulonephritis).

**Other Conditions**

In contrast to traditional ANA diseases, the presence of a positive ANA in other diseases remains largely unhelpful for diagnosis, although in Raynaud’s phenomenon (RP), juvenile rheumatoid arthritis (JRA), and antiphospholipid antibody syndrome (APS), it may predict the development or the presence of underlying SLE. 106 Other conditions in which ANAs have been found include other rheumatic diseases such as the vasculitides or sarcoidosis; autoimmune diseases such as multiple sclerosis or inflammatory bowel disease; and an ever-growing list of additional conditions such as dermatologic, infectious, psychiatric, neurologic, and cardiovascular diseases. In general, studies describing such associations lack comprehensive analyses regarding the significance of ANAs, such that use of the ANA outside of the rheumatic diagnoses indicated here (see Table 55-1) remains largely exploratory.

### CONCLUSION: CLINICAL UTILITY OF ANTINUCLEAR ANTIBODY TESTING

The ANAs encompass an ever-widening range of nuclear, nucleolar, and cytoplasmic autoantigen specificities. Within the ANA diseases—including SLE, SSc, PM/DM, SS, and MCTD—many autoantibodies possess unique rheumatologic associations, but the specificity of these associations has diminished as the sensitivity of clinical studies has increased the detection of these specificities in both rheumatic and nonrheumatic diseases. Consequently, tests for ANAs can greatly facilitate the clinical evaluation of patients in the context of their particular disease associations, but these studies retain only an adjunct role in rheumatologic diagnosis.

Figure 55-3 describes an algorithm for the rheumatologic evaluation of a patient for ANAs. In most clinical laboratories, the FANA serves as a screening test, and a positive result often prompts “cascade” testing by the laboratory, the ordering physician, or both with ELISAs or other assays for specific autoantibody specificities, such as anti-dsDNA, anti-Ro/SSA, anti-La/SSB, anti-RNP, and anti-Sm. 23,24 A negative or low-titer FANA in the setting of low clinical suspicion of rheumatic disease usually indicates the absence of significant ANAs and argues against diagnosis of one of the ANA diseases (see Table 55-1); however, if the clinical picture strongly suggests connective tissue disease,
Further investigation may involve specific assays for antigens that are often FANA negative, such as Ro, Jo-1, or phospholipids. On the other hand, because some specific ANAs possess diagnostic significance, positive FANA results usually warrant follow-up with specialized assays—but only in the setting of strong clinical suspicion, because the positive predictive value of an ANA in the absence of other clinical signs of CTD is low, in part because ANAs may precede clinical disease by many years, and in part because of the relatively high incidence of ANA in normal individuals. Thus, if SLE features are present, further work may focus on anti-DNA, anti-Sm, anti-U1 snRNP, and anti-Ro antibodies. Similarly, if MCTD, SS, SSc, or polymyositis is suspected, the serum may be tested, respectively, for anti-U1 snRNP; anti-Ro or anti-La; anti-topoisomerase I, anticentromere, or antinucleoli; or anti-tRNA synthetases. If these tests are negative in the setting of high clinical suspicion, repeat testing at a later date may be warranted, because titers of such autoantibodies can fluctuate over time, irrespective of the disease course. Positive results with these more specialized assays do not alone signify specific diseases, but rather add weight to diagnoses that should throughout the evaluation rely heavily on other clinical information. Indeed, many clinical and research studies upon which autoantibody–disease associations have been developed often utilized highly refined detection methods, such as immunoprecipitation and immunoblot, which often are unavailable in routine clinical laboratory testing. Thus, the relevance of many antibody test results in specific clinical settings continues to require careful, individualized interpretation by the referring physician.

References


The references for this chapter can also be found on www.expertconsult.com.
Autoantibodies in Rheumatoid Arthritis

FELIPE ANDRADE • ERIKA DARRAH • ANTONY ROSEN

Autoantibodies have proven to be useful tools for diagnosis and prediction in the autoimmune rheumatic diseases. Emerging data in several autoimmune rheumatic diseases have demonstrated that clinical evolution of disease from preclinical phase to overt clinical disease is marked by a change in the specificity of the immune response, with autoantibodies directed against distinct antigenic targets at different disease phases. Although many other autoimmune rheumatic diseases have traditionally been marked by highly phenotype-specific autoantibody responses (e.g., anti-dsDNA antibodies in systemic lupus erythematosus [SLE], high-titer anti-topoisomerase-1 antibodies in diffuse scleroderma), the discovery of highly specific autoantibody markers of rheumatoid arthritis (RA) lagged significantly behind these other diseases. During the past decade, there has been enormous progress in this area, largely fueled by the discovery that citrullinated proteins are specific targets of autoantibodies in RA. This chapter reviews the autoantibodies in RA, with emphasis on both rheumatoid factors (RFs) and anticitrullinated protein autoantibodies (ACPAs). The implications of these specificities for diagnosis and prediction of outcome, as well as for understanding pathogenic events in the disease, are highlighted.

RHEUMATOID FACTOR

The earliest assays (later called the Rose-Waaler agglutination test), which suggested the existence of autoantibodies in RA, were developed in the early 1940s. At this time serum from RA patients was shown to cause agglutination of sheep blood cells, which had been sensitized with subagglutinating doses of rabbit anti-sheep erythrocyte antibodies. These assays were later shown to detect immunoglobulin M (IgM) antibodies from RA patients that recognized the Fc portion of IgG. Numerous modifications of the agglutination assay evolved, particularly the use of IgG-coated latex beads instead of sheep erythrocytes. Subsequent developments established RF assays in radioimmunoassay, enzyme-linked immunosorbent assay (ELISA), and nephelometry formats, with increased convenience but similar performance in terms of sensitivity and specificity. RFs may be positive in healthy controls (1% in younger individuals, up to 5% in individuals older than 70 years) and in patients with numerous other non-RA diseases (including other rheumatic diseases such as Sjögren’s syndrome and cryoglobulinemia), as well as chronic infections (see Table 56-1). The pretest probability of diagnosing RA therefore greatly influences the performance of the RF test, with sensitivities and specificities both in the 50% to 90% range, depending on the patient groups studied. The existence of IgG and IgA RFs and evidence of somatic hypermutation have suggested that some RFs in RA are T cell dependent. Although there have been studies suggesting that IgG and IgA RFs increase the specificity of RFs for RA, a recent meta-analysis showed that assays of the different RF isotypes performed similarly in terms of sensitivity and specificity and therefore do not add much over standard RF assays.

Importantly, patients with RA also vary in terms of timing of RF appearance, with RFs preceding symptomatic disease in a significant subpopulation and following disease onset with variable kinetics in other patients. Indeed, the earlier onset of RF in patients with RA has been associated with more severe disease, highlighting a possible contribution of these antibodies to amplification. The observation that RF only appears after disease onset in some patients suggests that it may mark distinct, sequential events in pathogenesis that are close together in the subgroup in whom RF occurs early or are separated in time in patients in whom RF follows symptoms. The possible mechanisms whereby RF may play an amplifying role are numerous and include amplification of antigen capture, signaling, and effector functions, among others.
Table 56-1  Citrullinated and Noncitrullinated Autoantigens in Rheumatoid Arthritis (RA)

<table>
<thead>
<tr>
<th>Autoantigen</th>
<th>Size (kD)</th>
<th>Function (Location)</th>
<th>Cell/Tissue Type</th>
<th>Epitopes</th>
<th>Prevalence of Autoantibodies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrullinated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldolase</td>
<td>160</td>
<td>Glycolytic enzyme (cytosol)</td>
<td>Widely expressed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calreticulin</td>
<td>55</td>
<td>Ca^{2+} sensor/chaperone (endoplasmic reticulum)</td>
<td>Widely expressed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collagen type II</td>
<td>=140</td>
<td>Structural component (extracellular)</td>
<td>Cartilage</td>
<td>CII (aa 359-369) J1 (aa 551-564) U1 (aa 494-504) CEP-1 (aa 5-21; cit 9 and 15)</td>
<td>15%-25% to unmodified antigen 40% to citrullinated CII 50% of CCP2-positive patients</td>
</tr>
<tr>
<td>α-Enolase</td>
<td>49</td>
<td>Glycolytic enzyme (cytosol)</td>
<td>Widely expressed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibrinogen/fibrin</td>
<td>340</td>
<td>Coagulation cascade (extracellular)</td>
<td>Plasm RA synovium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FUSE-BP 1/2</td>
<td>71/78</td>
<td>mRNA Trafficking (nucleus)</td>
<td>Widely expressed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSP-60</td>
<td>60</td>
<td>Heat shock protein (cytosol, mitochondria)</td>
<td>Widely expressed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAD4</td>
<td>74</td>
<td>Citrullination enzyme (nucleus)</td>
<td>Granulocytes, monocytes</td>
<td>aa 1-120</td>
<td>30%-40%</td>
</tr>
<tr>
<td>Vimentin (Sa)</td>
<td>57</td>
<td>Intermediate filament (cytoskeleton)</td>
<td>Spleen, placenta, RA synovium</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Noncitrullinated     |           |                           |                     |                               |                             |
| G6Pi                 | 63        | Glycolytic enzyme (cytosol) | Widely expressed    |                               | 12%-29%                      |
| hnRNP-A2 (RA33)3     | 33        | mRNA Processing (nucleus) | Widely expressed    |                               | 35%                         |
| IgG (Fc)             | =50       | Antibody (extracellular)  | Widely expressed    |                               | 50%-90%                     |

ANTICITRULLINATED PROTEIN AUTOANTIBODIES

Antikeratin Antibodies and Antiperinuclear Factor in Rheumatoid Arthritis

Although RFs have been clinically useful in diagnosis of RA, particularly in cases in whom the index of clinical suspicion is high, these autoantibodies have suffered from two shortcomings: (1) their lack of specificity for RA and (2) unclear kinetic and mechanistic connection to pathogenesis. Defining additional autoantibodies specific for RA has therefore been a major priority. A significant discovery in this regard was made initially by Nienhuis and Mandema in 1964 (Figure 56-1). They recognized an autoantibody that stained keratohyalin granules surrounding the nucleus in cells of human buccal mucosa, which they called antiperinuclear factor (APF). These APF antibodies were found in 49% to 91% of RA patients; they have reported specificities of 73% to 99%. The assay was difficult to standardize because not every donor of buccal mucosal cells demonstrated the staining pattern. The assay therefore did not find its way into routine clinical practice, but the specificity for RA suggested that the assay might be reporting on a potentially important pathogenic pathway. A second research group subsequently demonstrated that RA sera recognized antigens in the stratified squamous layer of esophageal epithelium—a staining pattern that they termed antikeratin antibodies (AKAs). Subsequent iterations of this assay used other forms of stratified squamous epithelium including human skin. Although this AKA assay also had good specificity properties, it was of limited sensitivity and not particularly convenient or easy to standardize. Neither assay therefore was broadly used in the clinical diagnostic arena. Like the LE cell in SLE, however, the assays provided a framework for antigen discovery that would later lead to highly specific and convenient tools for RA diagnosis and classification.
Discovery of Autoantibodies That Recognize Peptidylcitrulline

The initial discovery of antibodies in RA patients, which recognized keratinized squamous epithelium, prompted studies to define the target of these antibodies. It was initially shown that APF only stained a fully differentiated squamous epithelial layer and that APF staining was identical to the pattern produced by a monoclonal antibody specific for human (pro)filaggrin in indirect immunofluorescence of permeabilized squamous epithelial cells3 (see Figure 56-1). Although the antigen could not be definitively identified using this co-localization alone, it suggested that a differentiation state might affect antigen recognition. Subsequently, RA sera were shown to recognize a 40-kD protein extracted from human epidermis; this was identified as a neutral/acidic isoform of filaggrin.38 IgG from RA sera affinity purified against the 40-kD filaggrin protein was reactive in both the APF and AKA tests, demonstrating that these autoantibodies were similar or identical.39 Filaggrin (filament-aggregating protein) is produced during the late stages of terminal differentiation of epithelial cells. It is synthesized as a heavily phosphorylated precursor protein (profilaggrin) that consists of 10 to 12 filaggrin repeats.30 Profilaggrin is deposited in granules, and it is released by proteolytic cleavage during differentiation of the cells, by proteases that remain to be fully defined. Coincident with cleavage, the protein is dephosphorylated and a significant proportion (~20%) of the arginine residues are deiminated (i.e., converted to citrulline).41,44 A posttranslational modification mediated by the peptidylarginine deiminase (PAD) enzymes (see later).41 Because RA sera appeared to specifically target the neutral isoform of filaggrin (which is present and citrullinated in fully differentiated squamous epithelia), Schellekens and colleagues44 addressed whether RA sera were positive in the AKA/APF assays specifically recognized citrullinated peptides. Thus regions within the deduced amino acid sequence of human profilaggrin with a high antigenicity index and the largest number of arginine residues were selected to generate synthetic peptides, where arginine residues were substituted with citrulline. Using these citrullinated peptides, Schellekens and colleagues demonstrated that AKA/APF antibodies are directed against citrullinated filaggrin. It is important to highlight that antibodies against citrullinated proteins in RA do not recognize free citrulline but do recognize citrulline residues (i.e., peptidylcitrullines) within the context of peptides or protein sequences.

These initial studies became the basis for the truly major discovery that protein sequences containing citrulline residues are one of the most prominent targets for autoantibodies in RA. However, because the epidermis is not a target of rheumatoid inflammation and there is no evidence that (pro) filaggrin is expressed in articular tissues, this molecule cannot be the autoantigen that drives the AKA/APF response in the joint. The facts that anticitrullinated filaggrin antibodies were enriched in synovial tissue compared with the serum or synovial fluid46 and that these antibodies are synthesized locally by plasmacytes within the rheumatoid pannus46 strongly suggest that anticitrullinated filaggrin cross-reacts with an antigen enriched in the RA joint/synovial tissue. Because citrullination is a frequent event in different tissues,47-51 great caution must be exercised when attempting to ascribe a primary role for a specific antigen in driving the antipeptidylcitrulline antibody response. The relevance of filaggrin in RA is therefore largely historical but unlikely of pathogenic relevance. Instead, its conformation and high content of citrulline residues make it an excellent surrogate with which to detect antibodies against citrullinated molecules. Many other citrullinated autoantigens found in the rheumatoid joint, which are currently being characterized, are much more likely to be of pathogenic relevance (see later).

Anti-CCP Antibodies: Clinical Relevance

In the initial characterization of AKA/APF antibodies, Schellekens and colleagues44 used a single C-terminal peptide derived from filaggrin (amino acids 306 to 324) to generate nine variants in which five arginine residues were changed to citrullines, either individually or in pairs, and these peptides were all assayed by ELISA. Interestingly, although the peptides were almost identical (the only difference being that the citrulline residues had different positions within the peptide), there were remarkable differences in the serum reactivity patterns toward each peptide (from 20% to 48% positivity), suggesting that although citrullination plays a critical role in antigen recognition by AKA/APF antibodies, the modification per se is not the only determinant that confers antibody binding. Instead, the data suggested that AKA/APF represent a pool of antibodies that recognize citrulline residues depending on the context of their surrounding amino acids. Indeed, when data from all peptides were pooled, the sensitivity increased to as much as 76%.44 Because peptides often adopt a β-turn conformation within the antibody-peptide complex52 and cysteine-bridged cyclic peptides have been shown to mimic the β-turn structure of the original antigenic determinant and can bind with enhanced affinity to antibodies,53 Schellekens and colleagues48 engineered a cyclic peptide (substituting the terminal serine residues with cysteines and cyclizing the peptide through the formation of a disulfide bond) to which ACPs reacted with higher affinity (see Figure 56-1). Using this cyclic citrullinated peptide (later named CCP1) and its linear counterpart, it was demonstrated that the cyclic structure increased the sensitivity of the assay (68% vs. 49%, without affecting specificity),44 although it was still less sensitive than the assays using the combination of linear peptides (i.e., 76% vs. 68%).44 Nevertheless, CCP1 became the antigen for the first generation of ELISAs designed to detect antibodies against citrullinated autoantigens. To improve the CCP1 test, libraries of ELISAs designed to detect antibodies against citrullinated molecules were used to construct the second-generation anti-CCP assay (CCP2), which was broadly adopted for clinical use.55,56 In 2005 a third generation of anti-CCP (CCP3) was made available for the laboratory diagnosis of RA. These assays have been reported to recognize additional citrullinated epitopes that are not identifiable with the second-generation CCP assays. In studies directly comparing second- and third-generation anti-CCP assays, the assays performed similarly,57,58 with a slightly increased sensitivity of CCP3 in some studies (e.g., sensi-
Anti-PAD and anti-CCP antibodies predict disease severity

Anti-PAD and anti-CCP antibodies precede disease onset

PAD4 polymorphisms are associated with RA and anti-CCP antibodies in some cohorts

PAD4 is an autoantigen

PAD4 can autocitrullinate

Inflammatory cells in the target tissue express PAD2/PAD4 and citrullination is activated by TNF

Citrullinated proteins accumulate in the target tissue

PADs convert peptidyl arginine to citrulline

Figure 56-2  Protein citrullination as a hub in rheumatoid arthritis (RA) pathogenesis. CCP, cyclic citrullinated peptide; PAD, peptidylarginine deiminase; TNF, tumor necrosis factor.

The clinical importance of anti-CCP antibody tests in RA stems from several favorable features (Figure 56-2): (1) Anti-CCP antibodies have high specificity for the diagnosis of RA. In systematic reviews and meta-analyses, anti-CCP antibody positivity has been shown to be as sensitive as, but more specific than, RF in distinguishing RA from other forms of inflammatory arthritis.17 (2) The presence of anti-CCP autoantibodies is an important predictor for development of RA. More than 90% of the patients with undifferentiated arthritis who tested positive for anti-CCP develop RA within 3 years, in contrast to only 25% of the anti-CCP negative patients.59 (3) Anti-CCP positivity has been associated with a more severe and destructive disease course, although the independent role of the antibody in comparison with RF has not yet been confirmed. Some investigators have shown that both RF and anti-CCP antibodies are independent predictors of severity, whereas others have shown that anti-CCP antibodies rather than RF are the better predictor of radiographic progression,60 particularly in those patients who were seronegative for RF.57

The early and accurate diagnosis of RA, together with early, adequate use of disease-modifying antirheumatic drugs, decreases accrual of joint damage in RA. Because anti-CCP antibodies are a marker of disease severity and are detectable early in the disease course, they are powerful tools to classify patients with early inflammatory arthritis who will benefit from treatment.68-70

Kinetics of Appearance of Anti-CCP Antibodies in Rheumatoid Arthritis

Recent studies have demonstrated that autoantibodies may precede onset of clinical disease in both tissue-specific and systemic autoimmune diseases.15,17-19 Defining events before clinical onset is challenging, and several different study designs have been employed to define the appearance of autoantibodies and their relationship with disease. The two general categories include (1) stored blood samples collected before disease onset (blood banks or military cohorts); (2) prospective studies examining emergence of autoantibodies and disease in high-risk individuals (often relatives of affected individuals).78 In RA, the former approach has clearly shown that autoantibodies precede RA in many individuals who subsequently develop seropositive RA, often by 2 to 6 years.76,77,79-80 In several studies, 20% to 60% of RA patients were RF positive before diagnosis and 30% to 60% of patients were positive for anti-CCP.76,77,79,80 With one exception,80 the prevalence of anti-CCP before diagnosis was almost twice as high as RF. Interestingly, although RF remained present consistently, anti-CCP could vary over time, even disappearing in some individuals. In one study, preclinical anti-CCP positivity was strongly associated with the development of erosive RA (odds ratio = 4.64; 95% confidence interval, 1.71 to 12.63; P < 0.01), whereas RF was not (P = 0.60).80 Together, these data highlight that the development of autoimmunity against citrullinated antigens is often asymptomatic and generally precedes the onset of clinical disease. This presymptomatic phase is potentially of great importance because it may identify individuals in whom important precursor conditions for the subsequent development of RA have been satisfied. The subsequent events that convert this RA precursor into a chronic, self-sustaining process are not yet known, but defining them is of great importance.

Anticitrullinated Protein Autoantibodies

Defining the primary antigen(s) that drive the generation of ACPAs has important implications in terms of RA etiology and pathogenesis. AKA/APF were the first ACPAs to be characterized, and they recognized citrullinated forms of filaggrin or its precursor profilaggrin.81

To date, several different approaches have been used to identify potential ACPA targets in RA. Although these approaches all use protein sequencing to identify autoantigens detected by ACPAs using immunoblotting assays, they vary in terms of the source of the antigens. Thus in addition to RA pannus, identification of citrullinated autoantigens has been extended to rheumatoid synovial fluid, as well as to cell lysates in which purified PADS have been used to generate in vitro citrullinated proteins. Using these approaches, several citrullinated autoantigen candidates have been identified. These include vimentin,82,83 fibronec-tin; α-enolase, elongation factor-1α, and adenyl cyclase-
Antibodies Recognizing Citrullinated Vimentin

The first local targets for ACPAs identified in RA synovial tissue were the citrullinated forms of the α and β chains of fibrin(ogen) \[92\] (see Table 56-1). Using protein extracts from rheumatoid synovial membranes and affinity-purified anti-citrullinated flaggrin antibodies, Masson-Bessière and colleagues \[92\] identified for the first time that RA synovial tissue contains several citrullinated proteins and defined among them two proteins that were specifically targeted by anti-citrullinated flaggrin antibodies. By amino-terminal sequencing the proteins were identified as the α- and β-chains of fibrin (although indeed, these sequences cannot allow distinguishing between fibrin and its precursor fibrinogen). \[92\] Fibrin is the cleavage product of fibrinogen, a molecule that is processed by thrombin in the final steps of the coagulation cascade. Fibrinogen is a dimeric glycoprotein, with each monomer composed of three polypeptides: the Aα (610 residues), the Bβ (461 residues), and the γ (411 residues). Thrombin cleaves fibrinogen at Aα Arg-16 and at Bβ Arg-14, which result in the release of fibrinopeptides A and B, and exposure of the E A α and E B β polymerization sites. Cleavage of fibrinogen leads to formation of fibrin monomers that spontaneuously polymerize to form a fibrin clot. Interestingly, fibrin deposition is usual during tissue inflammation and the presence of fibrin has long been known to be particularly abundant in rheumatoid synovial tissue,\[93\] likely as a result of an altered balance between coagulation and fibrinolysis.\[87,93\] Because circulating fibrinogen is not citrullinated, the presence of citrullinated fibrin in the rheumatoid synovium strongly suggested that after fibrin deposition, citrullination likely occurs in situ through the activity of locally expressed PAD. This proposal (i.e., that the joint is the site for autoantigen citrullination in RA) has been further supported and extended to many other citrullinated autoantigens in RA.\[97\] Citrullinated fibrinogen antibodies are among the most frequent ACPAs found in patients with RA (60% to 66%).\[96-98\] In total, 54 of the 81 arginines (66%) in human fibrinogen have been found to be susceptible to citrullination.\[99,101\] Using linear citrullinated peptides, major citrullinated epitope regions have been identified in this molecule, with some overlap among studies, but also with clear differences.\[100,102\] Although some of these peptides have been used to detect ACPAs against citrullinated fibrinogen, whole citrullinated fibrinogen is the more widely used antigen in these assays.\[96-98\]
Antibodies Recognizing Citrullinated Enolase

Antibodies against the immunodominant epitope of α-enolase, citrullinated α-enolase peptide 1 (CEP-1), occur in approximately half of anti-CCP2-positive patients (see Table 56-1). Enolase, also known as phosphopyruvate dehydratase, is a metalloenzyme responsible for the catalysis of the conversion of 2-phosphoglycerate to phosphoenolpyruvate, the penultimate step of glycolysis. It is also involved in various processes including hypoxia tolerance, growth control, and fibrinolysis. There are three subunits of enolase, α, β, and γ, each encoded by a separate gene that can combine to form five different isoenzymes: αα, αβ, αγ, ββ, and γγ. Three of these isoenzymes (all homodimers) are more commonly found in adult human cells than the others: αα or non-neuronal enolase (also known as enolase 1) is found in a variety of tissues including liver, brain, kidney, spleen, and adipose; ββ or muscle-specific enolase (enolase 3); and γγ or neuron-specific enolase (enolase 2). Citrullinated α-enolase is a target for ACPAs. However, whether this modification plays a role in the physiologic function of the enzyme or whether its citrullination only occurs under pathologic conditions (e.g., during cell death, either intracellularly or extracellularly) remains unclear. Using cyclic citrullinated α-enolase peptides covering 15 out of 17 arginine residues present in human α-enolase, CEP-1 was identified as the immunodominant B-cell epitope. This region comprised amino acids 5 to 21 of α-enolase, with arginine-9 and arginine-15 replaced by citrulline. Although antibodies against CEP-1 have not shown a better diagnostic or prognostic value than anti-CCP2, α-enolase has gained special attention as a potential target for disease initiation (see later).

Antibodies Recognizing Citrullinated Collagen Type II

Type II collagen is the predominant collagenous component of cartilage and is an autoantigen implicated in disease initiation in RA (see Table 56-1). Immunization with collagen type II is associated with development of autoimmune arthritis in several species, and the epitope-specific recognition of type II collagen by RA antibodies is shared by antibodies that are arthritogenic in collagen-induced arthritis (CIA) in the mouse. The antibody response to collagen type II is predominantly directed toward conformational triple-helical structures, and the identification of relevant B-cell epitopes has required the construction of recombinant triple-helical proteins and synthetic triple-helical peptides. The major B-cell epitopes on triple-helical collagen type II identified including C1 III (amino acids 359 to 369), J1 (amino acids 551 to 564), and U1 (amino acids 494 to 504) seem to share a common motif, comprising an arginine–glycine-hydrophobic amino acid motif. The observation that arginines occur in most of the collagen type II epitopes has raised the possibility that these may be citrullinated by PADs and that autoantibodies might bind specifically to the citrullinated forms. Indeed, using triple-helical molecules with the C1 epitope citrullinated, the diagnostic sensitivity for RA was 40%, compared with only 15% to 25% using the unmodified antigen.

Citrullinated Antigen Generation in Rheumatoid Arthritis

Peptidylarginine Deiminase Enzymes

The PADs, which are deiminating enzymes that hydrolyze guanidinium side chains in peptidylarginine to yield peptidylcitrulline and ammonia, belong to a larger group of guanidino-modifying enzymes called the amidinotransferase (AT) superfamily. Additional members of this superfamily include (1) the arginine deiminases (ADIs), enzymes found in both prokaryotes and the primitive eukaryote Giardia intestinalis, that act on nonpeptidyl arginine and that are involved in energy production; (2) the dimethylarginine dimethylaminohydrolases (DDAHs), enzymes found in both bacteria and mammals that convert nonpeptidyl methylguanidinerines into citrulline; (3) the ATs, enzymes involved in both creatine and streptomycin biosynthesis; and (4) the dihydrolases, bacterial enzymes involved in arginine catabolism that catalyze two successive hydrolytic steps.

The PADs, which are deiminating enzymes that hydrolyze guanidinium side chains in peptidylarginine to yield peptidylcitrulline and ammonia, belong to a larger group of guanidino-modifying enzymes called the amidinotransferase (AT) superfamily. Additional members of this superfamily include (1) the arginine deiminases (ADIs), enzymes found in both prokaryotes and the primitive eukaryote Giardia intestinalis, that act on nonpeptidyl arginine and that are involved in energy production; (2) the dimethylarginine dimethylaminohydrolases (DDAHs), enzymes found in both bacteria and mammals that convert nonpeptidyl methylguanidinerines into citrulline; (3) the ATs, enzymes involved in both creatine and streptomycin biosynthesis; and (4) the dihydrolases, bacterial enzymes involved in arginine catabolism that catalyze two successive hydrolytic steps. Whereas most superfamily members act on nonpeptidyl amino acids, PADs are highly specific for peptidylarginine residues and require at least one additional residue N-terminal to the site of modification. Although the chemical reaction catalyzed by PAD enzymes is officially called peptidylarginine deimination, the term citrullination is more generally used. PAD and PAD-like enzymes are present in numerous species, from bacteria to humans. Interestingly, the bacterial PADs are composed of only the approximately 40-kD catalytic domain, whereas mammals have larger multidomain enzymes (=75 kD), whose activity is regulated by calcium. The extra N-terminal domains in the mammalian enzymes include two immunoglobulin-like domains that are proposed to mediate protein–protein interactions and/or substrate targeting. To date, five human PAD homologs have been identified and the PADI genes are located at a single cluster on chromosome 1p36.1. For historical reasons, these isoenzymes are designated PADI-1 to PADI-6. Human PAD4 was initially named PAD5 but was later renamed PAD4 to reflect the fact that it is a true ortholog of this isoform. PAD4 is the only PAD that resides in the nucleus, a function of its nuclear localization sequence.

The PADs share 50% to 55% sequence identity with each other, with different PADs preferentially expressed in specific tissues. PAD1 is primarily expressed in uterus and skin. PAD2 is more widely expressed in muscle, skin, brain, spleen, rheumatoid synovium, and secretory glands. PAD3 is also expressed in skin, PAD4 is expressed in hematopoietic cells, and PAD6 is found in germ cells and peripheral blood leukocytes. Because of their prominent expression in rheumatoid synovial tissue and fluid, PAD2 and PAD4 have gained prominence as potential candidates that drive citrullination of self-antigens in RA.

Peptidylarginine Deiminase Structure, Activity, and Regulation

The three-dimensional structure of PAD4 has been solved and is likely representative of the broader family. PAD4 is a dimer formed by head-to-tail contact between the N-terminal domain of one molecule and the C-terminal domain of the second. Each monomer adopts an...
differentiate into granulocytes by retinoic acid\(^{13}\) and later described in peripheral blood granulocytes.\(^{5}\) The expression of PAD4 during granulocyte differentiation initially suggested that this enzyme may have a role during granulopoiesis. Histones H2A, H3, and H4, as well as nucleophosmin/B23, were the first PAD4 substrates to be identified.\(^{54}\) PAD4 appears to function as a transcriptional co-regulator, mediated by its ability to catalyze the deimination of specific residues present in the N-terminal tails of histones.\(^{148-150}\) Interestingly, mice deficient in PAD4 also have normal development,\(^{151}\) suggesting that this enzyme has no essential roles in steady-state cellular functions and/or development. Recent studies have demonstrated that PAD4-mediated citrullination of histones is required for NET (neutrophil extracellular traps) formation and bacterial clearance,\(^{150,151}\) suggesting a role for this enzyme in immune and inflammatory effector functions.

PAD6, the most recently defined member of the PAD family, was initially cloned from mouse oocytes and named egg PAD (ePAD).\(^{152}\) PAD6 is essential for the formation of cytoplasmic lattices, and female mice deficient in PAD6 are infertile but otherwise normal.\(^{13}\)

### Peptidylarginine Deiminases in Rheumatoid Arthritis

Multiple PADS are likely responsible for autoantigen citrullination in RA, particularly PADS 2 and 4\(^{48,133,136}\) (see Figure 56-2). Although PAD4 has some unique characteristics that have focused significant attention on this protein, it is important to be aware that there is no evidence to suggest that pathologic citrullination in RA is exclusively or preferentially mediated by this isoform.

It is of interest that several polymorphisms in PAD4 have been genetically associated with RA development, particularly in Asian populations\(^{153-156}\) (see Figure 56-2). In this regard, two common haplotypes of the PADI4 gene were initially identified\(^{155}\) and designated “susceptible” or “non-susceptible” on the basis of their relative frequency in patients with RA versus controls. In the initial population, the odds ratio for the susceptible haplotype was almost 2; this effect was not observed in most Caucasian populations. Changes in mRNA stability of the susceptible PAD4 reported in the initial paper were not associated with changes in enzyme level. Furthermore, although the susceptible haplotype generates a PAD4 molecule containing three amino acid substitutions in the N-terminal region (i.e., Gly55-Ser, Val82-Ala, Gly112-Ala), current evidence suggests that these changes have no effect on the function of the protein.\(^{56-2}\) Indeed, they appear to affect conformation only within the N-terminal domain, but not the active site located in the C-terminal domain.\(^{157}\) The fact that PAD4 acts as a head-to-tail dimer may nevertheless allow conformational differences in the N-terminus of one molecule to influence the C-terminal domain of another and influence catalysis or even inhibition by autocitrullination. It has been proposed that PAD4 genotype may exert its effects on RA disease susceptibility as a consequence of its immunogenicity, rather than its enzyme activity.\(^{7}\)

Consistent with this hypothesis is the observation that, in addition to its ability to citrullinate RA autoantigens, PAD4 itself is an RA autoantigen, even in the

**Structural and Functional Implications of Protein Citrullination**

Arginine residues often play a central role in the maintenance of tertiary structure in proteins because the positively charged guanidino group is a versatile interacting partner, forming multiple intramolecular hydrogen bonds to backbone carbonyl oxygens, and also intermolecularly between different proteins.\(^{13}\) Thus during the process of citrullination, the conversion of arginine into citrulline reduces the net charge of the protein by the loss of a positive charge per citrulline residue. In addition, by changing the guanidino group in arginine to an ureido group in citrulline, arginine deamination may modify intramolecular and intermolecular interactions, producing changes in protein structure.\(^{140}\)

These structural changes likely also have important functional implications. Indeed, deamination has been implicated in several physiologic processes. In the skin (which expresses PADS 1, 2, and 3\(^{80,131,141}\)), deimination of filaggrin (a component of the cornified cell envelope) is critical for its degradation into free amino acids, which act as a natural moisturizing factor in the stratum corneum.\(^{48,133}\) Interestingly, PAD2-deficient mice have normal development and show no abnormalities in the skin,\(^{142}\) suggesting functional redundancy in mouse development. Whether PAD2 functions in the skin during injury or environmental perturbations is currently unknown. In the immune system, PAD2 (and/or potentially other PADS) citrullinates chemokines, a process that decreases chemokine activity.\(^{143-146}\) Thus PAD2 may play important roles in controlling effector functions related to environmental triggers.

Human PAD4 (first called PAD5) was first found in human myeloid leukemia HL-60 cells induced to

elongated fold in which the N-terminal domain forms two immunoglobulin-like subdomains (named subdomain 1 and 2), and the C-terminal domain forms an α/β propeller structure. Five Ca\(^{2+}\)-binding sites were identified in the structure, and Ca\(^{2+}\) binding induces conformational changes required to generate the active site cleft. Thus PADS rely strongly on the presence of Ca\(^{2+}\) for activity and require millimolar amounts of calcium to be activated in vitro. Because calcium concentrations in cells do not rise above the low micromolar levels, the mechanisms that control PAD activation in vivo remain unclear. One possible explanation for the requirement of “extracellular” concentrations of calcium to achieve PAD activation is that protein citrullination occurs under extreme conditions where intracellular concentrations of Ca\(^{2+}\) approach extracellular concentrations (e.g., during necrosis or apoptosis).\(^{108}\) Whether other mechanisms regulate PAD enzyme activation, potentially by modifying the calcium requirement (e.g., allosteric effects of post-translational modifications or partner-binding), is not known. It is also important to note that although PAD activation provides one level of regulation, the activity of the PADS can also be negatively regulated through autocitrullination. This is not dissimilar to other enzymes, where automodification (e.g., autophosphorylation) modifies enzymatic function. Autocitrullination has been described for PADS 1, 2, 3, and 4.\(^{48,133,138}\) In the case of PAD4, direct citrullination of arginines surrounding the active site cleft appears to have a major impact on activity.\(^{10}\)
noncitrullinated form. (see Table 56-1 and Figure 56-2). The sensitivity of PAD4 antibodies for RA is in the 30% to 40% range, with specificity of greater than 95%. PAD4 autoantibodies are associated with more severe, erosive RA that persists despite treatment with tumor necrosis factor (TNF) inhibitors and, like anti-CCP, are frequently present early before onset of any symptoms. Interestingly, the susceptible PAD4 haplotype was strikingly associated with PAD4 autoantibodies, even in Caucasian populations in whom the RA disease association was not evident.

**Genetic Associations with Anti-CCP/ACPAs**

Although the association of specific HLA-DR alleles and RA has been known for decades, the relationship between genetics and the development of RA autoantibodies is just beginning to be elucidated. A subset of HLA-DR alleles termed the “shared epitope” (SE) alleles includes HLA-DR*0101, *0102, *0401, *0404, *0405, *0408, *0410, *1001, and *1402 and is united by a conserved sequence of amino acids (QRRAA, QKRAA, or RRRAA) on the α-helix of the DR-β chain peptide binding groove. SE alleles are strongly associated with the development of anti-CCP/ACPAs but are not independently associated with RF (Figure 56-3). Furthermore, there appears to be a gene dosage effect on the relative risk of anti-CCP development with an odds ratio of 3.3 to 4.7 for patients with one SE allele and an odds ratio of 11.8 to 13.3 for patients with two SE alleles. Analysis of individual SE alleles has revealed that anti-CCP/ACPAs are predominately associated with HLA-DR4 rather than HLA-DR1 SE alleles. Interestingly, smoking has been shown to skew the genetic association of SE alleles with anti-CCP development, with HLA-DR1 and HLA-DR10 SE alleles being more important in this patient group (see “Autoantibodies in Rheumatoid Arthritis: Insights into Disease Mechanism” later).

In addition to SE alleles, a single nucleotide polymorphism in the PTPN22 gene (1858C/T) is also associated with anti-CCP antibodies with an odds ratio of 3.80 (see Figure 56-3). PTPN22 encodes for lymphoid protein tyrosine phosphatase and has been shown to be associated with several autoimmune disorders including RA. The combination of PTPN22 1858C/T genotype and anti-CCP antibodies was 100% specific for RA and confers a relative risk of 130.3. The 1858C/T polymorphism was not associated with RF isotypes and appeared to act independently of SE alleles. Genetic factors such as the SE and PTPN22 may therefore play a role in the preclinical phase of RA, predisposing individuals to the generation of anti-CCP antibodies and susceptibility to the overt disease phenotype.

**OTHER AUTOANTIBODY SPECIFICITIES IN RHEUMATOID ARTHRITIS**

Anti-RA33 antibodies were discovered in 1989 as a novel autoantibody specificity recognizing a nuclear protein with an apparent molecular mass of 33 kDa. (see Table 56-1). These autoantibodies were initially detected in 35% of RA patients but in a small number of patients suffering from other autoimmune or degenerative rheumatic disorders. The antigen was termed RA33, and it was the first description of a nuclear antigen apparently specific for RA. Later, the antigen was identified as the A2 protein of the heterogeneous nuclear ribonucleoprotein complex (hnRNP-A2). Further characterization of this antibody showed that it was not strictly specific for RA but also found in approximately 20% of patients with SLE and in 40% to 60% of patients with mixed connective tissue disease (MCTD). However, in SLE and MCTD they usually occur together with antibodies to U1-snRNP or Sm. Therefore anti-RA33 without concomitant anti-U1-snRNP autoantibodies were found to have specificity of 96% for RA. Additionally, anti-RA33 antibodies in RA, SLE, and MCTD are distinguished by their recognition of different conformation-dependent epitopes in hnRNP-A2. Interestingly, despite their limited specificity, in the absence of RF and ACPAs, anti-RA33 antibodies in patients with very early arthritis (<3 months’ duration) are associated with a relatively mild nonerosive disease course, potentially identifying patients with a good prognosis who will respond well to treatment with disease-modifying antirheumatic drugs.

In studies in human samples, stimulation assays using hnRNP-A2 induced T cell proliferation in almost 60% of the RA patients (which were largely HLA-DR-restricted); T cell proliferation in response to hnRNP-A2 was present in only 20% of the controls, with substantially stronger responses in RA patients. Interestingly, immunohistochemical analyses revealed pronounced overexpression of hnRNP-A2 in synovial tissue of RA patients, placing the antigen at sites of pathology in RA. Another autoantigen of potential interest in RA is glucose-6-phosphate isomerase (G6PI). Arthritis in the K/BxN mouse model results from pathogenic immunoglobulins that recognize G6PI. Arthritis in the K/BxN mouse model results from pathogenic immunoglobulins that recognize G6PI, a glycolytic enzyme residing in the cytoplasm of all cells. Moreover, antibodies directed against G6PI can, alone, transfer arthritis to healthy recipients. Considering the pathogenic potential of this antibody, researchers were prompted to study it in human RA. However, although initial studies reported a high frequency of such antibodies in RA patient sera, these results have been the subject of debate.

Overall, anti-G6PI antibodies in RA patients range from 12% to 29% in different cohorts, with a higher prevalence in patients with active disease. Psoriatic arthritis, undifferentiated arthritis, and spondyloarthropathy patients also displayed anti-G6PI antibodies at similar frequencies (12% to 25%), and similar titers are detected in a proportion (5% to 10%) of control subjects or patients with Crohn’s disease or sarcoidosis. The potential relevance of these antibodies in RA is still uncertain.

More recently, homocitrulline-containing proteins (i.e., carbamylated proteins) have been recognized as an important target of autoantibodies in RA, with IgG and IgA antibodies recognizing carbamylated antigens identified in more than 40% of patients with RA. Most carbamylation in vivo is believed to take place during inflammation when myeloperoxidase is released from neutrophils. Although homocitrulline resembles citrulline (homocitrulline is one methylene group longer, but similar in structure), the finding that anticarbamylated protein antibodies are predictive for a more severe disease course in ACPA-negative patients has suggested that these antibodies are...
Figure 56-3  A feed-forward model of rheumatoid arthritis pathogenesis: a pivotal role for protein citrullination. A, During disease initiation, disease susceptibility genes interface with environmental factors to trigger aberrant protein citrullination and the development of autoantibodies. Citrullinated antigens are indicated (*). The asymptomatic initiation phase may persist for several months to decades before the onset of clinical symptoms, with circulating autoantibodies to citrullinated proteins being detectable during this time. B, Undefined transition events prompt the shift from the asymptomatic phase to the overt clinical phenotype. C, Once initiated, the immune response drives a feed-forward loop of target tissue destruction and disease propagation resulting in the clinical phenotype. Although aberrant protein citrullination may initially occur outside the joint, during disease propagation, citrullinated autoantigens are generated in the synovium by inflammatory cell peptidylarginine deaminases (PADs), leading to amplification of the immune response within the target tissue. anti-CCP, anticyclic citrullinated peptide; TNF, tumor necrosis factor.
not cross-reactive. Although the study of these novel autoantibodies is still in an early phase, it is likely that this is another post-translational modification that plays a central role in generating the autoantigens targeted in RA.

**AUTOANTIBODIES IN RHEUMATOID ARTHRITIS: INSIGHTS INTO DISEASE MECHANISM**

Although the pathogenesis of RA remains incompletely understood, current models employ a similar construct to other autoimmune rheumatic diseases; that is, in the setting of a complex genetic predisposition, environmental and stochastic events function to initiate an autoimmune process recognizing specific self-antigens, where the immune response itself participates in immune amplification and tissue injury (see Figure 56-3). The striking association of RA (and particularly ACPA) with particular major histocompatibility complex (MHC) class II alleles underscores at least one component of this framework. Whether RA autoantibodies are directly pathogenic or whether they are markers of inflammation and/or damage remains uncertain, particularly for RFs. However, ACPAs have several features that suggest direct involvement in RA pathogenesis (see Figure 56-2): (1) The targets of these antibodies (i.e., citrullinated proteins) are abnormally expressed and highly enriched in synovial tissue and fluid of RA patients; (2) the antibody response directed against these citrullinated antigens precedes RA onset, and is highly specific for RA and predictive of disease progression; (3) ACPA positivity is closely linked with the best known predisposing genetic risk factors for the development of RA: the HLA-DR shared epitope (SE) alleles, particularly, HLA-DRB1*04, linking genetic predisposition and autoantibody production; (4) plasma cells producing antibodies against citrullinated proteins are present in RA synovial tissue; and (5) about half of the anti-CCP-positive RA patients have circulating immune complexes containing citrullinated fibrinogen, and immunohistochemical staining has shown co-localization of fibrinogen, immunoglobulin, and complement component C3 in RA pannus tissue; and (6) immune complexes containing citrullinated fibrinogen can also activate monocyte-derived macrophages to produce TNF via engagement of Fcγ receptors and TLR4. Similar observations implying overlapping pathogenic properties are likely also operative in the case of RFs in RA. There is also recent evidence that Fc-linked glycosylation patterns of immunoglobulins differ in RA and may influence pathogenicity of these antibodies.

**Potential Environmental Factors in ACPA Production and Rheumatoid Arthritis Pathogenesis**

Although abundant citrullination is found in RA synovial tissue once disease is established, no studies have addressed whether abnormal synovial citrullination precedes the onset of clinical RA (a requisite to explain the presence of antibodies against citrullinated proteins long before clinical disease). Such evidence will be challenging to obtain. At this stage, it is therefore important to be aware that the joint is not the only possible site of abnormal citrullination that might initiate the RA-specific immune responses to citrullinated autoantigens, and it is possible that the joint only becomes targeted secondarily after the ACPA immune response has been initiated at another site, as a consequence of an inflammatory event triggered by a common environmental exposure such as periodontal infection/inflammation and/or smoking (see Figure 56-3). Although there is no evidence that viral proteins are citrullinated in vivo, a subset of ACPA-positive RA sera has been demonstrated to cross-react with in vitro citrullinated peptides from Epstein-Barr virus EBNA-1 and human papillomavirus (HPV)-47 E2, 145-162 proteins. Defining the kinetics of appearance of these antibodies during disease development and demonstrating directly that these modified proteins are generated during natural infection will be important to determine if this cross-reactivity is relevant to RA pathogenesis.

There are tantalizing similarities and epidemiologic associations between periodontitis and RA including presence of bone erosions, association with similar MHC class II alleles and smoking, and enrichment of severe periodontal disease in patients with RA. These connections have focused attention on the study of oral pathogens that express enzymes with PAD activity, specifically *Porphyromonas gingivalis*. This is a gram-negative bacterium commonly present as a biofilm in the gingival crevice and intracellularly in oral epithelial cells. This organism uses numerous virulence factors to evade host defenses as it establishes itself as one of the predominant pathogens in periodontal pockets. *P. gingivalis* expresses its own citrullinating enzyme named PgPAD, which produces ammonia that likely has negative effects on neutrophil function. PgPAD is not an ortholog of mammalian PADS, having no sequence homology with mammalian PADS and showing significant differences in citrullination activity: (1) PgPAD can citrullinate free L-arginine in addition to protein-bound arginine, (2) the enzyme does not require calcium or any other metal ion for activity, and (3) it has stronger preference for citrullination of C-terminal arginines. *P. gingivalis* also produces arginine-gingipains, endopeptidases that cleave substrates after arginine residues, a process that generates C-terminal arginine residues, which are then citrullinated by PgPAD. Interestingly, both arginine gingipain and PgPAD are secreted and have been shown respectively to cleave and citrullinate human fibrinogen and α-enolase in solution. *P. gingivalis* per se contains a large number of endogenous citrullinated proteins. Thus either as a source of bacterial citrullinated proteins and/or bacterial enzymes that generate host protein citrullination, *P. gingivalis* has been proposed as a potential factor responsible for initiating an immune response against citrullinated proteins, in individuals genetically predisposed to developing RA. Interestingly, when *P. gingivalis* enolase is citrullinated in vitro, it is recognized by human antibodies against CEP-1, the immunodominant peptide from citrullinated human α-enolase. The noncitrullinated *P. gingivalis* enolase was not similarly recognized. Although it remains to be demonstrated whether bacterial enolase is citrullinated by the bacterial enzyme or even by host PADS (a process that may occur during bacterial lysis by phagocytic cells expressing PADS), these initial findings raise the intriguing possibility that ACPA specific-
ity might be initially driven by bacterial products, which are subsequently driven by citrullinated endogenous proteins in the microenvironment of the inflamed joint (see Figure 56-3).

Another environmental factor with potential relevance in initiating production of ACPAs is cigarette smoking (see Figure 56-3). Interestingly, recent evidence has demonstrated (at least in some populations) a strong interaction between smoking and SE alleles in the development of ACPA-positive RA. However, although primarily HLA-DRB1*04 alleles have been associated with ACPA production, in smokers, HLA-DRB1*0101, *0102, and *1001 alleles showed the closest association with positive anti-CCP. The observations that cigarette smoking induces increase expression of PAD2 and protein citrullination in lungs and that anti-CCP antibodies of the IgA class are found in about one-third of patients with recent-onset RA have suggested that the lung or other mucosal surfaces may play a role in generating or sustaining autoantibodies to citrullinated proteins. These gene-environment interactions, playing out in microenvironments outside the joint, add important dimensions to the framework within which RA pathogenesis must be investigated and understood. Although it is possible that none of the detailed mechanisms currently defined represent the unifying mechanism of pathogenesis in RA, the broader view of potential immunizing environments and the search for a frequent worldwide environmental exposure are of great importance.

Role of Autoantibodies in Rheumatoid Arthritis: Insights from Animal Models

Although animal models are not ideal settings to define the role of human ACPA in RA, several in vivo models have provided evidence to support a pathogenic role for these antibodies in inflammatory arthritis in animals.

1. The conversion of peptidylarginine to peptidylcitrulline significantly increases affinity of the interaction of peptides with the RA-associated HLA-DRB1*0401 MHC class II molecule (shared epitope) and leads to the activation of CD4+ T cells in human HLA-DRB1*0401 transgenic mice.

2. Citrullinated human fibrinogen can induce arthritis in mice transgenic for human HLA-DRB1*0401. Transgenic and wild-type mice were immunized with unmodified or citrullinated human fibrinogen. Approximately one-third of transgenic mice immunized with citrullinated fibrinogen but none of those immunized with unmodified fibrinogen developed inflammatory arthritis, although with some histologic features that were not entirely typical of human RA. Interestingly, wild-type mice immunized with either form of fibrinogen did not develop arthritis. These data indicate the importance of both the susceptible human MHC II genotype, in this case DRB1*0401, and citrullination of the antigen, in this case the human antigen, in mediating pathology. Although this initial study strongly supports the potential pathogenic role of protein citrullination in RA, there appears to be specificity for the citrullinated molecules because not all citrullinated antigens are similarly able to induce/amplify disease. For example, noncitrullinated and citrullinated collagen type II are both potent inducers of arthritis in rats, although the citrullinated form induced arthritis with slightly higher incidence and earlier onset. In the same experiments, citrullinated rat albumin was also used to immunize rats. However, despite the fact that the rats developed a strong immune response against citrullinated albumin, the rats had no signs of disease, demonstrating that an immune response toward any citrullinated protein is clearly not enough to induce clinical arthritis.

3. Antibodies against citrullinated proteins enhance tissue injury in experimental autoimmune arthritis. In this model, murine monoclonal antibodies specific to citrullinated fibrinogen were shown to enhance arthritis when co-administered with a submaximal dose of anticollegen II antibodies. These antibodies were shown bound to targets within the inflamed synovium of treated mice.

Selected References

10. Anuradha V, Chopra A: In the era of nephelometry, latex agglutination is still good enough to detect rheumatoid factor, J Rheumatol 32(12):2343–2344, 2005.


816

PART 7

| DIAGNOSTIC TESTS AND PROCEDURES IN RHEUMATIC DISEASES

63. Mewar D, Coote A, Moore DJ, et al: Independent associations of
anti-cyclic citrullinated peptide antibodies and rheumatoid factor
with radiographic severity of rheumatoid arthritis, Arthritis Res Ther
64. Syversen SW, Gaarder PI, Goll GL, et al: High anti-cyclic citrullinated peptide levels and an algorithm of four variables predict radiographic progression in patients with rheumatoid arthritis: results from
65. De Rycke L, Peene I, Hoffman IE, et al: Rheumatoid factor and
antibodies to cyclic citrullinated peptides are associated with severe
extra-articular manifestations in rheumatoid arthritis, Ann Rheum Dis
measured at disease onset help identify seronegative rheumatoid
of anti-cyclic citrullinated peptide antibodies are associated with the
response to methotrexate in recent-onset arthritis, Ann Rheum Dis
inflammatory polyarthritis patients with anti-cyclic citrullinated
peptide antibodies versus those without antibodies, Arthritis Care Res
necrosis factor-alpha therapy in patients with rheumatoid arthritis,
rheumatoid arthritis. Antikeratin antibodies precede the clinical
73. Arbuckle MR, McClain MT, Rubertone MV, et al: Development of
the onset of systemic lupus erythematosus in northern Sweden,
76. Kokkonen H, Mullazehi M, Berglin E, et al: Antibodies of IgG, IgA
and IgM isotypes against cyclic citrullinated peptide precede the
development of rheumatoid arthritis, Arthritis Res Ther 13(1):R13,
2011.
77. Rantapaa-Dahlqvist S, de Jong BA, Berglin E, et al: Antibodies
against cyclic citrullinated peptide and IgA rheumatoid factor predict
the development of rheumatoid arthritis, Arthritis Rheum
to investigating the natural history of preclinical rheumatoid arthritis
(RA) using first-degree relatives of probands with RA, Arthritis
rheumatoid arthritis is predicted by presence of antibodies against
cyclic citrullinated peptide before and at disease onset, and by
IgA-RF at disease onset, Ann Rheum Dis 65(4):453–458,
2006.
80. Majka DS, Deane KD, Parrish LA, et al: Duration of preclinical
rheumatoid arthritis-related autoantibody positivity increases in subjects with older age at time of disease diagnosis, Ann Rheum Dis
81. Girbal-Neuhauser E, Durieux JJ, Arnaud M, et al: The epitopes targeted by the rheumatoid arthritis-associated antifilaggrin autoantibodies are posttranslationally generated on various sites of (pro)
filaggrin by deimination of arginine residues, J Immunol 162(1):585–
and activity of citrullinating peptidylarginine deiminase enzymes

83.
84.
85.
86.
87.
88.

89.

90.
91.

92.

93.
94.
95.
96.

97.

98.

99.

100.
101.
102.
103.
104.

in monocytes and macrophages, Ann Rheum Dis 63(4):373–381,
2004.
Vossenaar ER, Despres N, Lapointe E, et al: Rheumatoid arthritis
specific anti-Sa antibodies target citrullinated vimentin, Arthritis Res
alpha-enolase as a candidate autoantigen in rheumatoid arthritis,
collagen type I antibody is a target of autoimmunity in rheumatoid
eukaryotic translation initiation factor 4G1 as novel autoantigen in
rheumatoid arthritis, Biochem Biophys Res Commun 341(1):94–100,
2006.
Goeb V, Thomas-L’Otellier M, Daveau R, et al: Candidate autoantigens identified by mass spectrometry in early rheumatoid arthritis are
chaperones and citrullinated glycolytic enzymes, Arthritis Res Ther
peptidyl arginine deiminase type 4 regulates protein citrullination
to the glycolytic enzyme alpha-enolase in sera from patients with
early rheumatoid arthritis, Arthritis Rheum 46(5):1196–1201,
2002.
synovial targets of the rheumatoid arthritis-specific antifilaggrin autoantibodies are deiminated forms of the alpha- and beta-chains of
Andersen RB, Gormsen J: Fibrinolytic and fibrin stabilizing activity
Gormsen J, Andersen RB, Feddersen C: Fibrinogen-fibrin breakdown
Kinloch A, Lundberg K, Wait R, et al: Synovial fluid is a site of
citrullination of autoantigens in inflammatory arthritis, Arthritis
Nielen MM, van der Horst AR, van Schaardenburg D, et al: Antibodies to citrullinated human fibrinogen (ACF) have diagnostic and
prognostic value in early arthritis, Ann Rheum Dis 64(8):1199–1204,
2005.
Vander CB, Cantaert T, Nogueira L, et al: Diagnostic value of antihuman citrullinated fibrinogen ELISA and comparison with four
other anti-citrullinated protein assays, Arthritis Res Ther 8(4):R122,
2006.
to citrullinated antigens in sera from patients with rheumatoid arthritis: association with HLA-DRB1 alleles, Ann Rheum Dis 68(5):736–
743, 2009.
Kubota K, Yoneyama-Takazawa T, Ichikawa K: Determination of sites
citrullinated by peptidylarginine deiminase using 18O stable isotope
labeling and mass spectrometry, Rapid Commun Mass Spectrom
enzymatic properties between hPADI2 and hPADI4, Biochem Biophys
recognized by the rheumatoid arthritis-specific autoantibodies to
specific Sa autoantigen is present in high concentration in the target
novel antigen-antibody system specific for rheumatoid arthritis [see




Full references for this chapter can be found on www.expertconsult.com.
References


10. Anuradha V, Chopra A: In the era of nephelometry, latent agglutination test is still good enough to detect rheumatoid factor, J Rheumatol 32(12):2343–2344, 2005.


817.e4

136.
137.
138.
139.
140.

141.

142.
143.

144.

145.

146.

147.
148.
149.
150.
151.
152.

153.

154.
155.

156.
157.

PART 7

| DIAGNOSTIC TESTS AND PROCEDURES IN RHEUMATIC DISEASES

expressed in rheumatoid arthritis synovium in close association with
deiminase 4 (PADI4) and citrullinated protein in synovial tissue of
N-terminal recognition by human peptidylarginine deiminase 4, Proc
Mechin MC, Coudane F, Adoue V, et al: Deimination is regulated at
arginine in proteins: multiple hydrogen bonds to backbone carbonyl
Tarcsa E, Marekov LN, Mei G, et al: Protein unfolding by peptidylarginine deiminase. Substrate specificity and structural relationships
of the natural substrates trichohyalin and filaggrin, J Biol Chem
2005.
Loos T, Mortier A, Gouwy M, et al: Citrullination of CXCL10 and
CXCL11 by peptidylarginine deiminase: a naturally occurring posttranslational modification of chemokines and new dimension of
of the NH2-terminal region of CXCL5 by proteases or peptidylarginine deiminases (PAD) differently affects its biological activity, J Biol
Proost P, Loos T, Mortier A, et al: Citrullination of CXCL8 by peptidylarginine deiminase alters receptor usage, prevents proteolysis,
and dampens tissue inflammation, J Exp Med 205(9):2085–2097,
2008.
2009.
residues in nucleophosmin/B23 and histones in HL-60 granulocytes,
Hagiwara T, Hidaka Y, Yamada M: Deimination of histone H2A and
H4 at arginine 3 in HL-60 granulocytes, Biochemistry 44(15):5827–
5834, 2005.
arginine methylation levels via demethylimination, Science
Li P, Li M, Lindberg MR, et al: PAD4 is essential for antibacterial
innate immunity mediated by neutrophil extracellular traps, J Exp
Wright PW, Bolling LC, Calvert ME, et al: ePAD, an oocyte and
early embryo-abundant peptidylarginine deiminase-like protein
that localizes to egg cytoplasmic sheets, Dev Biol 256(1):73–88,
2003.
PADI4, encoding citrullinating enzyme peptidylarginine deiminase
4, are associated with rheumatoid arthritis, Nat Genet 34(4):395–402,
2003.
gene associated with increased rheumatoid arthritis susceptibility in
gene associated with rheumatoid arthritis in a Japanese population is
not associated in a United Kingdom population, Arthritis Rheum
Burr ML, Naseem H, Hinks A, et al: PADI4 genotype is not associated with rheumatoid arthritis in a large UK Caucasian population,

158.
159.

160.

161.

162.

163.
164.

165.

166.

167.

168.
169.
170.

171.

172.

173.

174.
175.

176.

2011.
(PADI4) identified as a conformation-dependent autoantigen in
Harris ML, Darrah E, Lam GK, et al: Association of autoimmunity
to peptidyl arginine deiminase type 4 with genotype and disease
severity in rheumatoid arthritis, Arthritis Rheum 58(7):1958–1967,
2008.
Halvorsen EH, Pollmann S, Gilboe IM, et al: Serum IgG antibodies
to peptidylarginine deiminase 4 in rheumatoid arthritis and associations with disease severity, Ann Rheum Dis 67(3):414–417,
2008.
Halvorsen EH, Haavardsholm EA, Pollmann S, et al: Serum IgG
antibodies to peptidylarginine deiminase 4 predict radiographic progression in patients with rheumatoid arthritis treated with tumour
necrosis factor-alpha blocking agents, Ann Rheum Dis 68(2):249–252,
2009.
Gregersen PK, Silver J, Winchester RJ: The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis, Arthritis Rheum 30(11):1205–1213,
1987.
van der Helm-van Mil AH, Huizinga TW: Advances in the genetics
of rheumatoid arthritis point to subclassification into distinct disease
Huizinga TW, Amos CI, van der Helm-van Mil AH, et al: Refining
the complex rheumatoid arthritis phenotype based on specificity of
of HLA-DR3 and the shared epitope alleles, Arthritis Rheum
van der Helm-van Mil AH, Verpoort KN, le Cessie S, et al: The
HLA-DRB1 shared epitope alleles differ in the interaction with
smoking and predisposition to antibodies to cyclic citrullinated
Johansson M, Arlestig L, Hallmans G, Rantapaa-Dahlqvist S:
PTPN22 polymorphism and anti-cyclic citrullinated peptide antibodies in combination strongly predicts future onset of rheumatoid
arthritis and has a specificity of 100% for the disease, Arthritis Res
sequencing of the nuclear autoantigen RA33 shows that it is indistinguishable from the A2 protein of the heterogeneous nuclear ribonucleoprotein complex, J Clin Invest 90(3):1061–1066, 1992.
Steiner G, Skriner K, Hassfeld W, Smolen JS: Clinical and immunological aspects of autoantibodies to RA33/hnRNP-A/B proteins—a
link between RA, SLE and MCTD, Mol Biol Rep 23(3–4):167–171,
1996.
Meyer O, Tauxe F, Fabregas D, et al: Anti-RA 33 antinuclear autoantibody in rheumatoid arthritis and mixed connective tissue disease:
comparison with antikeratin and antiperinuclear antibodies, Clin Exp
autoantibodies are directed to the RNA binding region of the A2
protein of the heterogeneous nuclear ribonucleoprotein complex—
differential epitope recognition in rheumatoid arthritis, systemic
lupus erythematosus, and mixed connective tissue disease, J Clin
Nell-Duxneuner V, Machold K, Stamm T, et al: Autoantibody profiling in patients with very early rheumatoid arthritis: a follow-up study,
Fritsch R, Eselbock D, Skriner K, et al: Characterization of autoreactive T cells to the autoantigens heterogeneous nuclear ribonucleoprotein A2 (RA33) and filaggrin in patients with rheumatoid




The inflammatory response is the body’s natural defense against unchecked tissue damage from infection or injury. It occurs during the acute phase of an inciting event and, if the stimulus is not eliminated, in a chronic, healing stage. When excessive or uncontrolled, these responses have the potential to cause significant harm to the host through processes such as autoimmune diseases, allergic reactions, and septic shock.

The concept of inflammation itself has evolved over thousands of years; Celsus first described the cardinal signs of redness, swelling, heat, and pain in the 1st century AD. The fifth sign, loss of function, is often attributed to Galen’s work during the 2nd century. Since that time, countless advances have been made in the understanding of inflammation, from histologic findings of inflammatory cells within tissues, to the discovery of hematologic and soluble mediators such as cytokines and complement. Also, the molecular signaling pathways that drive both its protective effects and the inappropriate injurious responses have been recently elucidated.

With increasing insight into mechanisms of the inflammatory response has come an appreciation of its significant complexity. Molecular and microscopic processes are different during acute and chronic stages, and diverse responses are induced by various types of exogenous and endogenous stimuli (e.g., bacteria, viruses, parasites, crystals, allergens, ischemia). More recently, the role of the milieu of the inflammatory response, especially at the level of the vascular endothelium, has taken on considerable importance. Furthermore, inherent redundancies of functions have been noted, as have interactions between the mediators of inflammation that allow for a broad, effective response, but these redundancies make it difficult to understand the pathways of inflammation in a linear fashion. These intricacies have resulted in a vague definition and on continued reliance on the final downstream macroscopic cardinal signs to tie together all of the ongoing processes. It is these same intricacies that have made evaluation of inflammation through laboratory tests imprecise.

Basic hematologic abnormalities may give clues to the presence of inflammation, but different patterns are often associated with underlying causes and are not universally found. Leukocytosis can be seen in infections, acute crystal diseases, and some autoimmune disorders such as adult Still’s disease. Anemia is often associated with certain diseases that cause chronic inflammation, such as rheumatoid arthritis (RA). Reactive thrombocytosis occurs secondary to the release of cytokines after an inciting infectious or inflammatory event, and the role of platelets and platelet-derived mediators in stimulating inflammation has been described at the molecular level.

Laboratory tests most commonly used by physicians to objectively obtain information about the extent of inflammation are those that measure the acute phase reaction. In response to injury, local inflammatory cells secrete cytokines that influence the liver to increase or decrease production of various proteins. The erythrocyte sedimentation rate...
(ESR) has been the classic marker of inflammation, with serum C-reactive protein (CRP) taking on an increasingly prominent role. Other novel inflammatory markers have been recognized but have not been found to be more clinically useful.

CRP elevations, especially minor ones, have been noted in numerous conditions that traditionally have not been considered inflammatory, most significantly involving the cardiovascular system. This has shed light on subclinical inflammation as a possible factor in the pathogenesis of a great number of diseases.

**ACUTE PHASE RESPONSE**

Within minutes of tissue injury, activation of the innate immune system induces cytokine production that results in a multisystem acute phase response involving the liver, vascular system, bone marrow, and central nervous system. Many elements of the reaction can be regarded as part of the innate response and are defensive or adaptive in nature. Mouse studies have shown that up to 7% of the regulatable gene pool undergoes significant changes in expression during inflammation, and that induction of liver acute phase genes is mediated by the transcription factor signal transducer and activator of transcription 3 (STAT3).

Although the acute phase response can trigger numerous neuroendocrine, hematopoietic, and metabolic effects, it is the changes in plasma proteins synthesized by hepatocytes that are monitored as signs of underlying inflammation (Tables 57-1 and 57-2). An acute phase protein is one whose plasma concentration changes from baseline by at least 25% during inflammation; responses vary in terms of concentration and kinetics (Figure 57-1). CRP and serum amyloid A (SAA) levels increase more than 1000-fold during acute infection, and peak at 2 to 3 days. Concentrations of other proteins peak at longer periods and can range from a 50% increase in complement and ceruloplasmin, to a several-fold amplification in haptoglobin, fibrinogen, α1-proteinase inhibitor, and α1-acid glycoprotein. Other proteins are negative acute phase proteins whose concentrations fall during the inflammatory response. These include antithrombin III, protein S, prealbumin, albumin, transferrin, and apolipoprotein A-I.

Hepatic stimulation of acute phase proteins is induced by cytokines released by activated monocytes, macrophages, neutrophils, natural killer (NK) cells, and endothelial cells acting at the front lines of the inflammatory response. The main cytokine influencing the liver is interleukin (IL)-6, once called the “hepatocyte-stimulating factor.” It likely mediates protein expression via the Janus-activated kinase (JAK) and STAT3 pathways, as well as C/EBP family members and Rel proteins (nuclear factor κB [NFκB]). During initial stages, IL-1 and tumor necrosis factor (TNF) synergize with IL-6 and trigger further IL-6 production, but their roles are limited. The soluble IL-6 receptor amplifies IL-6 effects both locally and systemically. IL-6 also performs a protective role during disease, inducing the expression of an IL-1 receptor antagonist.

Acute phase protein levels are not uniform in their expression; this is likely related to the underlying pathophysiologic state and is regulated by different combinations and interactions of cytokines. The roles of the acute phase proteins themselves will be discussed throughout the chapter but have been found to include direct involvement in host defense by activation of the complement, proteinase inhibition, and antioxidant activity. However, some of the described in vitro effects of proteins may not be relevant in vivo.

**Table 57-1 Human Acute Phase Proteins**

<table>
<thead>
<tr>
<th>Patients Whose Plasma Concentrations Increase</th>
<th>Complement System</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>C3</td>
</tr>
<tr>
<td>C4</td>
<td>C4</td>
</tr>
<tr>
<td>C9</td>
<td>C9</td>
</tr>
<tr>
<td>Factor B</td>
<td>Factor B</td>
</tr>
<tr>
<td>C1 inhibitor</td>
<td>C1 inhibitor</td>
</tr>
<tr>
<td>C4b-binding protein</td>
<td>C4b-binding protein</td>
</tr>
<tr>
<td>Mannose-binding lectin</td>
<td>Mannose-binding lectin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coagulation and Fibrinolytic System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibrinogen</td>
</tr>
<tr>
<td>Plasminogen</td>
</tr>
<tr>
<td>Tissue plasminogen activator</td>
</tr>
<tr>
<td>Urokinase</td>
</tr>
<tr>
<td>Protein S</td>
</tr>
<tr>
<td>Vitronectin</td>
</tr>
<tr>
<td>Plasminogen-activator inhibitor 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Antiproteases</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Protease inhibitor</td>
</tr>
<tr>
<td>α-Antichymotrypsin</td>
</tr>
<tr>
<td>Pancreatic secretory trypsin inhibitor</td>
</tr>
<tr>
<td>Inter-α-trypsin inhibitors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transport Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ceruloplasmin</td>
</tr>
<tr>
<td>Haptoglobin</td>
</tr>
<tr>
<td>Hemopexin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participants in Inflammatory Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secreted phospholipase A2</td>
</tr>
<tr>
<td>Lipopolysaccharide-binding protein</td>
</tr>
<tr>
<td>Interleukin-1–receptor antagonist</td>
</tr>
<tr>
<td>Granulocyte colony-stimulating factor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-reactive protein</td>
</tr>
<tr>
<td>Serum amyloid A</td>
</tr>
<tr>
<td>α-Acid glycoprotein</td>
</tr>
<tr>
<td>Fibronectin</td>
</tr>
<tr>
<td>Ferritin</td>
</tr>
<tr>
<td>Angiotensinogen</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proteins Whose Plasma Concentrations Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albumin</td>
</tr>
<tr>
<td>Transferrin</td>
</tr>
<tr>
<td>Transthyretin</td>
</tr>
<tr>
<td>α-HS glycoprotein</td>
</tr>
<tr>
<td>Alpha-fetoprotein</td>
</tr>
<tr>
<td>Thyroxine-binding globulin</td>
</tr>
<tr>
<td>Insulin-like growth factor I</td>
</tr>
<tr>
<td>Factor XII</td>
</tr>
</tbody>
</table>

**Erythrocyte Sedimentation Rate**

Although ESR is an indirect screen for elevated concentrations of acute phase proteins, it has been the most widely used marker of inflammation for almost a century.
Measurement of ESR is performed when blood is placed in a vertical tube and the rate of fall of erythrocytes is measured. The ancient Greeks recognized increased red blood cell (RBC) sedimentation as a way to detect “bad bodily humors,” but our modern understanding and use of RBC sedimentation as a test date back to the German scholar Fahraeus in 1918. He determined that certain plasma proteins, especially fibrinogen, are able to lower the electrostatic charge on RBC surfaces so they can aggregate, form rouleaux, and fall faster.

Several factors are involved in acceleration of ESR. Asymmetric plasma proteins such as fibrinogen and, to a lesser extent, alpha, beta, and gamma globulins decrease the negative charge of erythrocytes (zeta potential) that prevents rouleaux formation. Red cell factors also play a role in that changes in plasma ratios in anemic states also favor rouleaux. However, microcytosis, polycythemia, and abnormally shaped RBCs (e.g., sickle cells, spherocytes) hinder aggregation and lower the ESR. Conditions that elevate fibrinogen, even if they are not necessarily considered inflammatory, can raise ESR. These include pregnancy, diabetes, end-stage renal disease, and heart disease. Major increases in the concentration of a single molecular species, such as a monoclonal immunoglobulin in multiple myeloma, also cause increased sedimentation. The ESR is elevated in obesity, as is CRP, presumably as a result of IL-6 secretion by adipocytes. Factors such as glucocorticoids, croglobulins, hypofibrinogenemia, and hyperviscosity have been shown to lower the value. The physiochemical dynamics that allows for sedimentation has been a continued source of debate, with disparate models presented to explain how proteins on cell surfaces interact to cause RBC aggregation.

Although novel and rapid tests for ESR have proved promising, the International Committee for Standardization in Hematology continues to recommend the Westergren technique of testing anticoagulated blood. The usual accepted upper limits of normal are 15 mm/hr for males and 20 mm/hr for females; however, the ESR increases with age and varies by race, calling the reliability of the test into question. A simple formula for calculating the upper limit of normal ESR at any age has been used regularly: In men, age in years divided by 2; in women, 10 plus age in years divided by 2. Despite the ability to control for age, other limitations of the test have been noted and are listed in Table 57-3. The relative virtues of CRP determination have diminished some of the importance of ESR, but it remains an easy, inexpensive test with a wealth of background literature. Therefore the sedimentation rate will continue to play a prominent role in clinical practice.

**C-Reactive Protein**

C-reactive protein is an acute phase protein whose serum concentration reflects ongoing inflammation better than other tests in most, but not all, diseases. CRP was identified in 1930, when sera obtained from patients with *Streptococcus pneumoniae* infection were found to contain a protein that could bind to the “C” polysaccharide of the bacterial cell wall. This protein circulates as a 115-kD pentamer of noncovalently linked 23-kD subunits, which has

---

**Table 57-2** Other Acute Phase Phenomena

<table>
<thead>
<tr>
<th>Neuroendocrine Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever, somnolence, and anorexia</td>
</tr>
<tr>
<td>Increased secretion of corticotropin-releasing hormone, corticotropin, and cortisol</td>
</tr>
<tr>
<td>Increased secretion of arginine vasopressin</td>
</tr>
<tr>
<td>Decreased production of insulin-like growth factor I</td>
</tr>
<tr>
<td>Increased adrenal secretion of catecholamines</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hematopoietic Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anemia of chronic disease</td>
</tr>
<tr>
<td>Leukocytosis</td>
</tr>
<tr>
<td>Thrombocytosis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metabolic Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss of muscle and negative nitrogen balance</td>
</tr>
<tr>
<td>Decreased gluconeogenesis</td>
</tr>
<tr>
<td>Osteoporosis</td>
</tr>
<tr>
<td>Increased hepatic lipogenesis</td>
</tr>
<tr>
<td>Increased lipolysis in adipose tissue</td>
</tr>
<tr>
<td>Decreased lipoprotein lipase activity in muscle and adipose tissue</td>
</tr>
<tr>
<td>Cachexia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hepatic Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased metallothionein, inducible nitric oxide synthase, heme oxygenase, manganese superoxide dismutase, and tissue inhibitor of metalloproteinase-1</td>
</tr>
<tr>
<td>Decreased phosphoenolpyruvate carboxykinase activity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changes in Nonprotein Plasma Constituents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypozincemia, hypoferremia, and hypercupremia</td>
</tr>
<tr>
<td>Increased plasma retinol and glutathione concentrations</td>
</tr>
</tbody>
</table>

Table 57-3  Comparison of Erythrocyte Sedimentation Rate and C-Reactive Protein

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Erythrocyte Sedimentation Rate</th>
<th>C-Reactive Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much clinical information in the literature</td>
<td></td>
<td>Rapid response to inflammatory stimuli</td>
</tr>
<tr>
<td>May reflect overall health status</td>
<td></td>
<td>Wide range of clinically relevant values are detectable</td>
</tr>
<tr>
<td>Disadvantages</td>
<td></td>
<td>Unaffected by age and gender</td>
</tr>
<tr>
<td>Affected by age and gender</td>
<td></td>
<td>Reflects value of a single acute phase protein</td>
</tr>
<tr>
<td>Affected by red blood cell morphology</td>
<td></td>
<td>Can be measured on stored sera</td>
</tr>
<tr>
<td>Affected by anemia and polycythemia</td>
<td></td>
<td>Quantitation is precise and reproducible</td>
</tr>
<tr>
<td>Reflects levels of many plasma proteins, not</td>
<td></td>
<td>None</td>
</tr>
<tr>
<td>all of which are acute phase proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responds slowly to inflammatory stimuli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requires fresh sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May be affected by drugs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

been highly conserved over hundreds of millions of years of evolution. In contrast to immunoglobulins and complement components, CRP deficiency in humans has not been described. Genome-wide associated studies performed recently have shown that at least seven distinct loci are involved in the basal expression of CRP, which is upregulated upon stimulation by the transcription factors C/EBP and Rel. It is present in trace concentrations in the plasma of all humans (roughly 1 mg/L, with higher concentrations in women and the elderly). Plasma C-reactive protein is synthesized by hepatocytes, although other sites of local production and possibly minimal secretion have been suggested.

The precise function of CRP is unknown and may be varied, but it exhibits important recognition and activation capabilities, and it binds to numerous ligands. CRP recognizes phosphocholine, phospholipids, fibronectin, chromatin, and histones, all of which are exposed at sites of tissue damage and by apoptotic cells; CRP may target them for clearance. CRP protein bridges the gap between innate and adaptive immunity by activating the classical complement pathway and interacting with cells of the immune system through binding of Fcγ receptors. CRP induces inflammatory cytokines, tissue factors, and shedding of the IL-6 receptor, all of which result in a complement-dependent increase in tissue damage. Other CRP functions are anti-inflammatory, including promoting the noninflammatory clearance of apoptotic cells and preventing neutrophil adhesion to the endothelium. Thus CRP may play many pathophysiologic roles during the course of the inflammatory process.

After an acute inflammatory stimulus, CRP concentration increases rapidly and peaks at 2 to 3 days at levels that reflect the extent of tissue injury. If the stimulus has been removed, serum CRP levels drop rapidly, with a half-life of roughly 19 hours. Persistent elevations in CRP are seen in chronic inflammatory states such as active RA, pulmonary tuberculosis, or extensive malignant disease.

Immunoassays and laser nephelometry are used at modest cost to quantify serum CRP levels. Most healthy adults have levels less than 0.3 mg/dL. The significance of minor elevations in CRP is being debated and will be discussed subsequently. However, usual methods of CRP determination are less precise at concentrations in the range of 0.3 to 1 mg/dL, so high-sensitivity (hs)CRP methods are used to accurately measure these levels. Generally, concentrations greater than 1 mg/dL reflect clinically significant inflammatory disease. Concentrations of 1 to 10 mg/dL can be considered to represent moderate increases, and concentrations greater than 10 mg/dL show marked increases. Most patients with extremely high levels (e.g., >15 mg/dL) have bacterial infection; one study found that in patients with CRP concentrations greater than 50 mg/dL, infection was present in 88% of subjects. Clinical conditions associated with varying degrees of elevation of CRP are listed in Table 57-4, and the range of CRP concentrations in many rheumatologic diseases is shown in Figure 57-2.

Several limitations associated with the use of C-reactive protein measurement must be acknowledged. No uniformity in reporting concentrations has been noted between laboratories, and values can be conveyed in mg/L, µg/mL, or mg/dL. Similar to ESR, population studies show a skewed, rather than Gaussian, distribution, leaving parametric statistical tests inappropriate for interpretation of CRP data. Population differences in CRP levels in the United States have been reported between sexes and among racial groups.

Table 57-4  Conditions Associated with Elevated C-Reactive Protein Levels

<table>
<thead>
<tr>
<th>Normal or Minor Elevation (&lt;1 mg/dL)</th>
<th>Moderate Elevation (1-10 mg/dL)</th>
<th>Marked Elevation (&gt;10 mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vigorous exercise</td>
<td>Myocardial infarction</td>
<td>Acute bacterial infection (80%-85%)</td>
</tr>
<tr>
<td>Common cold</td>
<td>Malignancies</td>
<td>Major trauma</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>Pancreatitis</td>
<td>Systemic vasculitis</td>
</tr>
<tr>
<td>Gingivitis</td>
<td>Mucosal infection (bronchitis, cystitis)</td>
<td></td>
</tr>
<tr>
<td>Seizures</td>
<td>Most connective tissue diseases</td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>Rheumatoid arthritis</td>
<td></td>
</tr>
<tr>
<td>Insulin resistance and diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Several genetic polymorphisms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
was higher than 95% after 6 years for individuals whose median SAA was less than 10 mg/L, but 40% if median SAA remained above 10 mg/L. However, reliable testing for acute phase SAA is not yet widely available, and data about levels expected in disease are limited.

Other Acute Phase Proteins

Measurement of other acute phase proteins has been of limited value clinically, because their responses to tissue injury are often slower, and the magnitude of concentration change is smaller than with CRP and SAA. Serum ferritin is moderately increased, triggered by cytokines such as IL-1, IL-6, IL-18, and TNF. Levels are frequently high in adult-onset Still’s disease and systemic lupus erythematosus (SLE), and they correlate with disease activity. Hepcidin, a liver-derived antimicrobial peptide and regulator of iron homeostasis, is induced by inflammation, and by IL-6 in particular. Its levels rise in parallel with ferritin, and it is important in the development of anemia of chronic disease, acting as a negative regulator of iron absorption and macrophage iron release teleologically, to deprive microbes of iron. Transferrin, which binds and transports iron, is a negative acute phase protein.

The prohormone of calcitonin, procalcitonin, is produced by many cell types during severe infection and has been used as a marker to distinguish bacterial infection from other types of inflammatory processes, with some caveats. Apolipoprotein a-I, the principal protein constituent of HDL, is another negative acute phase protein. In chronic inflammatory diseases such as RA and SLE, decreased levels may contribute to increased risk of thrombotic events. Serum albumin and prealbumin (transthyretin) are also negative acute phase proteins, although their measurement has not been shown to be more helpful in diagnosis or prognosis than standard tests. Serum complement fractions become depressed when the system is activated in certain autoimmune disorders but otherwise rise during the acute phase response.
Table 57-5  Products of Inflammatory, Endothelial, and Resident Target Tissue Cells/Matrix

<table>
<thead>
<tr>
<th>Cytokines and Related Molecules</th>
<th>Products of Inflammatory and Endothelial Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytokines:</td>
<td></td>
</tr>
<tr>
<td>IL-1</td>
<td>Calprotectin</td>
</tr>
<tr>
<td>IL-6</td>
<td>von Willebrand factor</td>
</tr>
<tr>
<td>IL-12</td>
<td>Soluble adhesion molecules (e.g., sVCAM and sE-selectin)</td>
</tr>
<tr>
<td>Interferon-α</td>
<td>Hyaluronic acid</td>
</tr>
<tr>
<td>Tumor necrosis factor</td>
<td>Collagen and aggrecan degradation products</td>
</tr>
<tr>
<td>Granulocyte-macrophage colony-stimulating factor</td>
<td>Osteocalcin</td>
</tr>
<tr>
<td>IL-1 receptor antagonist</td>
<td></td>
</tr>
</tbody>
</table>

IL, interleukin.

Cytokines

Although not acute phase proteins in the classic sense, cytokines display the most striking acute phase behavior of any circulating proteins. IL-6 responds dramatically to tissue injury, with concentration changes that are faster and greater than those of CRP or SAA. Acute inflammation and chronic inflammation have been associated with increases in IL-6, and serum levels of this cytokine have been correlated with the severity and course of disease in RA, juvenile arthritis, ankylosing spondylitis, and polymyalgia rheumatica (PMR). It is also more sensitive than ESR for detecting disease activity in giant cell arteritis (GCA). IL-6 levels may be useful in monitoring inflammation if hepatocytes are damaged to the point of not being able to synthesize acute phase proteins. The importance of cytokines such as TNF and IL-1 has been inferred by the successful reduction of inflammation by their therapeutic inhibitors. Certain diseases, such as the TNF receptor-associated periodic syndrome (TRAPS) and the autoinflammatory syndromes that involve mutations of the inflammasome controlling IL-1, point to the importance of these cytokines.

Increased levels of several other cytokines and circulating cytokine receptors have been associated with inflammation or disease activity as well (Table 57-5). Different patterns of cytokine responses have been reported in different diseases, suggesting that cytokine determinations are potentially useful clinically. However, their quantitation presents several problems related to their short plasma half-lives, the presence of blocking factors and natural inhibitors, and other technical considerations. At present, high costs, limited availability, and absence of standardization discourage measurement of plasma cytokines and their receptors in clinical practice.

**ACUTE PHASE REACTANTS IN THE MANAGEMENT OF RHEUMATIC DISEASES**

Measurement of ESR and CRP has no role in the diagnosis of any particular disease, including RA, osteoarthritis, SLE, PMR, GCA, or other inflammatory arthropathies. However, these measurements can be clinically helpful in three ways:

1. In evaluating the extent or severity of inflammation,
2. In monitoring changes in disease activity over time,
3. In assessing prognosis.

**Rheumatoid Arthritis**

ESR and CRP cannot be used in the diagnosis of RA because 45% of patients may have normal serum levels at presentation, although these values represent part of the diagnostic syndrome or classification criteria sets. These tests are more appropriately applied in RA for monitoring disease activity and response to therapy. Although ESR traditionally has been more widely used for these purposes, many studies have suggested that CRP levels correlate better with disease activity. Some recent reports state that CRP levels may overestimate disease response compared with ESR; others claim that differences between the two are minimal. The existence of patients with depressed CRP concentrations caused by carrying low-CRP–associated genetic variants must be taken into account when this test is used universally. Matrix metalloproteinase (MMP)-3, pro-MMP-3, and soluble E-selectin have also been proposed as markers for RA disease activity; their measurements correlate with CRP levels but do not provide more information than is provided by standard tests.

CRP levels average 2 to 3 mg/dL in adult RA patients with moderate disease activity. However, variation is considerable: At least 5% to 10% of patients have values in the normal range, whereas a few patients with severe disease activity have levels greater than 10 mg/dL. ESR values have been found to remain stable over the years. ESR and CRP have long been used to follow the response to therapy; in general, effective disease-modifying antirheumatic drug therapy decreases CRP by about 40%. Inhibition of joint damage by these agents usually is accompanied by marked improvement in acute phase reactants. Progression of joint damage can occur while the patient is on therapy, however, despite decreases in ESR and CRP. Even more striking improvement has been seen with biologic agents introduced since the 1990s, providing objective laboratory support for the encouraging clinical responses observed. In early reports of anti-TNF therapy, CRP and SAA levels declined by 75% and 85% in about 1 week. Treatment with abatacept, the T cell CD80/CD86:CD28 co-stimulation modulator, resulted in significant decreases in CRP at both 90 and 360 days of therapy. In one study, failure to suppress CRP levels 2 weeks after initiation of infliximab therapy identified most patients who would prove to be clinical nonresponders after 12 weeks. In contrast to traditional disease-modifying antirheumatic drugs, TNF inhibitors have been found to inhibit joint damage even while clinical activity, reflected by CRP levels, remains high. Tocilizumab, a human IL-6 receptor antibody, improves RA by inhibiting effects of the cytokine. However, owing to its mechanism of action, inflammatory markers such as ESR and CRP drop to negative values, so that tracking them may not reflect the actual effect of the drug; care must be taken when monitoring disease activity during tocilizumab therapy.

ESR and CRP also have value as prognostic indicators in RA. Elevated acute phase reactant levels are associated with early synovitis and erosions as detected by magnetic resonance imaging, with inflammatory cellular infiltrates in...
Systemic Lupus Erythematosus

Although serum levels of CRP often parallel disease activity in autoimmune disorders, it has been recognized that SLE is an exception. Although marked CRP responses are seen in subsets of patients, such as those with serositis or chronic synovitis, many (such as patients with nephritis) show mild or no elevation during periods of activity. Serum levels of SAA are also relatively low in comparison with those of RA patients. In contrast, ESR correlates with disease activity and accrued tissue damage in SLE. Fibrinogen levels increased over time in patients, regardless of disease activity. Data are insufficient to evaluate the potential use of some of the other newer markers described previously, but many SLE patients with normal CRP levels show elevated IL-6 concentrations. Therefore a deficiency in IL-6 does not explain the muted CRP response in SLE.

Although the concomitant decrease in SAA may point against it, several investigations have raised the possibility that low CRP levels may be related to the pathogenesis of SLE: (1) An association has been noted between SLE and a genetic polymorphism associated with low CRP levels, (2) it has been observed that low CRP levels may contribute to defective clearance of autoantigens during apoptosis, and (3) the therapeutic efficacy of CRP has been reported in mouse models of SLE. Recent studies have also raised the possibility that type I interferon (IFN), which has been shown to be expressed significantly in SLE, may inhibit CRP expression.

Substantial CRP elevation in SLE patients is more likely to result from superimposed infection than from activation of lupus. CRP levels greater than 6 mg/dL in these patients should serve as an impetus to exclude the possibility of infection, just as they should in other diseases. Such levels should not be regarded as proof of infection, however; as indicated earlier, marked CRP elevation related to active SLE can be seen in the absence of infection.

Carotid plaque and intima-media wall thickness, correlates of atherosclerotic vascular disease, have been found in association with minor CRP elevation in women with SLE, as they have in patients with RA. Polymyalgia Rheumatica and Giant Cell Arteritis

The diagnosis of PMR or GCA is supported by an elevated ESR, often greater than 100 mm/hr. However, such elevation is no longer regarded as a sine qua non of these disorders; continuing reports suggest that 10% to 20% of patients with PMR can have “normal” ESRs, depending on which value is taken as the limit of normal. Such patients tend to have fewer systemic symptoms and less severe, less frequent anemia. They have the same frequency of positive temporal artery biopsy results, however, as patients with elevated ESR.

Only about 5% of patients with GCA had ESR values less than 40 mm/hr; these patients had fewer visual and systemic symptoms than patients with high ESR values. In contrast to these findings, ESR and CRP were found to be significantly lower in patients with ocular involvement, most commonly in the range of 70 mm/hr to 100 mm/hr. Patients with ESR greater than 100 mm/hr had decreased incidence of visual ischemic events.

In PMR and GCA, CRP and ESR have been regarded in the past as equally valuable in assessing disease activity. However, recent reports suggest that CRP is more sensitive for both conditions and should be included routinely in the diagnostic workup. The report that IL-6 is more sensitive than ESR for indicating disease activity in GCA is of particular interest. Subsets of patients with PMR who have been noted to have persistently elevated levels of CRP and IL-6 despite corticosteroid treatment have been shown to have a higher risk of relapse. A polymorphism at the IL-6 gene promoter has characterized these PMR patients with persistently elevated levels of the cytokine. Clinical manifestations of disease, even in the presence of a normal ESR or CRP level, should not be ignored. Extreme elevation of the ESR in the absence of symptoms of PMR or GCA should raise suspicion of other disorders, such as infection, malignancy, or renal disease.

Numerous markers of endothelial perturbation, although not acute phase reactants in the strict sense, are elevated in plasma in various inflammatory disorders of vessels, particularly PMR, GCA, and other vasculitides. These molecules include von Willebrand factor, thrombomodulin, some vasoactive prostanoids, and a variety of adhesion molecules, such as vascular cell adhesion molecule-1.

Adult-Onset Still’s Disease

Markedly elevated concentrations of ferritin, disproportionately high compared with those of other acute phase reactants, have long been noted in adult-onset Still’s disease but are not specific. Only a low percentage—commonly less than 20%—of ferritin is glycosylated in adult-onset Still’s disease—a criterion included in recently proposed classification criteria for this condition. Extremely elevated ferritin levels have been found in macrophage activation syndrome, and 40% of individuals with this condition meet the criteria for adult-onset Still’s disease, suggesting that many of the two disorders are not distinct entities. Concentrations of serum IL-18 were extremely elevated in patients with active adult-onset Still’s disease compared with those of patients with other connective tissue diseases or of healthy individuals and were correlated with serum ferritin values and disease severity. The cytokine profile in the sera of patients suggests a type 1 T helper cell (Th1) response, with significantly higher levels of TNF, IL-6, and IL-8, in addition to IL-18. The role of IL-1 has been
inferred from significant improvement in disease activity by the IL-1 receptor antagonist, anakinra. It has been suggested that interferon-α may be responsible for the hyperferritinemia of adult-onset Still's disease. CRP levels are usually markedly elevated in this disease.

**Ankylosing Spondylitis**

Ankylosing spondylitis ordinarily does not lead to a substantial increase in ESR or CRP. Median ESR and CRP levels are 13 mm/hr and 1.6 mg/dL, respectively, in patients with only spinal involvement, and 21 mm/hr and 2.5 mg/dL in patients with peripheral involvement or associated inflammatory bowel disease. Treatment with infliximab led to an average decrease of 75% in CRP concentration after 12 weeks. Patients with lower CRP levels showed little improvement, however, raising the possibility that patients with higher CRP values show better responses to anti-TNF treatment than do patients with lower CRP levels.

High-sensitivity CRP may better correlate with clinical disease activity compared with standard CRP testing. It has been reported that levels of IL-8, IL-17, and IL-23 are elevated in the serum of patients with active ankylosing spondylitis, and polymorphisms in the IL-23 receptor gene are associated with the disease.

**Osteoarthritis**

Minor CRP elevations of 0.3 to 1.0 mg/dL have been reported in patients with osteoarthritis, particularly those with progressive joint damage. However, no evidence supports this association independent of body mass index (BMI), because obesity is a common accompaniment of osteoarthritis. Although local inflammation likely plays a role in the pathogenesis of osteoarthritis, systemic inflammation likely does not. However, CRP levels are higher in patients with erosive osteoarthritis of the hand than in those with nonerosive osteoarthritis.

**Other Rheumatic Diseases**

Acute phase markers are elevated in numerous rheumatic diseases as the inflammatory cascade commences. Elevated ESR and CRP can be found in systemic vasculitides, crystal arthropathies, psoriatic and reactive arthritides, and infectious joint diseases. Monitoring for elevated SAA levels has been proposed as a tool for diagnosis and medication adjustment in familial Mediterranean fever. CRP is often normal in patients with primary Sjögren’s syndrome, and those with elevated responses do not differ clinically from patients with normal levels. Oligoarticular-onset juvenile idiopathic arthritis is classically considered as not associated with elevated inflammatory markers, although they are present in patients most at risk for developing systemic disease.

**C-REACTIVE PROTEIN AND HEALTH: ASSOCIATIONS WITH NONRHEUMATOLOGIC CONDITIONS**

Although most ostensibly healthy individuals have CRP concentrations of 0.3 mg/dL or less, some have concentrations greater than 1 mg/dL. Such minor CRP elevation has long been attributed to trivial tissue injury or to minimal inflammatory processes, such as gingivitis. However, recent data indicate that CRP concentrations between 0.3 and 1 mg/dL have clinical relevance. This has led to an explosion of published literature measuring CRP levels in cardiac, neurologic, neoplastic, pulmonary, and even psychiatric disease.

The foundation of these investigations is the well-established notion that when a high-sensitivity CRP assay is used, serum CRP levels greater than 0.3 mg/dL indicate increased relative risk of atherogenesis and future myocardial infarction. The statistical strength of this association has been shown to be as robust as, but not more robust than, established risk factors such as hypertension, diabetes, and hypercholesterolemia. The primary unanswered questions remain: Why does CRP predict myocardial infarction? Is it a pathogenic mediator itself? The observation that CRP is associated with many “noninflammatory” conditions, such as low levels of physical activity, low intake of fruits and vegetables, a variety of other “unhealthy” diets, smoking, hypertension, obesity, sleep deprivation, and low alcohol intake, indicates that classic inflammation does not invariably underlie a

**PRACTICAL USE OF ACUTE PHASE REACTANTS**

When surveyed in the 1990s, rheumatologists were found to use the ESR more than twice as frequently as they did CRP levels, although ESR reflects many complex, poorly understood changes in the physical and chemical characteristics of blood not associated with inflammation. As indicated earlier, the reference normal values for ESR are unclear. It is well established that mean ESR values increase substantially with age and differ between men and women. Difficulties associated with interpretation of the ESR and an increasingly positive clinical experience with CRP suggest that rheumatologists may benefit from relying more on CRP than ESR testing. No single ideal test can be used to evaluate the acute phase response, however. The relative virtues of these tests in following patients with RA have been discussed.

Discrepancies between ESR and CRP may result from effects of blood constituents that are not related to inflammation, but that can influence the ESR, as was discussed earlier. In addition, patterns of acute phase protein changes differ in different conditions. ESR may be markedly elevated in many patients with active SLE, whereas CRP is normal. Undoubtedly, numerous other clinical situations exist in which similar discrepancies occur. Although falsely high ESR has many noninflammatory physicochemical causes (some known, and many unknown), CRP values greater than 1 mg/dL almost invariably reflect a clinically significant inflammatory process. In light of these considerations, many authors believe that several tests, rather than a single test, should be performed and interpreted in their clinical context. It has been suggested that ESR, which is associated with anemia and immunoglobulin levels, may reflect “general severity” in RA, whereas CRP is a better test of active inflammation per se.
CRP response. Many of these CRP-associated conditions are known to be risk factors for cardiovascular disease, confounding the situation and suggesting that CRP predicts myocardial infarction because of its association with these risk factors. In addition, elevated CRP may not reflect inflammation, but rather a response to the presence of distressed, metabolically disturbed cells. This reverse causation offers a potential explanation of the laboratory result, because atherosclerosis might trigger an elevation in CRP.

Nevertheless, CRP has been known for many years to bind to low-density lipoprotein (LDL) and to activate complement—a potentially proinflammatory response; it has also been detected in atherosclerotic plaques. These observations have raised the possibility that CRP may play a direct causal role in coronary heart disease. Proatherogenic effects secondary to CRP injections given to mice, however, may have been caused by contaminants in commercial CRP preparations, not by CRP itself. Because statin drugs lower CRP and LDL, some consider their effects to provide indirect evidence of a causal role of CRP. In the recently published JUPITER trial, patients with normal LDL and elevated CRP were prescribed rosuvastatin 20 mg daily or placebo and were followed prospectively. Those in the treatment arm had significant reductions in major cardiovascular events. However, whether the effects of the study can be solely dependent on CRP reduction remains unclear; alternative explanations have been put forth, including the question of the importance of lowering LDL even among patients with “normal” levels.

Whether to target CRP levels in cardiovascular disease is an ongoing topic of intense debate.

Because no “CRP inhibitor” drugs are available to directly measure the effect of lowering the protein, recent studies have centered on observing patients with genetic variations that result in different baseline levels of CRP. To date, results have been conflicting with regard to the role that genetically determined CRP levels play in coronary heart disease. However, additional, more robust analyses are ongoing.

Epidemiologic studies describing an association of CRP levels with morbidity and mortality in many chronic diseases and even in normal aging have become a “cottage industry.” It is important to remember that these are observational population studies. Although such associations may have broad and intriguing implications, particularly at a societal level, they reflect probabilities; this limits their clinical value when they are applied to the individual patient.

Selected References


*Full references for this chapter can be found on* www.expertconsult.com.
References

ACUTE PHASE REACTANTS AND THE CONCEPT OF INFLAMMATION

829.e3


159. De Torres JP, Pinto-Pata V, Casanova C: C-reactive protein levels and survival in patients with moderate to very severe COPD, Chest 133:1336–1343, 2008.


For decades, imaging in rheumatology has been synonymous with conventional radiography (CR). However, new imaging modalities such as magnetic resonance imaging (MRI) and ultrasonography (US) have dramatically increased the amount and scope of information obtainable by imaging. In rheumatology, imaging may be used for multiple reasons that include establishing or confirming the diagnosis, determining extent of disease, monitoring change in disease (e.g., activity, structural damage), selecting patients for specific therapies (e.g., surgery, injections), identifying complications of disease or treatment, and assessing therapeutic efficacy in trials. These entirely different contexts may favor different imaging approaches.

The present chapter focuses on inflammatory joint diseases, such as rheumatoid arthritis (RA), psoriatic arthritis (PsA), ankylosing spondylitis (AS), other types of axial spondyloarthritis (SpA), gout, and osteoarthritis (OA). Other aspects, including nuclear medicine and capillaroscopy in connective tissue diseases and vasculitides, are also briefly discussed. The reader is kindly referred to the chapters on individual diseases for imaging aspects of other rheumatologic diseases, and to textbooks of musculoskeletal radiology for a more detailed description of the different imaging modalities, including the technical aspects. This chapter outlines the virtues of CR and describes its current major importance in diagnosis and follow-up of rheumatologic diseases, but also, by putting emphasis on newer imaging modalities, particularly MRI and US, stresses that rheumatology has entered a time of exciting and expanding therapeutic and imaging possibilities.

**CONVENTIONAL RADIOGRAPHY**

**KEY POINTS**

X-ray is relatively inexpensive, easily available, and reliable.

X-ray provides information on bone damage and, indirectly through joint space narrowing, on cartilage damage, whereas x-ray is neither sensitive nor specific for soft tissue change.

X-ray findings are important parts of the classification criteria for many rheumatic diseases, including RA, AS, SpA, PsA, and OA.

X-ray often should be the first imaging investigation in arthritis because the significance of positive and negative findings is well known.

X-ray can be used to follow structural damage progression in inflammatory and degenerative joint diseases but is less sensitive to change than MRI.

The main disadvantage of x-ray is its low sensitivity, particularly for soft tissue changes.

---

**KEY POINTS**

**Practical Use of Imaging in Inflammatory Joint Diseases**

**A. Peripheral joints**

*Use in clinical practice:*
- To establish a diagnosis of rheumatoid arthritis (RA): x-ray, magnetic resonance imaging (MRI), ultrasonography (US)
- To assist with the diagnostic workup in suspected, but not definite, inflammatory joint disease and early, unclassified inflammatory joint disease (by detection of presence/absence of synovitis, enthesitis, bone erosions etc.): x-ray, MRI, US
- To monitor disease activity: MRI, US
- To monitor structural joint damage: x-ray, MRI
- To assist with prognostic stratification of patients with early RA: x-ray, MRI, (US*), digital x-ray radiogrammetry (DXR*)
- To help define the presence or absence of imaging remission: MRI, US
- To guide aspirations and injections in joints, bursae, and tendon sheaths: US

*Use in clinical research:*
- To assess structural joint damage in RA trials: x-ray, MRI
- To assess the anti-inflammatory effectiveness of a new compound MRI, US
- For pretrial selection of patients most likely to progress ("enrichment"): x-ray, MRI

**B. Axial joints**

*Use in clinical practice:*
- To establish a diagnosis of ankylosing spondylitis (AS)/axial spondyloarthritis (SpA): x-ray and MRI
- To monitor disease activity: MRI
- To monitor structural joint damage: x-ray, MRI, computed tomography (CT*)

*Use in research:*
- To assess structural progression AS/SpA trials: x-ray, MRI
- To assess the anti-inflammatory effectiveness of a new compound: MRI
- For pretrial selection of patients most likely to progress: x-ray, MRI

*Promising, but more data needed.*

*CT allows this but cannot be used owing to radiation exposure.*
The first roentgen ray image was an “x-ray” of the hand. Since then, imaging of musculoskeletal structures has always been an important role of conventional radiography (x-rays). The simple radiograph is relatively cheap, is available worldwide, and produces images that are almost identical regardless of technical parameters or whether the image is analogue or digital. The reliability of the image means that, despite the limitations of radiography, advances in clinical practice can rely on the good old x-ray, with the knowledge that it is largely independent of technologic advances. For many years to come, radiographs will appear much the same as they do now.

**Technical Aspects**

The conventional radiograph is a two-dimensional summation image that is dependent on variable absorption of x-rays by different tissues for its inherent contrast. It has a very high spatial resolution that is rarely surpassed by other modalities, but radiography offers high contrast only between a limited number of structures, namely, bone (calcium), soft tissue, and air. Fat is visible as a separate density, but the distinction between soft tissue and fat is often subtle, and radiography cannot distinguish between the other soft tissues because cartilage, muscle, tendon, ligament, synovium, and fluid all appear at the same density. These characteristics give the radiograph its inherent advantages and disadvantages.

In its favor, the x-ray shows skeletal structures very well, and because it is a summation image, it allows excellent overall assessment of skeletal trauma and alignment. The limited number of images produced facilitates rapid review, and the high bone–soft tissue contrast often produces radiographic manifestations of disease in specific patterns that make the test particularly useful in daily clinical practice. The biggest disadvantage of radiography is its inherent lack of soft tissue contrast, which makes it insensitive for the detection of soft tissue abnormalities. The first radiographic sign of inflammatory arthritis may be permanent structural damage (bony erosion); in degenerating joints, cartilage damage is not usually visualized until sufficient overall cartilage loss allows approximation of the bone ends, resulting in “joint space narrowing.”

Although x-rays use ionizing radiation, they should be regarded as relatively safe, especially in older patients. An exception may be spine and sacroiliac joint radiography, for which doses are higher to penetrate the trunk; in a younger population, MRI offers a safer and more informative alternative. In most examinations, two or more projections are required to adequately visualize the joint in question, and radiographic quality is improved by strict adherence to standard imaging protocols.

**Rheumatoid Arthritis**

Rheumatoid arthritis (RA) is the archetypal inflammatory joint disease that primarily targets synovium and peripheral joints. It is a systemic inflammatory disorder in which the typical clinical manifestations are usually symmetric, and the radiographic signs usually follow this pattern. As in all of radiology, the observed distribution of disease is often characteristic of the underlying cause; in RA, the metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints of the hands, the wrists, and the metatarsophalangeal (MTP) joints of the feet are most often involved.

Juxta-articular osteoporosis is a characteristic feature of RA best seen in early disease in small peripheral joints. Later, generalized osteoporosis is usually present and is exacerbated by disuse. Erosion of bone (Figure 58-1) is a characteristic feature of RA and usually appears first at the margins of the joint, where the proliferating synovium lies directly on the surface of the bone between the edge of the cartilage-covered articular surface and the capsular attachment—the “bare areas.” These marginal erosions may be subtle and first appear as disruption of the thin white cortical line, especially at the radial aspect of the metacarpal heads, where they may be seen best on dedicated radiographic projections such as the ball-catcher’s view of the hands. These erosive changes are a good indication of the aggressiveness of the arthritis. In large joints, synovial proliferation may be very severe before bony erosion is detectable on x-ray. This is especially notable in the knee, where pain and swelling due to arthritis and bursitis occur long before erosion. In addition to bare-area erosions, two other types of erosions have been described in RA. Compressive erosion refers to remodeling of inflamed and osteoporotic bone with gradual invagination of one bone into another, typified by protrusio acetabuli of the hip (Figure 58-1E). Surface erosion of bone may also be seen, usually resulting from inflammation of an adjacent tendon sheath. A typical location for this is at the outer margin of the ulnar styloid process secondary to extensor carpi ulnaris tenosynovitis.

Joint space involvement is characteristic of RA. In many cases, inflammatory processes result in progressive destruction of articular cartilage, which in turn causes the radiographic finding of concentric joint space narrowing. This is usually diffuse in RA because all cartilage is involved at the same time; this feature may allow differentiation from the focal or asymmetric type of joint space loss that occurs with degenerative disorders. Occasionally cartilage destruction precedes synovial erosion of bone. Continued cartilage damage may result in partial or complete fibrous ankylosis, but progression to bony ankylosis is uncommon, although it may occur in end-stage disease in the wrist or midfoot. Subchondral radiolucent areas are common in RA and are often referred to as cysts, geodes, or pseudocysts. They may develop as a result of intraosseous extension of pannus, injury to the bone of any kind, or true intraosseous rheumatoid nodules. Mechanical factors accentuate their development, and very large cystic lesions may be seen in the elbow, femoral neck, or knee, occasionally precipitating pathologic fracture.

Symmetric soft tissue swelling around the small joints of the hands or feet is often the first clinical and radiographic sign of RA. Soft tissue swelling is caused by joint effusion, synovial proliferation, and periarticular inflammation, but the radiographic findings are nonspecific, and x-rays are much less sensitive for such changes than MRI and US. Eccentric soft tissue swelling may be due to adjacent bursitis, tenosynovitis, or rheumatoid nodules. Involvement of tendons and tendon sheaths is common; however, soft tissue changes usually are poorly visualized and are of lesser diagnostic importance.
PART 7 | DIAGNOSTIC TESTS AND PROCEDURES IN RHEUMATIC DISEASES

DMARD therapy has been shown to reduce cervical radiologic progression. The upper cervical region is most affected, although abnormalities throughout the cervical spine are frequent. Upper cervical changes include abnormalities of apophyseal and discovertebral joints, subluxation, and dislocation.

Adequate radiographic evaluation of the cervical spine requires lateral views in flexion and extension and should be performed in all RA patients with neck pain. Flexion-extension views are especially important for demonstration of the degree of atlantoaxial instability.

Spine

Cervical spine involvement is common in RA and merits specific attention. It occurs mainly after several years of disease in rheumatoid factor–positive patients with severe peripheral RA and may cause severe pain, instability, and, ultimately, spinal cord compression. Prevalences of radiologic cervical involvement up to 70% have been reported but are much lower in recent studies in accordance with the fact that intensive disease-modifying antirheumatic drug (DMARD) therapy has been shown to reduce cervical radiologic progression.

The upper cervical region is most affected, although abnormalities throughout the cervical spine are frequent (Figure 58-2). Upper cervical changes include odontoid process (dens) erosion and atlantoaxial subluxations (anterior, vertical [known as cranial settling], lateral, and posterior); subaxial manifestations include abnormalities of apophyseal and discovertebral joints, subluxation, and dislocation.

Adequate radiographic evaluation of the cervical spine requires lateral views in flexion and extension and should be performed in all RA patients with neck pain. Flexion-extension views are especially important for demonstration of the degree of atlantoaxial instability.

Atlantoaxial subluxation is caused by laxity or rupture of the transverse ligament. The characteristic finding is anterior subluxation (i.e., abnormal separation between the anterior arch of the atlas and the odontoid process of the
CHAPTER 58 | IMAGING MODALITIES IN RHEUMATIC DISEASES

Other joint conditions such as osteoarthritis, psoriatic arthritis, and neoplasm.

Monitoring. In routine clinical management and in clinical trials of RA, x-ray evaluation focuses on joint space narrowing and bone erosion in hands, wrists, and forefeet as measures of structural joint damage. Validated scoring methods for radiologic damage (the Larsen method and the Sharp method and their modifications) are available and are used extensively in clinical trials. The van der Heijde and Genant modifications of the Sharp score are generally considered the methods that are most sensitive to change, but they are also the most time-consuming. For clinical practice, the less time-consuming simple erosion narrowing score (SENS), derived by counting joints with bone erosion and joints with joint space narrowing, is available but is seldom used.

Prognostication. Early bone erosion correlates with poor long-term radiographic and functional outcomes, and early progression in x-ray erosion is related to future impairment in physical function. In early undifferentiated arthritis, the presence of x-ray erosion increases the risk of developing persistent arthritis. However, radiographic erosions are present in only a minority of patients with early RA, with prevalences of 8% to 40% at 6 months, and the absence of erosion in early disease is not necessarily associated with good outcomes. Therefore at this stage, x-ray is not effective for identifying future “nonprogressors” (i.e., patients who will not show increasing structural joint damage).

axis). Reports of the exact lower limit of abnormal distance vary, but a distance greater than 2.5 mm is considered abnormal, as is distance that changes significantly between flexion and extension. Cranial settling is diagnosed by radiographic measurement of the relationship between the tip of the odontoid process and landmarks at the base of the skull. It may lead to protrusion of the dens through the foramen magnum and compression of neurologic structures. Significant atlantoaxial subluxations may require surgery.

Subaxial involvement is frequent and manifests as varying degrees of bone erosion, joint space narrowing, and subluxation of facet joints, as well as disk space narrowing and spinous process erosion. Subaxial subluxations may also lead to neurologic deficits.

MRI is useful for providing supplementary information on cervical spine involvement (Figure 58-2C), including spinal cord compression (see MRI section).

Use in Diagnosis, Monitoring, and Prognostication

Diagnosis. Characteristic x-ray findings are part of the American College of Rheumatology (ACR) 1987 classification criteria for RA. X-ray also has a role in the recent ACR/European League Against Rheumatism (EULAR) 2010 classification criteria, in that patients who display bone erosions typical of RA plus at least one clinically swollen joint fulfill the criteria for and are classified as having RA. X-ray can be helpful in differentiating RA from other joint conditions such as osteoarthritis, psoriatic arthritis, and neoplasm.

Monitoring. In routine clinical management and in clinical trials of RA, x-ray evaluation focuses on joint space narrowing and bone erosion in hands, wrists, and forefeet as measures of structural joint damage. Validated scoring methods for radiologic damage (the Larsen method and the Sharp method and their modifications) are available and are used extensively in clinical trials. The van der Heijde and Genant modifications of the Sharp score are generally considered the methods that are most sensitive to change, but they are also the most time-consuming. For clinical practice, the less time-consuming simple erosion narrowing score (SENS), derived by counting joints with bone erosion and joints with joint space narrowing, is available but is seldom used.

Prognostication. Early bone erosion correlates with poor long-term radiographic and functional outcomes, and early progression in x-ray erosion is related to future impairment in physical function. In early undifferentiated arthritis, the presence of x-ray erosion increases the risk of developing persistent arthritis. However, radiographic erosions are present in only a minority of patients with early RA, with prevalences of 8% to 40% at 6 months, and the absence of erosion in early disease is not necessarily associated with good outcomes. Therefore at this stage, x-ray is not effective for identifying future “nonprogressors” (i.e., patients who will not show increasing structural joint damage).
**Ankylosing Spondylitis**

Ankylosing spondylitis (AS) is the archetypal inflammatory joint disease that primarily targets fibrocartilage and the axial skeleton. Cartilaginous joints and sites of enthesis (tendon and ligament insertion) are involved early, and the characteristic involvement of the axial skeleton includes a predilection for the sacroiliac joints and all articulations and entheses of the spine.

**Sacroiliac Joints**

Erosion and ankylosis of the sacroiliac (SI) joints are the hallmarks of spondyloarthritis. Sacroiliitis is usually the first manifestation and is characteristically bilateral and symmetric in AS. Early radiographic findings predominate on the iliac side of the cartilage compartment, with erosion of subchondral bone causing loss of definition of the articular surfaces, usually accompanied by variable degrees of adjacent osteoporosis and surrounding reactive sclerosis. Bone erosion may result in the radiographic observation of focal joint space widening (Figure 58-3A), and as the disease progresses, the definition of the joint is completely lost with radiographic superimposition of erosion, sclerosis, and new bone formation, which fills in the erosions and the original cartilaginous “joint space.” The joint may disappear completely in late disease with ankylosis and remodeling of the bone (Figure 58-3B). The ligamentous compartment of the SI joint is frequently affected by bony erosion and enthesal proliferation, although these may be difficult to see radiographically.

**Spine**

Traditional descriptions of the initial sites of spinal involvement are enormously influenced by the inability of radiography to adequately visualize many parts of the spine. Radiographic reports indicate that the lumbosacral and thoracolumbar junctions are first affected, whereas the MRI literature clearly indicates a predilection for early involvement of the midthoracic spine—an area that is extremely difficult to evaluate radiographically. The cervical spine is rarely affected first, but this can occur occasionally in women, and spinal disease rarely occurs in the absence of significant SI joint involvement.

Early radiographic manifestations of AS in the spine are most often due to enthesitis at the edges of the discovertebral joints. Focal sclerosis (the “shiny corner”) and erosion (the “Romanus lesion”) develop at the attachment of the annulus fibrosus to the anterior corner of the vertebral end plate and are characteristic features of early AS (Figure 58-4A). The anterior borders of the vertebrae may appear straight or “squared” owing to periosteal proliferation of new bone filling in the normal concavity or erosion at the anterosuperior and anteroinferior vertebral margins. This observation is much easier to make in the lumbar spine, where normal vertebrae are always concave anteriorly in comparison with the thoracic and cervical spine, where the normal contour is much more variable and may be square or occasionally convex.

The hallmark of spinal disease in AS is the development of characteristic bony spurs known as syndesmophytes (see Figure 58-3C). These start as thin, vertically oriented

---

**Figure 58-3** Ankylosing spondylitis (x-ray). **A,** Sclerosis (arrows) is seen along the iliac sides of the sacroiliac joints, along with loss of portions of the iliac subchondral bone indicating erosion. **B,** Bilateral hip joint space narrowing is present. A ring of osteophytes is noted at the synovial insertion (arrows) on each femoral head. The sacroiliac joints are fused (ankylosis). **C,** Syndesmophytes (black and white arrows), some of which extend from the edge of one vertebral body to the next (bridging syndesmophytes, ankylosis; black arrows), are seen in this lateral view of the spine.
Ankylosing spondylitis (x-ray and magnetic resonance imaging (mRI)), Figure 58-4C and D, and has a horizontally oriented segment of variable length at the discovertebral junction, is typically triangular in shape, and has projections of bone that develop as the result of ossification within the outer fibers of the annulus fibrosus of the intervertebral disk. Syndesmophytes are radiographically visible on the anterior and lateral aspects of the spine, starting from the anterosuperior corners of the vertebral bodies (arrows), corresponding to early osteitis. C and D, Fracture in ankylosing spondylitis. C, X-ray shows disruption of the previously fused C6-C7 facet joints (arrow) and slight anterior subluxation. D, Sagittal T2-weighted MR image confirms a high signal fracture line (arrow) through the superior aspect of C7.

Use in Diagnosis, Monitoring, and Prognostication

Although the modified New York criteria are actually classification criteria, they are the most commonly used criteria for the diagnosis of AS and are based on clinical features and radiographic sacroilitis. According to these criteria, AS may be diagnosed if, in addition to the presence of one clinical criterion, grade 2 sacroilitis (minimal sacroilitis: loss of definition of the joint margins, minimal sclerosis, joint space narrowing, and erosions) or higher occurs bilaterally, or grade 3 (moderate sacroilitis: definite sclerosis on both sides of the joint, erosions, and loss of joint space) or grade 4 (complete bony ankylosis) occurs unilaterally. Owing to the requirement for these radiographic structural changes, the median duration of disease before diagnosis has been 7 to 10 years. Definitions of radiographic changes according to the New York criteria are included in the Assessment of Spondyloarthritis International Society (ASAS) classification criteria for axial spondyloarthritis, the European Spondyloarthritis Study Group (ESSG) criteria for spondyloarthritis, and the modified New York criteria.

Radiography of the spine is not included in the classification criteria but may be useful for following structural disease progression in patients with spinal involvement. Bone changes seen in patients with axial SpA develop slowly and often are not present in patients with early disease; generally, only minor changes can be observed in 1 to 2 years. Different scoring methods, all based on assessment of lateral views, have been developed to quantify changes in the spine of patients with AS—the Stoke AS Spine Score (SASSS), the Bath AS Radiology Index (BASRI), and the modified Stoke AS Spine Score (mSASSS). A comparative study of these three methods concluded that all measures were reliable, but that mSASSS was more sensitive to change. These spine scores are used primarily in clinical research.
evidence of joint damage on radiographs within 2 years of presentation. The hand and the wrist are most often involved and are affected in up to three-quarters of patients with PsA, but the pattern of joint involvement varies from patient to patient and over time in the individual patient. Radiographic appearances are usually asymmetric and may have a ray distribution, with affected joints involving a single digit, including the distal interphalangeal (DIP) joints (Figure 58-5), which are seldom involved in RA. Clinically, this manifests as dactylitis and is a common early presentation of PsA. Osteoporosis is frequently absent, and a distinguishing feature of PsA is the propensity for bone proliferation. This may be seen in the form of periostitis of the shafts of the phalanges, which is sometimes the first radiographic manifestation, or irregular bony spurring at joints or entheses. Fluffy new bone formation adjacent to marginal joint erosions may result in a particularly characteristic “whiskering” appearance. Severe marginal erosion of the heads of metacarpals or phalanges may produce the appearance of a whittled pencil; if combined with deep central erosion of phalangeal bases, this may be referred to as the “pencil-in-cup” appearance. Ankylosis of joints occurs frequently in advanced disease. Large joints are less often affected, and the findings are usually similar to those of RA. Spinal involvement is frequent in PsA. It may be seen in early disease, and sacroiliitis may be demonstrable in up to 75% of patients. Changes may be extensive and are more likely to be asymmetric than in AS.

Use in Diagnosis, Monitoring, and Prognostication

Although RA is characterized mainly by osteodestructive lesions, in PsA osteodestructive and osteoproliferative manifestations may coexist not only in the same patient, but even in the same joint. In particular, osteoproliferative lesions on radiography are characteristic and are included in the new classification criteria (CASPAR) for PsA. The presence of juxta-articular new bone formation, which appears as ill-defined ossification near joint margins (but excluding osteophytes) on radiographs of the hand or foot, is one of five criteria.

Structural joint damage on conventional radiography is an important outcome measure in PsA. Different radiographic scoring methods of peripheral joints have been developed (e.g., the Sharp-van der Heijde modified scoring method for PsA, which is a detailed scoring system for evaluating erosion and joint space narrowing), and osteolysis and pencil-in-cup phenomena are assessed separately. Scoring systems are used primarily in clinical trials. No specific scoring systems are available for spine and sacroiliac joints in PsA; these areas can be monitored as in AS (see later).

Gout

A variety of microcrystals can deposit in and around the joints, inducing an inflammatory response. In this chapter, the characteristics of gout (i.e., arthritis related to monosodium urate deposition) and of calcium pyrophosphate dihydrate (CPPD) crystal deposition disease will be presented. The reader is referred to chapters on individual diseases for descriptions of the remaining conditions, such as calcium hydroxyapatite crystal deposition disease, hemochromatosis, ochronosis, and Wilson’s disease.

Radiography is not helpful in the diagnosis of acute gouty arthritis because findings are limited to the soft tissues and are nonspecific. Chronic tophaceous gout is an asymmetric arthritis that frequently affects the feet, hands, wrists, elbows, and knees. The most common site of involvement is the first metatarsophalangeal (MTP) joint (Figure 58-6).
Use in Diagnosis, Monitoring, and Prognostication

The 1977 Criteria for the Classification of Acute Arthritis of Primary Gout include the conventional radiographic features of asymmetric swelling and subcortical cysts without erosion. However, conventional radiographic features of acute gout have low clinical utility owing to lack of sensitivity and specificity. In the chronic phase, the characteristic pattern with asymmetric, erosive polyarticular disease and tophi appearing as nodular soft tissue masses with amorphous increased density or patchy calcification may be diagnostically useful.

A specific gout radiographic scoring method has been developed and validated and may improve sensitivity to change in longitudinal studies. However, radiography is

Radiologic changes in long-standing gout are related to tophi and their effects on soft tissue and bone. Tophi appear as eccentric, nodular soft tissue masses around the joints with an amorphous increased density or patchy calcification. Intraosseous tophi cause well-margined subchondral cystic lucencies or intraosseous calcifications. Erosion of adjacent bone is common at the periosteal or endosteal surface. Erosions may be intra-articular, para-articular, or remote from the joint and typically are well defined with smooth sclerotic borders. An overhanging margin of bone at the edge of the erosion is a characteristic feature of gout. The joint space and bone density are usually preserved until late in the disease. Bursal inflammation commonly produces soft tissue swelling around the olecranon and in a prepatellar location.

Figure 58-6 Tophaceous gout (x-ray, computed tomography [CT], and magnetic resonance imaging [MRI]). A, X-ray of forefoot in tophaceous gout. Extensive bone destruction is seen at the first metatarsophalangeal (MTP) joint with overhanging edges (arrowhead) and soft tissue swelling. Smaller erosions are present involving the first tarsometatarsal and second MTP joints (arrows). B, CT scan shows hyperdense tophus around the extensor digitorum muscle (arrow), with smaller deposits around the other extensor tendons. Axial precontrast (C) and coronal short tau inversion recovery (D) MR images show a large tophus (arrow) medial to the first metatarsal bone. Ca, calcaneus; Na, navicular. (C and D courtesy Professor Fiona McQueen, Auckland, New Zealand.)
not a sensitive method for monitoring gout manifestations, compared with US and MRI, because inflammatory changes and tophi volumes cannot be assessed.

**Calcium Pyrophosphate Dihydrate Crystal Deposition Disease**

The term CPPD crystal deposition disease refers to a variety of clinical situations, including asymptomatic crystal deposition, pseudogout, pyrophosphate arthropathy, and other, less frequent presentations, such as pseudo-RA. Chondrocalcinosis is a general term that refers to calcification of hyaline or fibrocartilage regardless of its origin (Figure 58-7).

An acute attack of pseudogout is often a monoarthritis, and typical radiographic findings of soft tissue swelling and joint effusion are nonspecific, mimicking gout and infection. Often chondrocalcinosis is not radiographically visible. X-ray changes in more chronic pyrophosphate arthropathy are very similar to those in OA, with joint space narrowing, sclerosis, and subchondral cysts, with or without intra-articular or periarticular calcinosis.

The patellofemoral, radiocarpal, and second and third MCP joints are typical sites of involvement in pyrophosphate arthropathy. Distribution is usually bilateral and may be symmetric. In addition to the usual target sites, arthritis changes may be observed in non–weight-bearing joints and sacroiliac joints.

Pyrophosphate arthropathy may be associated with extensive and rapid subchondral bone collapse and fragmentation with intra-articular loose bodies, resembling neuropathic osteoarthropathy (pseudo-Charcot). Tumorous CPPD crystal deposits that resemble gouty tophi are observed occasionally, most frequently in digits, and are referred to as tophaceous pseudogout.

**Septic Arthritis**

Septic arthritis usually is monoarticular and typically is associated with rapid onset of symptoms, systemic illness, obvious local clinical signs, and laboratory evidence of acute inflammation. The earliest clinical sign is symmetric soft tissue swelling around a joint due to soft tissue edema, hypertrophied synovium, and joint effusion. However, this may be difficult to discern on x-ray, and periarticular osteopenia may be the first radiographic feature. Initial widening of the joint space may be seen, as in septic arthritis of the pediatric hip. Marginal erosions develop quickly and have a very similar appearance to other forms of erosive inflammatory arthritis. As the infection progresses, hyaline cartilage and the subchondral cortex are rapidly destroyed, with progressive narrowing of the joint space (Figure 58-8). In late stages, ankylosis of the joint may be seen. Adjacent osteomyelitis occurs with increasing frequency as the infection becomes chronic. With granulomatous infection, the characteristic triad (Phemister’s triad) of radiographic findings consists of marked osteoporosis, marginal erosions, and absent or mild joint space narrowing. Periostitis and bone production are less common.

**Use in Diagnosis, Monitoring, and Prognostication**

Radiography should be performed routinely but is insensitive for diagnosis. When clinical suspicion of septic arthritis arises, joint aspiration should be performed without delay. The utility of radiography is limited in septic arthritis, and immediate joint aspiration (directly or by image-guided techniques) is required for early diagnosis; more advanced imaging techniques are required for assessment of associated osteomyelitis or other complications. However, radiography still should be routinely performed because it may assist in interpretation of other studies and offers a satisfactory overall assessment of bone morphology and background disease.

**Osteoarthritis**

Osteoarthritis (OA) primarily targets hyaline cartilage, and its distribution is profoundly influenced by mechanical
projection, particularly in the lower limbs. Sensitive detection of cartilage thinning requires apposition of the most affected surfaces, which requires the joint to be optimally positioned with sufficient mechanical stress to bring the surfaces into contact. In the knee, for example, angular deformity tends to reduce on supine radiography, and severe cartilage loss may not be detected at all (see Figure 58-9).

Bilateral weight-bearing projections are more consistent for the detection and measurement of cartilage loss, and the erect extended anteroposterior (AP) view has been established as the preferred projection for diagnostic utility. However, no view is completely reliable until global loss of cartilage occurs in late disease. Weight bearing in flexion is factors. OA is typically asymmetric in distribution both within and between joints. The most characteristic sites of osteoarthritis include hips, knees (Figure 58-9), proximal and distal interphalangeal joints of the hand, first carpometacarpal and trapeziocapitate joints of the wrist (Figure 58-10), and first metatarsophalangeal joints.

Joint space narrowing due to thinning of hyaline cartilage is a key feature of OA. Unlike the inflammatory arthropathies, resultant narrowing of the radiographic joint space is asymmetric, with cartilage thinning more pronounced in areas subject to greater mechanical stress, such as one compartment of the knee. Accurate assessment of cartilage loss is hugely influenced by radiographic projection, particularly in the lower limbs. Sensitive detection of cartilage thinning requires apposition of the most affected surfaces, which requires the joint to be optimally positioned with sufficient mechanical stress to bring the surfaces into contact. In the knee, for example, angular deformity tends to reduce on supine radiography, and severe cartilage loss may not be detected at all (see Figure 58-9). Bilateral weight-bearing projections are more consistent for the detection and measurement of cartilage loss, and the erect extended anteroposterior (AP) view has been established as the preferred projection for diagnostic utility. However, no view is completely reliable until global loss of cartilage occurs in late disease. Weight bearing in flexion is
the most sensitive projection for cartilage thinning in the knee and is most reliable for detection of change, although the ideal degree of flexion for a specific individual varies according to the distribution of cartilage loss (see Figure 58-9).\textsuperscript{43,44}

Osteophytes, the most characteristic abnormality of OA, are more specific than joint space narrowing and often are the first radiographic sign of the disease. They start as osseous metaplasia of the hyaline cartilage best seen at the articular margins but also occur centrally on the articular surface. They can become large and often dominate the radiographic appearance. Thus, osteophytes are the key criterion in defining the presence of OA.\textsuperscript{45}

Bone density is usually normal or sclerotic. Bone erosion does not occur, except at the articular surface in late disease and occasionally earlier in the interphalangeal (IP) joints in inflammatory OA. The radiographic appearance of periarticular soft tissues is relatively normal in OA. Synovitis and effusion do occur but only rarely dominate the radiographic presentation.

Some deformity is common but tends to occur in particular joints such as the thumb and the knee. The deformity is

![Figure 58-9](https://example.com/figure589.png)
Computed tomography (CT) has been available for 40 years, and each decade has brought continued development in computing and engineering, resulting in remarkable changes in the technology. Although still limited in terms of soft tissue contrast, CT offers fast and reliable acquisition, high resolution, and multiplanar capabilities, which have enhanced its use in recent years. The modern CT scanner is a remarkable tool that might have an increasingly important role in arthritis imaging in the future.

Technical Aspects

CT image acquisition is no longer restricted to the axial plane of imaging, and its multiplanar capability is so versatile that many CT scans of the body are now interpreted primarily from coronal or sagittal images in the same way as MRI. Data acquisition is so fast that patient motion is rarely a problem because CT scans are now routinely acquired in a few seconds. In fact, CT is often faster than taking multiple radiographs. Complex requirements for

Use in Diagnosis, Monitoring, and Prognostication

Conventional radiography is the standard method for diagnosis and follow-up of OA. Radiographic changes are essential in the ACR criteria for classification of OA. A patient with radiographic osteophytes, knee pain, and age above 50 years, stiffness for less than 30 minutes, or crepitus should be classified as having OA. However, newer imaging modalities, in particular MRI, are providing additional information, which can be used in clinical trials and in enhancing our understanding of the disease.

Various scoring systems have been developed to grade the severity of OA. Among the most well known is the Kellgren and Lawrence five-point scale, which has been used widely in OA research studies. Atlas-based scoring systems have been developed that provide more accurate discrimination between individual radiographic features of OA. However, whichever system is used, the reliability is variable. In addition, quantitative measurement of joint space width as a surrogate for cartilage thickness has been employed in research studies, particularly in cases of knee OA, but also in hip and hand OA.

Computed tomography (CT) is useful mainly for detection of bone abnormalities in the axial skeleton (e.g., erosion, sclerosis, new bone formation, fracture).

CT is more sensitive for these purposes than x-ray and on most occasions MRI.

The role of CT of the axial skeleton is limited by a relatively large amount of ionizing radiation.

In practice, CT is rarely used unless x-ray is unclear and MRI is unavailable.

The main disadvantages of CT are the low sensitivity for soft tissue changes and the high exposure to ionizing radiation.

KEY POINTS

CT is a tomographic radiographic imaging technique that visualizes calcified tissue with high resolution. CT is a standard reference for detecting bone destruction.
patient positioning in musculoskeletal CT are now largely irrelevant in that almost all studies can be reconstructed to eliminate positioning problems. Patient tolerance is excellent and, unlike with MRI, no absolute contraindications to CT are known. Spatial resolution is high, usually higher than that of MRI, and contrast resolution between soft tissue and bone is unsurpassed by any other modality. However, despite all the advantages just listed, applications of CT in arthritis imaging will remain somewhat limited. CT still has two fundamental flaws that prevent its universal application. First, CT is constrained in the same way as radiography by its limited soft tissue contrast capability. Second, ionizing radiation is used (with increasing dose proportional to the size of the body part and the requirements for spatial detail).

Although this is not a problem with the more distal extremities because the exposure doses are smaller and the tissue more radioresistant, it remains an issue for spine, hip, and shoulder CT. Consequently, in most routine clinical situations, radiography provides sufficient information of a similar nature as CT for clinical decision making, and x-ray is cheaper than CT and is readily available; ultrasonography is a better and cheaper way to visualize and quantify superficial soft tissue pathology, and MRI offers superior soft tissue contrast and bone marrow imaging.

Rheumatoid Arthritis

The exquisite inherent contrast between bone and soft tissue that is afforded by CT makes it a gold standard reference for the detection of bone erosion; as such, CT is ideally suited to the investigation of erosion in inflammatory arthritis. Modern CT with isotropic voxel acquisition and three-dimensional visualization allows accurate detection and quantification of bone erosion with good intraobserver agreement, whereas radiography is limited by its two-dimensional projection and superimposition of structures. However, CT is very limited in its ability to visualize soft tissue changes, and even with contrast enhancement and complex subtraction techniques, CT remains inferior to MRI and ultrasound for assessment of synovial changes such as thickening and hyperemia. Furthermore, detection of erosion on CT and MRI shows very good agreement generally, although CT is slightly more sensitive.48,49

Use in Diagnosis, Monitoring, and Prognostication

CT is not currently used in routine clinical practice for the diagnosis of RA. However, it could be useful for diagnosis in that CT appears to be the most sensitive technique for detection of erosion,48,49 and CT of the feet is very simple with a very low radiation dose. CT is used for problem solving in specific cases, such as in examination of the cervical spine if MRI is unavailable or contraindicated.

Use of CT for longitudinal assessment of damage progression has potential merit.49,50 A problem with CT in comparison with MRI or US is its inability to demonstrate improvement in soft tissue pathology. No CT data are available for prognostication in RA. Current use of CT in RA is very limited, although its ability to define erosion so clearly offers opportunities for future developments in early diagnosis and clinical trial research.

Ankylosing Spondylitis/Axial Spondyloarthritis

CT allows visualization of the same pathologic processes as conventional radiography—erosion, osteoporosis/sclerosis, and new bone formation/ankylosis—with the added benefit of multiplanar imaging free from superimposition of overlying structures (Figure 58-11).

In AS, the pathologic processes start in the bone marrow and at sites of enthesis. However, CT has poor ability to detect soft tissue change and is usually normal until structural damage is present. CT can detect osteoporosis or osteosclerosis quite well, but these changes are very nonspecific. The primary value of CT in AS involves its ability to detect and clearly define erosion of bone at any joint or enthesis. New bone formation is also well visualized in the form of syndesmophytes, ligamentous ossification, and periarticular and intra-articular ankylosis, but use of CT in this regard is limited. CT can show these findings equally well in the axial and peripheral skeleton but is used primarily in areas where radiographic visualization or interpretation is problematic.

Use in Diagnosis, Monitoring, and Prognostication

Diagnosis. The diagnosis of AS is based primarily on the radiographic observation of bilateral moderate or unilateral severe sacroiliitis. When good quality radiographs of the sacroiliac joints are normal or radiographic changes meet diagnostic criteria, CT has no role. Early detection of AS is better investigated by MRI, and if clearly defined structural damage is present on x-ray, CT has no additional diagnostic utility. However, CT of the sacroiliac joints is much easier to interpret than radiography, which is notoriously subject to poor observer reliability. When radiographic findings are
unclear, CT will usually resolve this uncertainty. Because CT shows bone erosion in exquisite detail, it may have a role to play in the further investigation of equivocal MRI findings. It should be noted that classification criteria for AS depend on radiographic findings and more recently MRI but not CT specifically. In the spine, CT is useful in the diagnosis of complications of late disease such as spondylodiskitis or spinal fracture when patients may be unable to tolerate MRI because of pain or spinal deformity.

Monitoring Disease Activity and Damage. CT has no useful role in monitoring disease activity or damage. CT cannot show active inflammation, and the relatively high radiation dose of CT precludes its routine use for assessment of damage progression.

Prognostication. The prognostic value of CT findings of sacroiliitis requires further investigation.

Psoriatic Arthritis

CT is rarely used for the investigation of PsA, and very few data have been published on this topic. In the peripheral skeleton, it would be expected that the application of CT for PsA would be similar to that for RA, and in the axial skeleton, use of CT would be similar to that for AS. In both cases, use of CT is limited by an inability to directly visualize inflammation in bone or soft tissue, and although it likely has superior ability versus radiography in detecting bone erosion or proliferation, in most circumstances the extra radiation dose is not warranted.

Use in Diagnosis, Monitoring, and Prognostication

CT visualizes bone erosion and bone proliferation, but no systematic studies of specific diagnostic utility or prognostic value have been performed. Because of radiation concerns, CT is not used for disease monitoring in PsA.

Gout

CT allows clear visualization of bone erosion, and recent developments in technology offer exciting prospects for crystal imaging.

Acute and chronic tophaceous gout is associated with deposition of urate crystals in soft tissues and the development of tophi. Chronic tophi are often partially calcified, and all CT scanners are sensitive for the deposition of calcium in soft tissues (see Figure 58-6). The condition frequently involves intraosseous deposition of crystals, which is much better visualized on CT than on radiography. CT shows quite well soft tissue swelling due to tophi, and because anatomic bone detail is very good on CT, erosions are clearly delineated. Dual-energy CT is a new technologic development whereby images can be reconstructed based on a two-material decomposition algorithm that can separate calcium from monosodium urate (Figure 58-12). It does not provide imaging of synovitis, tenosynovitis, or osteitis. However, CT scanning with dual-energy acquisition is a recent technical development that appears to offer accurate detection of the presence of monosodium urate and likely will have an important role in diagnosis and research where available.

CT is reliable for monitoring tophus size. A bone erosion scoring method has recently been developed. These methods may be useful in trials.

Calcium Pyrophosphate Dihydrate Crystal Deposition Disease

CT is not used in the management of uncomplicated peripheral CPPD, although calcifications can be visualized. CPPD can involve the cervical spine and may cause neck pain. In such patients, CT may visualize CPPD deposition at the craniovertebral junction, often in the transverse liga-
ment of the atlas (crowned dens syndrome). Calcific deposits may also be found in other periodontoid structures and in the ligamenta flava in some patients, as may osseous abnormalities of the odontoid process such as subchondral cysts or erosions. CT may also be used for verifying the presence of calcifications at other rare locations (e.g., the temporomandibular joints).

Septic Arthritis

CT is rarely used for the investigation of septic arthritis. Acute septic arthritis of large joints is a surgical emergency, and although x-ray, fluoroscopy, and ultrasound may be useful for diagnosis or for image-guided aspiration, no use for CT in the peripheral skeleton is known. In chronic septic arthritis and infectious diskitis, CT may be used to investigate complications, particularly osteomyelitis. However MRI or a variety of scintigraphic studies are generally preferred. In complex cases where MRI scanning is problematic, contrast-enhanced CT can be very useful in detecting abscess formation in deep soft tissues that are not easily assessed by ultrasound. CT guidance may be preferred for spine biopsy as well.

Osteoarthritis

Even though CT can detect bone changes such as osteophytes more accurately than x-ray, overall it offers little additional diagnostic utility in peripheral joints. It generally has poor soft tissue contrast capability, hyaline cartilage cannot be directly visualized, and spatial resolution of bone detail is inferior to that of radiography. It can be used in specific problem-solving situations where anatomy is complex and/or poorly visualized on x-ray, such as in the lumbar spine facet joints, but this is not usually the case in the peripheral joints. CT is used occasionally to assess extent of disease, as in the orthopedic assessment of large joints when detailed evaluation of bone morphology may be required preoperatively. CT scanning with intra-articular contrast material (CT arthrogram) may be the single most accurate modality for detecting cartilage thinning, fissuring, flaps, or volume, as well as loose intra-articular bodies. This requires both a minimally invasive procedure and radiation exposure and is not widely used.
CT arthrography may be used as a problem-solving tool in patients who are not able to undergo MRI.

**Use in Diagnosis, Monitoring, and Prognostication**

CT is not currently used in routine clinical practice for the diagnosis of OA; it has no role in monitoring disease because it cannot quantify change in the target tissue (cartilage) without an invasive procedure. Also, no published data on the independent prognostic use of CT are available because good quality radiography usually provides the same or more information.

**MAGNETIC RESONANCE IMAGING**

**KEY POINTS**

MRI allows visualization and assessment of peripheral inflammatory and destructive joint and soft tissue involvement in degenerative and inflammatory rheumatic diseases.

MRI is by far the best available method for detecting and monitoring inflammation in the spine and sacroiliac joints in AS and other spondyloarthritides.

MRI allows monitoring of inflammatory soft tissue changes (e.g., synovitis, tenosynovitis, enthesitis) during treatment of patients with peripheral inflammatory joint disease.

Through detection of early inflammatory changes, MRI can contribute to an earlier diagnosis of RA and SpA.

MRI bone edema is uniquely visualized by MRI; this provides prognostic information in RA, SpA, OA, undifferentiated inflammatory arthritis, and possibly other rheumatic arthritides.

The main disadvantages of MRI are its cost and lack of availability.

MRI provides multiplanar tomographic imaging with unprecedented soft tissue contrast, without the use of ionizing radiation, and it allows assessment of all structures...
involved in musculoskeletal disease. MRI is more sensitive than clinical examination and x-ray for detection of inflammation and damage in inflammatory and degenerative rheumatologic disorders. Disadvantages of MRI include higher costs and lower availability than radiography, longer examination times, and restriction to a limited anatomic area per session. It should be remembered, however, that costs of MRI represent only a small fraction of the cost of biologic treatment or of the indirect costs of sick leave and early retirement.

After the following section on key technical aspects, characteristics of MRI with respect to diagnosis, monitoring, and prognostication and its clinical utility in RA, SpA, PsA, gout, and OA will be described.

Technical Aspects

MRI is very safe. It involves no ionizing radiation and no increased risk of malignancy, and adverse effects of contrast agents are very rare. However, use of contrast agent should be avoided in patients with severely impaired renal function because of the risk of nephrogenic systemic fibrosis.

T1-weighted (T1w) imaging sequences are favored by relatively short imaging times, good anatomic detail, and the ability to visualize tissues with high perfusion and permeability, including the inflamed synovium, after intravenous contrast (paramagnetic gadolinium compound [Gd]) injection. Fat and Gd-enhanced tissues have high signal intensity on T1w images (Figure 58-14). T2-weighted (T2w) images depict both fat and fluid/edematous tissues with high signal intensity. They are used together with T1w images in degenerative spine disease, while in inflammatory conditions they are particularly useful when fat saturation (FS) techniques (Figure 58-15), in which the signal from fat is suppressed, are applied. This enhances detection of edematous tissue/liquid located in areas with fatty tissue (e.g., bone marrow edema). FS requires a homogeneous field and high magnetic field strength, which are not available on low-field (<1.0 T) MRI units. The only fat-suppressed sequence possible on low-field MRI, the short tau inversion
recovery (STIR) technique, which is based on relaxation time differences, can provide information on bone marrow edema, but with less detail.

Whereas MRI examination of axial joints requires whole body MRI units, MRI of the peripheral joints can be performed with whole body MRI units or with dedicated extremity MRI units (E-MRI). E-MRI increases the potential for widespread rheumatologic use through markedly lower costs, more comfortable patient positioning, and elimination of claustrophobia.

Some E-MRI units may provide information on synovitis and bone destruction that is not markedly inferior to what is obtained by standard sequences on high-field units, but different machines may perform very differently, and for some units, smaller field of view, longer imaging times, and lack of certain imaging techniques (particularly FS at low field) should be considered.

Most MRI studies of the sacroiliac joint have used only one imaging plane—semi-coronal (i.e., parallel with the axis of the sacral bone). A supplementary T1w FS sequence may improve the evaluation of erosions, and sequences designed for cartilage evaluation (e.g., three-dimensional [3D] gradient echo sequences) may be added. For maximal sensitivity to changes in the ligamentous portion of the sacroiliac joints, imaging in the semi-axial plane is required. This may be recommended when MRI is used for diagnostic purposes, although it is probably not essential when used as an outcome measure in trials. On some indications (e.g., suspected disk herniation), MRI should include axial images; however, MRI of the spine in SpA generally involves only sagittal images, but these should extend sufficiently laterally to include the frequently involved facet and costo-vertebral and costotransversal joints. Bone marrow abnormalities in both sacroiliac joints and spine are detected almost equally well with STIR and contrast-enhanced T1w FS sequences in patients with SpA, so contrast injection generally is not needed. In contrast, injection of contrast increases sensitivity for synovitis in peripheral joints.

For evaluation of structural (sometimes referred to as chronic) changes, such as bone erosion, new bone formation, and fat infiltration, T1-weighted images are mandatory. A supplementary T1w FS sequence may improve the evaluation of erosions.

In peripheral and axial joints, specific sequences designed for optimal cartilage evaluation (e.g., 3D gradient echo sequences) can be applied.

**Rheumatoid Arthritis**

Most MRI studies in RA have investigated knee, wrist, and finger joints. Although the knee joint is an excellent model joint for methodological studies, the clinical value of MRI is mainly dependent on its power to evaluate wrists, hands, and feet, which is also the primary focus of this section. Reports on other peripheral joints are few and are not essentially different. Definitions of key pathologies in RA and PsA are provided in Table 58-1.

MRI allows assessment of all structures involved in RA (i.e., synovial membrane, intra-articular and extra-articular...
Imaging Modalities in Rheumatic Diseases

847

CHAPTER 58 | IMAGING MODALITIES IN RHEUMATIC DISEASES

and soft tissue abnormalities, which can serve as a valuable supplement to findings revealed by radiographic evaluation (see Figure 58-2C). MRI is able to directly visualize pannus tissue (e.g., around the odontoid process). Cord compression on MRI seems a better predictor of deterioration than initial clinical and plain radiographic features, supporting MRI as the imaging method of choice for evaluation of spinal cord involvement in RA.

Use in Diagnosis, Monitoring, and Prognostication

Diagnosis. Two recent large follow-up studies of undifferentiated arthritis have documented an independent predictive value of MRI in the diagnosis of RA (see Figure 58-2C). MRI is able to directly visualize pannus tissue (e.g., around the odontoid process). Cord compression on MRI seems a better predictor of deterioration than initial clinical and plain radiographic features, supporting MRI as the imaging method of choice for evaluation of spinal cord involvement in RA.

Figure 58-15  Rheumatoid arthritis (RA) in clinical remission (magnetic resonance imaging [MRI]) and undifferentiated arthritis (MRI). A and B, RA patient in clinical remission. Despite clinical remission, coronal short tau inversion recovery (STIR) (A) and postcontrast T1-weighted fat-suppressed (B) MRI show marked bone marrow edema (osteitis, arrows in A) and synovitis (arrows in B). C through E, Patient with undifferentiated arthritis. Axial T1-weighted precontrast (C) and postcontrast (D) T1-weighted MRI views show synovitis in second and third metacarpophalangeal joints (arrows), whereas the STIR sequence (E) shows bone edema (arrow), X-ray was normal. One year later the patient had developed RA, according to the American College of Rheumatology 1987 criteria. (Courtesy Dr. Anne Duer-Jensen, Glostrup, Denmark.)

fluid collections, cartilage, bone, ligaments, tendons, and tendon sheaths) (see Figure 58-14). MRI and histopathologic and miniarthroscopic signs of synovial inflammation are closely correlated. MRI bone marrow edema (see Figure 58-14) represents inflammatory infiltrates in the bone marrow (i.e., osteitis), as has been demonstrated by comparison with histologic samples obtained at surgery in patients with RA. Whereas erosion reflects bone damage that has already occurred, bone marrow edema appears to represent the link between joint inflammation and bone destruction. A high level of agreement for detection of bone erosion in RA wrists and MCP joints (concordance at 77% to 90% of sites) between MRI and CT, the gold standard reference for detection of bony destruction, documents that MRI erosions signify true bone damage.

In the cervical spine, the primary imaging modality is CR, but MRI can provide detailed information on bone and soft tissue abnormalities, which can serve as a valuable supplement to findings revealed by radiographic evaluation (see Figure 58-2C). MRI is able to directly visualize pannus tissue (e.g., around the odontoid process). Cord compression on MRI seems a better predictor of deterioration than initial clinical and plain radiographic features, supporting MRI as the imaging method of choice for evaluation of spinal cord involvement in RA.
Table 58-1 Magnetic Resonance Imaging Definitions of Inflammatory and Structural Lesions in Rheumatoid Arthritis, Psoriatic Arthritis, and Axial Spondyloarthritis

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Peripheral Joints in RA and PsA</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Inflammatory Lesions</strong></td>
<td>Synovitis: area in the synovial compartment that shows increased postgadolinium (post-Gd) enhancement* of a thickness greater than the width of the normal synovium.</td>
</tr>
<tr>
<td></td>
<td>Tenosynovitis: signal characteristics consistent with increased water content* or abnormal post-Gd enhancement† adjacent to a tendon, in an area with a tendon sheath.</td>
</tr>
<tr>
<td></td>
<td>Periarticular inflammation: signal characteristics consistent with increased water content* or abnormal post-Gd enhancement† at extra-articular sites, including the periostium (<em>periostitis</em>) and the entheses (<em>enthesitis</em>), but not the tendon sheaths‡.</td>
</tr>
<tr>
<td></td>
<td>Bone marrow edema: A lesion* within trabecular bone, with signal characteristics consistent with increased water content* and often with ill-defined margins.</td>
</tr>
<tr>
<td><strong>Structural Lesions</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bone erosion: a sharply margined bone lesion, with typical signal characteristics,* which is visible in two planes with a cortical break seen in at least one plane.</td>
</tr>
<tr>
<td></td>
<td>Bone proliferation: abnormal bone formation in the periarticular region, such as at the entheses (enthesophytes) and across the joint (ankylosis).</td>
</tr>
<tr>
<td><strong>B. Axial Disease: Ankylosing Spondylitis and Other Spondyloarthritides</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Inflammatory Lesions</strong></td>
<td>Bone marrow edema: increase in bone marrow signal* on STIR images.</td>
</tr>
<tr>
<td><strong>Structural Lesions</strong></td>
<td>Bone erosion: full-thickness loss of dark appearance of the cortical bone and change in normally bright appearance of adjacent bone marrow on T1-weighted images†.</td>
</tr>
<tr>
<td></td>
<td>Fat infiltration: focal increased signal in bone marrow on T1-weighted images.</td>
</tr>
<tr>
<td></td>
<td>Bone spur: bright signal on T1-weighted images extending from the vertebral end plate toward the adjacent vertebra (spine).</td>
</tr>
<tr>
<td></td>
<td>Ankylosis: bright signal on T1-weighted images extending across the sacroiliac (SI) joints or extending from one vertebra and being continuous with the adjacent vertebra (spine).</td>
</tr>
</tbody>
</table>

---

*OMERACT MRI in Inflammatory Arthritis Task Force recommendations for MRI definitions of important pathologies in rheumatoid arthritis and peripheral psoriatic arthritis.10,121

‡Canada-Canada MRI Working Group and MORPHO Group recommendations for MRI definitions of important pathologies in spine and sacroiliac joints in axial spondyloarthritis.10,104,105

From OMERACT MRI in Inflammatory Arthritis Task Force.86,121 Canada-Canada MRI Working Group, and MORPHO Group.104,105,107

Arthritis, morning stiffness, positive rheumatoid factor (RF) and MRI, and bone edema score in MTP and wrist joints correctly identified the development of RA or non-RA in 82% of patients.83

According to recent ACR/EULAR 2010 criteria for RA,6 classification as definite RA is based on the presence of definitive clinical synovitis (swelling at clinical examination) in one or more joints, absence of an alternative diagnosis that better explains the synovitis, and achievement of a total score of 6 or greater (of a possible 10) from the individual scores in four domains. In the joint involvement domain, which can provide up to 5 points of the 6 needed for an RA diagnosis, MRI and US synovitis count. In other words, MRI and US can be used to determine joint involvement in the ACR/EULAR 2010 criteria for RA.6,84,85

Monitoring Disease Activity and Structural Damage. To be valuable for monitoring joint inflammation and destruction, a method must be reproducible and sensitive to change. MRI allows quantitative (volume or, for synovitis, early contrast enhancement after intravenous contrast injection) and less detailed (qualitative: presence/absence; semi-quantitative: scoring) evaluation of synovitis, bone edema, and bone erosions. In observational and randomized clinical trials, semi-quantitative scoring by the OMERACT (Outcome Measures in Rheumatology) RA MRI scoring system (RAMRIS) has been used most frequently. It involves...
sensitivity to change and the discriminatory ability of the three most frequently used spine scoring systems (the Ankylosing Spondylitis Disease Activity Index (BASDAI) and Functional Index (BASFI), because disease activity could not be assessed in a sensitive manner by biochemical (mainly C-reactive protein [CRP]) or physical evaluation.

Several systems have been proposed for assessment of disease activity in the sacroiliac joints and in the spine (see a recent review at reference 109 for details). Reproducible and responsive methods are available, Sensitivity to change and the discriminatory ability of the three most frequently used spine scoring systems (the Ankylosing Spondylitis spine Magnetic Resonance Imaging-activity [ASpiMRI-a] score, the Berlin modification of the
ASpiMRI-a score, and the Spondyloarthritis Research Consortium of Canada [SPARCC] scoring system\(^\text{111-113}\) have been demonstrated in clinical trials, and these systems have been tested against each other by the ASAS/OMERACT MRI in AS group.\(^\text{110}\) All methods were feasible, reliable, sensitive to change, and discriminative. The SPARCC method showed the highest sensitivity to change, as judged by Guyatt’s effect size, and the highest reliability, as judged by the inter-reader intraclass correlation coefficient (ICC).\(^\text{110}\)

MRI is much less established for assessment of structural changes (often referred to as chronic changes) than inflammatory changes. MRI undoubtedly provides otherwise inaccessible information on inflammatory activity, and just “equality” of MRI with radiography concerning structural damage assessment is a step forward, because radiography and the ensuing need for two examinations and exposure to ionizing radiation could then be avoided. Scoring methods assess erosion, sclerosis, fat deposition, and/or bone bridges separately or by a global score.\(^\text{105,111,114}\) Validation of the methods used for damage assessment is limited, and their value has not yet been clarified.

**Prognostication.** Three spine studies have documented an association between the presence of bone marrow edema at the anterior corners of the vertebrae on MRI and subsequent development of syndesmophytes on radiography after 2 years of follow-up. Presence as opposed to absence of MRI anterior inflammation provides relative risks of 3 to 5 for a new anterior radiographic syndesmophyte at that level.\(^\text{115-117}\) In two studies, the association was even more pronounced in those vertebral corners in which inflammation had resolved following institution of anti–tumor necrosis factor (TNF) therapy, possibly explained by TNF in an active inflammatory lesion restricting new bone formation, whereas reduction of TNF achieved by applying a TNF antagonist allows tissue repair to manifest as new bone formation.\(^\text{116,117}\)

One study suggests that with early inflammatory back pain, severe sacroiliac MRI bone marrow edema together with human leukocyte antigen (HLA)-B27 positivity is a strong predictor of future AS, whereas mild or no sacroilitis, irrespective of HLA-B27 status, is a predictor of not developing AS.\(^\text{118}\) Data on the value of MRI for predicting therapeutic response in SpA are very limited. A high spine MRI inflammation score and a short disease duration have been reported as statistically significant predictors of clinical response (BASDAI improvement >50%) to anti-TNF therapy.\(^\text{119}\) Additional and larger studies are needed to

---

**Figure 58-16** Sacroiliitis in spondyloarthritis (magnetic resonance imaging [MRI] and x-ray). MRI (A through C) and x-ray (D) of the sacroiliac joints of a 26-year-old male with inflammatory back pain for 4 years. Semi-coronal (A) and semi-axial (B) short tau inversion recovery images show massive iliac and sacral bone edema around the left sacroiliac joint (i.e., demonstrate severe active sacroilitis) (long arrows in A and B). Inflammation (sacroilitis) is also present in the right sacroiliac joint (short arrow in A). T1-weighted semi-coronal image shows fat infiltration (white arrows in C) and bone erosion (black arrows in C). Anteroposterior radiograph (D) shows bilateral mild sclerosis on the iliac side, and bilaterally the articular surface is less well defined than normal, consistent with erosion. However, this feature is exaggerated by overlying bowel gas. (Courtesy Susanne Juhl Pedersen, Glostrup, Denmark.)
Figure 58-17 Spine inflammation in ankylosing spondylitis (magnetic resonance imaging [MRI]). A 43-year-old man with HLA-B27-positive ankylosing spondylitis with deteriorating symptoms, including inflammatory back pain, had an MRI scan before starting biologic therapy. Baseline sagittal short tau inversion recovery (STIR) MRI (A) shows diffuse increased signal (edema) in the T2 vertebral body and multiple foci of corner inflammation anteriorly at T5 and T6, and posteriorly at T7, T8, T9, and T10 (arrows). Other images confirmed extensive active inflammation in the spine. The patient responded very well, and after 6 months of therapy, a repeat STIR MRI (B) showed complete resolution of bone marrow inflammation. Subsequently, the patient experienced recurrence of symptoms, and a third MRI (C) was performed (2 months after anti-TNF therapy was stopped). This MRI shows no edema at T5-T6, a conspicuous new lesion anteriorly at T7, and recurrent inflammation posteriorly in the lower thoracic spine (arrows).

Figure 58-18 The spine in ankylosing spondylitis (AS) (magnetic resonance imaging [MRI]). A through C, T1-weighted sagittal MRI of three different patients with AS shows different structural lesions. A, Fat infiltration (arrowheads) in the bone marrow of several lumbar vertebral corners and anterior fusion (arrows) at L3-L4 and L4-L5. B, Fatty infiltration (arrows) in the bone marrow of multiple vertebral corners in the cervical spine, indicative of the diagnosis of spondyloarthritis. C, Extensive increased marrow fat signal (arrows) crossing the costovertebral joints (thoracic ankylosis) is seen, as are changes in several facet joints and other posterior elements. D, Sagittal short tau inversion recovery MRI shows intense inflammation in the bone marrow of the transverse processes and facet joints of the lower thoracic spine (arrows), as well as less intense inflammation in several discovertebral joints.
clarify the role of MRI in the prediction of disease course and therapeutic response.

Psoriatic Arthritis

The clinical appearance of PsA is highly diverse and involves the spine, SI joints, peripheral joints, and/or entheses; accordingly, MRI findings vary. PsA shares clinical manifestations with RA and SpA, and this also applies to its MRI features. Peripheral PsA synovitis and erosions do not have disease-specific MRI features, and MRI bone edema can involve any bone. General agreement is lacking regarding which joints should be imaged for assessment of PsA activity. This decision may have to be individualized according to the disease pattern observed.

MRI can visualize both peripheral and axial musculoskeletal anatomy and PsA disease manifestations. Findings include synovitis, tenosynovitis, periarticular inflammation, enthesitis, bone edema, bone erosion, and bone proliferation (Figure 58-20). As with other types of SpA, enthesitis, dactylitis, and spondylitis can be seen. Dactylitis has been shown on MRI to be due to tenosynovitis with effusion, sometimes associated with diffuse soft tissue edema and/or synovitis in nearby finger or toe joints (see Figure 58-20). Few MRI studies have examined axial PsA; findings are similar to AS findings but are more frequently asymmetric.

Entheses have attracted attention as a possible primary location of disease. Nail disease is common in PsA, and distal interphalangeal (DIP) joint inflammation on MRI has been described to extend to the nail bed.

PsA can be clinically silent. In patients with psoriasis without arthritic signs or symptoms, pathologic findings on MRI (including periarticular edema, tendon sheath effusion, intra-articular effusion, synovial pannus, bone erosion, bone cysts, subchondral changes, and joint subluxation) have been reported in more than two-thirds (68% to 92%) versus no to one-twelfth or less of healthy controls. The clinical importance of these findings has not yet been clarified.

Use in Diagnosis, Monitoring, and Prognostication

Diagnosis. As has been described, MRI can detect the different pathologies involved in PsA. However, no studies have documented that MRI in an early undifferentiated arthritis cohort can be used to differentiate PsA from other arthritides.

Monitoring. Data on monitoring activity and damage are limited. Most studies have reported only qualitative MRI assessments of the different pathologies of PsA (including periarticular edema, tendon sheath effusion, intra-articular effusion, synovial pannus, bone erosion, bone cysts, subchondral changes, and joint subluxation) have been reported in more than two-thirds (68% to 92%) versus no to one-twelfth or less of healthy controls.

Prognostication. No longitudinal studies of the prognostic value of MRI findings in PsA are available.

Gout

MRI can directly visualize the inflammatory (synovitis, tenosynovitis, bone edema, and soft tissue inflammation) and destructive (bone erosion) aspects of gout arthropathy. MRI can also visualize tophi and can provide
Imaging modalities in Rheumatic Diseases

T1-weighted and proton density images and may simulate meniscal degeneration or tear (see Figure 58-7). As in other diseases, MRI is a sensitive method for subchondral cysts and inflammatory changes such as synovitis and effusion.

Septic Arthritis

The MRI appearance of septic arthritis is nonspecific, and similar findings can be observed in other inflammatory arthritides. Thus in case of clinical suspicion of septic arthritis, a joint aspiration must be performed without delay to avoid irreversible articular damage. Characteristic findings include synovitis, joint effusion, and soft tissue edema (see Figure 58-8). MRI may be helpful in diagnosing complications of septic arthritis, such as abscesses and osteomyelitis. It should be remembered that most individual findings on MRI are not specific, and bone marrow edema alone does not necessarily denote osteomyelitis.

Contrast-enhanced MRI is particularly useful for identifying abscess/necrosis in soft tissue or bone that may confirm the diagnosis and/or requirement for surgical débridement. The painful diabetic foot is a specific clinical presentation for which MRI may now be routinely performed to help distinguish between infection and trauma and to identify nonviable tissues, although scintigraphic studies also play an important role in assessment of these complex patients.

Calcium Pyrophosphate Dihydrate Crystal Deposition Disease

MRI detects calcification of hyaline cartilage as an area of low signal intensity, especially on gradient echo images; meniscal chondrocalcinosis exhibits increased signal on T1-weighted and proton density images and may simulate meniscal degeneration or tear (see Figure 58-7). As in other diseases, MRI is a sensitive method for subchondral cysts and inflammatory changes such as synovitis and effusion.

Osteoarthritis

Based on its tomographic nature and ability to visualize cartilage, bone, and various soft tissues, MRI is very well suited for assessment of inflammatory changes, structural and compositional changes in the cartilage, and.
other structural lesions in OA. Most studies have been undertaken in knee joints (Figure 58-21) or, less frequently, in hip joints, but small joints of the hand (see Figure 58-10) have recently been studied in generalized OA.

MRI allows direct assessment of the thickness, surface contour, and internal architecture of articular cartilage in OA145-147 (see Figure 58-21), making possible the staging and monitoring of OA development. Osteophytes may be seen at the joint margins or beneath the articular cartilage. Subchondral changes include bone edema, sclerosis, and bone cysts. MRI “bone marrow lesions” (sometimes referred to as bone edema lesions), which are areas with inhomogeneous, intermediate to low signal on T1w and high signal in water-sensitive techniques (STIR/T2FS) (Figure 58-22), have, by comparison with histologic samples obtained by surgery in advanced OA, shown trabecular microfracture and bone marrow fibrosis and/or necrosis but limited interstitial edema.148-151 Synovitis is seen frequently on MRI in patients with OA, albeit to a lesser degree than in those with RA.152 Synovitis scores obtained by contrast-enhanced MRI have demonstrated a good correlation with arthroscopic and microscopic synovitis scores.153

Figure 58-21 Osteoarthritis of the knee (x-ray and magnetic resonance imaging [MRI]). A, Anteroposterior x-ray view of the left knee shows large osteophyte of the medial and lateral tibiofemoral joint and mild medial joint space narrowing. B and C, Coronal proton density-weighted MRI confirms the presence of large osteophytes in the tibiofemoral joint, shows diffuse cartilage loss of the medial tibia and femur, and discloses a denuded bone with complete loss of cartilage in the lateral tibial plateau (arrow in B). A small focal cartilage defect of the lateral femoral condyle (arrow in C) is also evident. Both lateral and medial menisci are partially macerated, and a subluxation of the medial meniscus is seen, along with severe cartilage loss on the medial and lateral tibial plateau and the medial femoral condyle. D, Coronal T2-weighted fat-suppressed MRI shows medial tibial plateau subchondral bone marrow edema (arrow in D). (Courtesy Professor Ali Guermazi, Boston, Mass.)
Quantitative measurements apply computer-aided image processing to quantitate various aspects (e.g., volumes of cartilage, bone, bone marrow lesions, menisci, or synovium). These have been reported to have excellent reproducibility.

Quantitative measurements can also provide insights into the composition of cartilage, such as the quantification of glycosaminoglycan content. Semi-quantitative methods have been used to perform semi-quantitative “multiple feature” (“whole organ”) assessments in the knee, based on conventional MRI.

**Use in Diagnosis, Monitoring, and Prognostication of OA**

**Diagnosis.** Classification criteria for OA are based on clinical and conventional radiographic findings. However, the use of MRI in the diagnosis of OA offers the advantages described previously (i.e., that MRI can sensitively depict all involved pathologic changes).

**Monitoring.** Various quantitative and semi-quantitative techniques have been used to measure structural abnormalities and change on MRI in patients with OA. Quantitative measurements apply computer-aided image processing to quantitate various aspects (e.g., volumes of cartilage, bone, bone marrow lesions, menisci, or synovium). These have been reported to have excellent reproducibility. Measures of cartilage composition (e.g., quantification of glycosaminoglycan content) are also available. Semi-quantitative methods have been used to perform semi-quantitative “multiple feature” (“whole organ”) assessments in the knee, based on conventional MRI.
acquisitions. These measures have high reliability and better sensitivity to change than have been noted with x-ray methods.

Prognostication. In a recent systematic literature review, quantitative cartilage volume change and the presence of cartilage defects or bone marrow lesions (bone edema) in three of three studies were significantly related to subsequent total knee replacement (i.e., a predictive value was demonstrated).

Furthermore, enlargements in bone marrow lesions (BMLs) have been demonstrated to be related to increased pain (and improvements in bone marrow lesions to decreased pain) in follow-up studies. In contrast, inconsistent and generally weak relations between cartilage loss and symptom change and a weak relation between change in synovitis and change in pain have been reported. Finally, the presence of meniscal damage, cartilage defects, and/or BMLs predicts subsequent MRI progression.

ULTRASONOGRAPHY

KEY POINTS

- US allows sensitive visualization and assessment of peripheral inflammatory and destructive joint and soft tissue involvement in a variety of degenerative and inflammatory rheumatic diseases.
- US can be performed easily by trained rheumatologists in relation to the clinical examination.
- US can, through detection of early inflammatory changes, contribute to earlier diagnosis of RA.
- Grayscale and/or power Doppler US allows monitoring of soft tissue changes (e.g., synovitis, tenosynovitis, enthesitis) during treatment of patients with peripheral inflammatory joint disease.
- US allows guidance of invasive procedures, making possible precise needle positioning for aspirations and injections (e.g., in joints, bursae, or tendon sheaths or at entheses).
- The main disadvantages of US include the need for a skilled investigator, the inter-reader and interscanner variability, and the need for an “acoustic window.”

Ultrasonography (US) is an evolving imaging technique that is used increasingly by rheumatologists in daily clinical practice. This trend has been confirmed over the past few years by extensive publications and a wide body of literature. In some countries, including Belgium, Italy, Germany, and Spain, US has become an integral part of rheumatology training. Absence of radiation, good visual assessment of tissue vascularity with Doppler US, and this may have an impact on the detection of subclinical synovitis. However, the use of ultrasound contrast agents for assessment of rheumatic diseases remains restricted to the research field in terms of ethical and cost/efficacy reasons.

US is very useful as a source of guidance for invasive procedures. US-guided needle placement within the selected target area is an accurate and safe approach in the treatment of patients requiring aspiration of synovial fluid, injection therapy, or biopsy. Under US guidance, progression of the needle in the soft tissue toward the selected target area can be carefully controlled, thereby avoiding injury induced by the tip of the needle and/or complications such as subcutaneous or intratendinous injection of corticosteroid, which introduces a high risk of tissue damage. Moreover, US-guided injections for inflammatory arthritis are significantly less painful than palpation-guided models, and even though recent studies have not been able to demonstrate significantly improved clinical outcomes, US guidance may improve clinical outcomes and enhance cost-effectiveness.

US can be used in daily rheumatology practice for a wide spectrum of indications (see “Key Points: Ultrasonography”).

Technical Aspects

Musculoskeletal US requires high-quality equipment. Low-frequency transducers (3.5 to 5 MHz) are useful for exploring deep targets (e.g., hip, sacroiliac joints). Medium-frequency transducers (6 to 13 MHz) represent the best choice for exploring large joints (e.g., shoulder, knee, elbow). High-frequency transducers (14 to 22 MHz) have a low penetration power of the ultrasonic beam but provide an accurate assessment of superficial structures (e.g., tendons, finger joints), reaching a resolution power of less than 0.1 mm.

US is an operator-dependent technique. Thus, the quality of the pictures and the interpretation of sonographic images depend on the quality of the US equipment, the technical conditions of the examination, and the skills and experience of the examiner.

Color Doppler and power Doppler US explore blood perfusion and may play a key role in monitoring inflammatory disease activity and in assessing response to therapy. Color Doppler provides information on the direction of blood flow. Power Doppler has greater sensitivity in detecting blood flow but is not able to determine flow direction or velocity. Use of US contrast agent enhances the detection of tissue vascularity with Doppler US, and this may have an impact on the detection of subclinical synovitis. However, the use of ultrasound contrast agents for assessment of rheumatic diseases remains restricted to the research field in terms of ethical and cost/efficacy reasons.

US is a continuously evolving technique; recent advances in technology promise additional improvements, especially in the fields of three-dimensional (3D) and four-dimensional (4D) US, elastosonography, and fusion imaging. Three-dimensional US has the potential to dramatically reduce the duration of the US examination and to improve...
CHAPTER 58 | IMAGING MODALITIES IN RHEUMATIC DISEASES

857

Imaging Modalities In Rheumatic Diseases

Another limitation of US is seen in the comprehensive assessment of some anatomic areas (spine, sacroiliac joint, and hip) caused by lack of adequate “acoustic windows.” Pitfalls and errors most often are related to misinterpretation of normal anatomy, inadequate regulation of the US beam, and inappropriate examination technique, mainly depending on pressure, position, and inclination of the probe.

US findings provide a wide range of useful information concerning the main tissue changes that occur in patients with rheumatic disease. However, standardization about scanning methods and definitions of basic sonographic abnormalities remain limited. The most widely accepted definitions for sonographic pathology were published in 2005 by the OMERACT special interest group (Table 58-2).

Examination Technique

US examination of the musculoskeletal system should be performed with multiplanar scans. European League Against Rheumatism (EULAR) guidelines have been produced to standardize the approach to musculoskeletal US in rheumatology.

Patient position is critical in performing a US examination correctly. Particular attention must be paid to the angle of the US beam, which must be perpendicular to the examined structure to avoid anisotropy-dependent artifacts (spatial asymmetry). These artifacts may generate diagnostic errors, especially when fibrillar or fascicular structures such as tendons and nerves are studied. Both longitudinal and transverse scans should be performed by slightly moving the transducer from radial to ulnar and from proximal to distal sides to enable maximum coverage of the anatomic surface area.

Operator dependency is the main shortcoming of US. Adequate supervised training is essential to ensure the correct use of US and appropriate interpretation of results. Another limitation of US is seen in the comprehensive assessment of some anatomic areas (spine, sacroiliac joint, and hip) caused by lack of adequate “acoustic windows.” Pitfalls and errors most often are related to misinterpretation of normal anatomy, inadequate regulation of the US beam, and inappropriate examination technique, mainly depending on pressure, position, and inclination of the probe.

US findings provide a wide range of useful information concerning the main tissue changes that occur in patients with rheumatic disease. However, standardization about scanning methods and definitions of basic sonographic abnormalities remain limited. The most widely accepted definitions for sonographic pathology were published in 2005 by the OMERACT special interest group (Table 58-2).

Rheumatoid Arthritis

RA is the inflammatory condition most frequently studied by US. US abnormalities detected in patients with RA include joint space widening, fluid collection, synovial hypertrophy, cartilage defects, bone erosions, tendon sheath widening, and tendon tears. Joint cavity widening is the earliest characteristic US finding of synovitis. Even minimal intra-articular fluid collection and synovial proliferation may be revealed by US. Fluid collection shows a typical anechoic pattern that can easily be distinguished from the soft echogenicity of synovial hypertrophy. This appears as a homogeneous thickening of the synovial layer or as irregularly shaped clusters of echoes (bushy and villous

Figure 58-23 Various regional pain syndromes (ultrasonography). A, Shoulder (anterior transverse scan): fluid collection in the subdeltoid bursa (asterisk). B, Achilles tendon (at) (longitudinal scan): large tendon tear (arrow). C, Knee (posterior longitudinal scan): massive fluid collection inside the semi-membranous gastrocnemius bursa (popliteal cyst) (asterisk). D, Elbow: epicondylitis characterized by intense power Doppler signal within the common extensor tendon of the fingers (tec). bt, long head of the biceps tendon; c, calcaneus; d, deltoid muscle; g, gastrocnemius muscle; o, lateral epicondyle; r, radial capitellum.
Detection of a highly perfused synovial pannus in patients with early arthritis should be regarded as a potential predictor of the future appearance of anatomic damage (Figure 58-24). US can also be used in assessing cartilage damage at metacarpophalangeal joint level in patients with RA (Figure 58-25).

The high spatial resolution of US allows recognition of even minimal bone erosion. Sensitivity is highest in easily US-accessible areas, including the frequently involved second metacarpophalangeal and fifth metatarsophalangeal joints. Bone erosions are viewed on US as an interruption of the sharp hyperechoic bone profile. US depicts the wall and the floor of the erosions that in most cases are filled by a hyperperfused synovial pannus. It has been demonstrated that US can be used to detect more erosions than x-ray, especially in early RA, and that US erosions are true erosions when compared with CT used as the gold standard.

Tendon involvement is frequent and characteristic in patients with RA. The spectrum of pathologic changes detectable in tendons by US is wide and heterogeneous. Tendon sheath widening, loss of the normal fibrillar echotexture, and loss of definition of tendon margins in partial or total tears are the abnormalities that characterize tenosynovitis. Circumscribed or extensive collection of anechoic fluid within the tendon sheath is the most characteristic feature of exudative tenosynovitis. The presence of echoes within the tendon sheath can be related to synovial proliferation and/or to aggregates of cells and proteins.

Table 58-2  Definitions of Ultrasonographic Pathology in Inflammatory Joint Disease

| **Bone erosion:** an intra-articular discontinuity of the bone surface that is visible in two perpendicular planes |
| **Synovial fluid:** abnormal hypoechoic or anechoic (relative to subdermal fat, but sometimes may be isoechoic or hyperechoic) intra-articular material that is displaceable and compressible but does not exhibit Doppler signal |
| **Synovial hypertrophy:** abnormal hypoechoic (relative to subdermal fat, but sometimes may be isoechoic or hyperechoic) intra-articular tissue that is nondisplaceable and poorly compressible and that may exhibit Doppler signal |
| **Tenosynovitis:** hypoechoic or anechoic thickened tissue with or without fluid within the tendon sheath, which is seen in two perpendicular planes and may exhibit Doppler signal |
| **Enthesopathy:** abnormally hypoechoic (loss of normal fibrillar architecture) and/or thickened tendon or ligament at its bony attachment (may occasionally contain hyperechoic foci consistent with calcification), seen in two perpendicular planes that may exhibit Doppler signal and/or bony changes, including entheseophytes, erosions, or irregularity |

From OMERACT Ultrasonography Task Force.

Figure 58-24  Rheumatoid arthritis (ultrasonography). A, Metacarpophalangeal joint (dorsal longitudinal scan): joint cavity widening, synovial hypertrophy (sh), and intense power Doppler signal. B, Fifth metatarsophalangeal joint (dorsal longitudinal scan): large bone erosion (arrows) associated with power Doppler signal inside the erosion. C, Tibialis posterior tendon (tp): tendon sheath widening with homogeneous anechoic aspect of the content (fluid collection) (asterisk). D, First compartment of the finger extensor tendons (longitudinal scan): chronic tenosynovitis characterized by synovial hypertrophy and intense power Doppler signal. a, astragalus; at, abductor pollicis longus tendon; et, common extensor tendon of the finger; m, metacarpal bone; mh, metatarsal head; mm, medial malleolus; pp, proximal phalanx.
Very subtle changes in tendon structure can be detected in patients with early RA. Their predictive value in detecting more aggressive or rapidly progressive forms of disease has not been clearly defined, and follow-up investigations are needed. Early circumscribed tendon tears appear as minimal fragmentation of small groups of contiguous fibrils; this represents a characteristic loss of the normal fibrillar echotexture of the tendon. In more advanced stages of structural damage, tendons can be subjected to large partial tears or to complete rupture. If the US beam and the major axis of the tendon are not perfectly perpendicular, certain areas of the tendon appear anechoic (anisotropy artifacts) and thus may be mistakenly interpreted as possible tendon ruptures.

US is sensitive for the detection of arthritis when several joints are examined. A significantly higher number of inflamed joints are found by US than by clinical assessment. However, US examination of 78 joints requires 70 minutes, and examination of even a seven-joint set can be challenging in daily clinical practice.180,181

Use in Diagnosis, Monitoring, and Prognostication of RA

Diagnosis. The ability of US to allow visualization of intra-articular and extra-articular changes suggests that it can be used to assist clinicians in reaching a specific diagnosis (e.g., in patients with early undifferentiated arthritis). However, data describing the differential diagnostic value of US in patients with undifferentiated arthritis are very limited. In the seronegative (RF and anticitrullinated protein antibody (ACPA) negative) subgroup of 50 early undifferentiated inflammatory arthritis patients, certain clinical and radiographic features (elevated CRP, swollen joints, and radiographic erosion) raise the probability of persistent inflammatory arthritis from 6% to 30%, whereas the addition of one to three specific US features (grade 3 grayscale synovial hypertrophy, at least grade 2 synovial power Doppler signal, and at least one erosion) increased the probability to 50% to 94%, suggesting that combining ultrasonographic evaluation with routine assessment in seronegative early arthritis may markedly increase the certainty of diagnosis.182 In seropositive disease (RF and/or ACPA positive), US had no predictive value.

Monitoring Disease Activity and Damage. Even though international consensus has not yet been reached on the best possible scoring system and joints to score in the assessment of synovitis and other joint pathologies in RA,181,183 the OMERACT Ultrasound Task Force has developed definitions for various joint pathologies that can be used for further development and validation. US measures that correlate with synovitis (Doppler signal and B-mode synovial membrane thickness) decrease in parallel with other markers of disease activity when glucocorticoids171,172,184-186 or TNF antagonists187-192 are administered, indicating their potential for monitoring joint inflammation in RA.193 US allows visualization of cartilage changes and bone erosion in accessible areas,194-197 but data are lacking on its reproducibility and sensitivity to change. Thus, US is probably a valid method for monitoring synovitis and, in accessible areas, damage progression, but the relative advantage of US over x-ray for showing damage progression remains unclear, and more data are needed on reproducibility (e.g., probe positioning, selection of site for the picture, machine selection, machine/pressure adjustments) and sensitivity to change when the entire US examination is repeated.

Prognostication. Data on the predictive value of US findings are ambiguous. One study found US synovitis to be predictive of future radiographic progression in patients...
treated with conventional DMARDs but not with anti-TNFs; another found no significant correlation between baseline clinical, laboratory, functional, and US parameters and 1-year follow-up DAS28 (measure of disease activity in RA), Health Assessment Questionnaire (HAQ) score, and radiographic scores in DMARD-treated patients.

The Leeds group has reported that US (and MRI) evidence of joint inflammation are frequent in patients in clinical remission. Baseline US synovial hypertrophy, US power Doppler, and MRI synovitis scores in individual joints were significantly related to progressive radiographic damage. Furthermore, a significant association was noted between power Doppler score at baseline and structural progression over 12 months in asymptomatic MCP joints and 12 times higher odds of structural progression in joints with increased power Doppler signal. In a recent study of the same group of RA patients in clinical remission on anti-TNF, US could not predict who could successfully stop anti-TNF therapy. It remains to be determined whether US can predict long-term disease progression, structural damage progression, and preservation of function better than traditional clinical or serologic scores.

**Ankylosing Spondylitis/Axial Spondyloarthritis**

Despite the fact that contrast-enhanced Doppler US has been reported to have a high negative predictive value for the detection of sacroilitis, the role of US in assessment of sacroiliac and spine involvement in AS and other types of axial SpA is minimal.

AS and other types of axial SpA frequently involve peripheral joints and entheses. US of peripheral involvement in SpA is described in the following section on PsA.

**Psoriatic Arthritis**

US assessment of patients with PsA should be focused on joints, tendons with and without a synovial sheath, entheses, skin, and nails (Figure 58-26). Basic sonographic

---

**Figure 58-26** Psoriatic arthritis (ultrasonography). A, Wrist (dorsal longitudinal scan): marked joint cavity widening of both radiocarpal and intercarpal joints with intense power Doppler signal. B, Proximal interphalangeal joint (longitudinal volar scan): chronic tenosynovitis with areas of low tendon echogenicity (arrowheads) and power Doppler signal. C, Metacarpophalangeal joint (dorsal longitudinal scan): peritendinous inflammation characterized by hypoechoic swelling surrounding the extensor digitorum tendon (asterisk) associated with power Doppler signal and subcutaneous edema (circle). D, Distal patellar enthesis (longitudinal anterior scan): thickening of the enthesis associated with intense power Doppler signal and bone erosion (arrowhead). E, Psoriatic plaque: hypoechoic subepidermal thickening (circle) with intense power Doppler signal (asterisk) indicating vasodilatation and angiogenesis. F, Psoriatic onychopathy: loss of the normal trilaminar aspect of the nail plate (arrow), thickening of the nail bed (asterisk), and intense power Doppler signal. c, capitate; cet, common extensor digitorum tendons; dp, distal phalanx; et, common extensor tendon of the finger; ft, flexor finger tendons; l, lunate; m, metacarpal bone; mp, middle phalanx; pa, patella; pp, proximal phalanx; pt, patellar tendon; r, radius.
changes in joint and synovial tendons can be similar to those of RA. However, a highly hypervascularized pannus seems to be more prominent in PsA patients in both small and large joints. This feature appears to be related to histopathologic findings.\textsuperscript{205} Tendons without synovial sheaths are characteristically involved in PsA. Inflammatory changes detectable by US include loss of fibrillar echotexture and focal or diffuse hypoechoic tendon thickening due to intratendinous edema, with or without power Doppler signal. Moreover, peritendinous inflammation may appear as a hypoechoic swelling of soft tissue surrounding the tendon, which may show power Doppler signal.\textsuperscript{206}

US can detect a wide range of abnormalities of the enthesis. They consist of enthesal thickening, focal hypochochogenicity, loss of the homogeneous fibrillar echotexture, irregularity of the bone profile (enthesophytes and/or bone erosions), and power Doppler signal. Evidence supports the use of US as a very sensitive tool for detecting subclinical enthesopathy in patients with psoriasis who have no clinical signs of arthritis and/or enthesitis. Hyperchoic thickening of the epidermis with or without acoustic shadowing and hypoechoic swelling of the dermis with evident power Doppler signal are the main US features of psoriatic plaque.

Both the nail plate and the nail bed can be accurately depicted by US, and a wide range of abnormalities can be detectable, including thickening and loss of the typical trilaminar aspect of the nail plate with or without increased Doppler signal at the nail bed.\textsuperscript{204}

\textbf{Use in Diagnosis, Monitoring, and Prognostication of PsA}

\textbf{Diagnosis.} US is more sensitive than clinical examination for detection of synovitis, tenosynovitis, and enthesitis in patients with PsA,\textsuperscript{207,209} but no studies have yet documented that US in an early undifferentiated arthritis cohort can be used to differentiate PsA from other arthritides. Because enthesitis is prominent in patients with SpA and may precede joint symptoms, it has been of interest to evaluate entheses by using US as a means of diagnosing SpA and therefore also PsA.\textsuperscript{210} A recent study indicates that early US enthesitis will predict the development of SpA.\textsuperscript{211} Further work is needed to clarify the role of US in the diagnosis of PsA.

\textbf{Monitoring.} Most studies that aim to monitor treatment response have applied semi-quantitative scoring systems for grayscale and/or Doppler changes.\textsuperscript{212,213} TNF blocker treatments have provided significant reductions in grayscale and Doppler semi-quantitative synovitis scores,\textsuperscript{212} in enthesal morphologic abnormalities, in power Doppler signal, and in bursitis.\textsuperscript{214} Several systems for assessment of enthesitis have been proposed.\textsuperscript{215,216,217} No system has been sufficiently validated to be used for monitoring PsA or other spondyloarthritides. More research is needed before standardized, reliable, and responsive US outcome measures in PsA are available.

\textbf{Prognostication.} No studies have evaluated the role of US in prognosticating PsA. Enthesal involvement in patients with psoriasis but without clinical PsA indicates that enthesitis may be a predictor of development of PsA.\textsuperscript{218} However, longitudinal studies are required to clarify this.

\textbf{Gout}

US can detect monosodium urate crystal deposition on the cartilage surface, inside the joint cavity, and around and inside the tendons. The spectrum of the US appearance of urate crystal aggregates can vary from homogeneously punctate to sharply defined hyperechoic densities of variable size and eventually to dense topographic material with posterior acoustic shadows.\textsuperscript{219} Joint cavity widening is the most frequent US finding in patients with acute gout.

Monosodium urate crystal deposition on the surface of the articular cartilage results in hyperechoic enhancement of the chondrosynovial margin, which can range from homogeneous thickening of the chondrosynovial interface to areas of focal deposition.\textsuperscript{220} Bone erosions are frequently detectable in patients with chronic gout.

Asymptomatic and symptomatic involvement of tendons is common in patients with gout. Urate crystal aggregates of variable size and shape are frequently detectable, especially at the patella and the Achilles tendons. The normal fibrillar echotexture of tendons can be completely altered by the presence of intratendinous monosodium urate deposits, which appear as areas of inhomogeneous echogenicity, sometime generating an acoustic shadow (Figure 58-27).\textsuperscript{225}

\textbf{Use in Diagnosis, Monitoring, and Prognostication}

The detection of tophi could be helpful in diagnosing gout, especially when these lesions are not detectable clinically. However, although detection of a nodule with a characteristic appearance on US, MRI, or CT is highly suggestive of gout, joint aspirate confirmation of monosodium urate crystals is usually required to rule out infection or other space-occupying lesions. Tophus size can be reliably assessed by US, and this may be useful for monitoring treatment efficacy in clinical trials.\textsuperscript{242,219} No US scoring system has been developed for overall assessment of gout. The prognostic value of US findings is not known.

\textbf{Calcium Pyrophosphate Dihydrate Crystal Deposition Disease}

US has been found to be highly sensitive in detecting aggregates of calcium pyrophosphate dihydrate (CPPD) crystals.\textsuperscript{220,221} CPPD aggregates range from tiny circumscribed hyperechoic spots to extended deposits with or without acoustic shadow. As indicated by EULAR recommendations for calcium pyrophosphate deposition: “US can demonstrate CPPD in peripheral joints, appearing typically as thin hyperechoic bands within hyaline cartilage and hyperechoic sparkling spots in fibrocartilage. Sensitivity and specificity appear excellent and possibly better than those of conventional x-rays.” Moreover, floating aggregates of crystals can be observed inside fluid collections (see Figure 58-27).

\textbf{Septic Arthritis}

Septic arthritis is a major emergency in rheumatology. The role of US is to provide early findings that can raise the suspicion of joint infection. Synovial fluid in septic arthritis may show a variable degree of reflectivity. The effusion can
irregularity of the bone profile to multiple erosions can be detected in areas that appear normal on conventional x-ray in both erosive and nonerosive OA. A slight increase in cartilage thickness caused by inflammatory edema in the early phases of OA has been described.

Osteophytes are the most frequent and characteristic abnormalities of the bone profile in patients with OA. They appear as irregularities of the bone contour with posterior acoustic shadow. Joint effusions are commonly found in patients with OA. Minimal fluid collections that are not detectable on clinical examination are easily demonstrated by US. The inhomogeneous echogenicity of synovial fluid can be related to proteinaceous material, cartilage fragments, aggregates of crystals, and calcified loose bodies. Osteoarthritis

US is exquisitely sensitive in detecting structural changes in the articular cartilage. Irregularity and/or loss of the thin and sharp contour of the outer border of the hyaline cartilage and/or increased cartilage echogenicity with patchy or diffuse loss of clarity is the earliest US feature of OA. These changes seem to reflect structural alterations such as fibrillation of cartilage and cleft formation. Variable loss of cartilage is detectable in patients with more advanced disease. Alteration of subchondral bone ranging from tiny irregularity of the bone profile to multiple erosions can be detected in areas that appear normal on conventional x-ray in both erosive and nonerosive OA. A slight increase in cartilage thickness caused by inflammatory edema in the early phases of OA has been described.

Osteophytes are the most frequent and characteristic abnormalities of the bone profile in patients with OA. They appear as irregularities of the bone contour with posterior acoustic shadow. Joint effusions are commonly found in patients with OA. Minimal fluid collections that are not detectable on clinical examination are easily demonstrated by US. The inhomogeneous echogenicity of synovial fluid can be related to proteinaceous material, cartilage fragments, aggregates of crystals, and calcified loose bodies. Osteoarthritis

US is exquisitely sensitive in detecting structural changes in the articular cartilage. Irregularity and/or loss of the thin and sharp contour of the outer border of the hyaline cartilage and/or increased cartilage echogenicity with patchy or diffuse loss of clarity is the earliest US feature of OA. These changes seem to reflect structural alterations such as fibrillation of cartilage and cleft formation. Variable loss of cartilage is detectable in patients with more advanced disease. Alteration of subchondral bone ranging from tiny irregularity of the bone profile to multiple erosions can be detected in areas that appear normal on conventional x-ray in both erosive and nonerosive OA. A slight increase in cartilage thickness caused by inflammatory edema in the early phases of OA has been described.

Osteophytes are the most frequent and characteristic abnormalities of the bone profile in patients with OA. They appear as irregularities of the bone contour with posterior acoustic shadow. Joint effusions are commonly found in patients with OA. Minimal fluid collections that are not detectable on clinical examination are easily demonstrated by US. The inhomogeneous echogenicity of synovial fluid can be related to proteinaceous material, cartilage fragments, aggregates of crystals, and calcified loose bodies. Osteoarthritis

US is exquisitely sensitive in detecting structural changes in the articular cartilage. Irregularity and/or loss of the thin and sharp contour of the outer border of the hyaline cartilage and/or increased cartilage echogenicity with patchy or diffuse loss of clarity is the earliest US feature of OA. These changes seem to reflect structural alterations such as fibrillation of cartilage and cleft formation. Variable loss of cartilage is detectable in patients with more advanced disease. Alteration of subchondral bone ranging from tiny irregularity of the bone profile to multiple erosions can be detected in areas that appear normal on conventional x-ray in both erosive and nonerosive OA. A slight increase in cartilage thickness caused by inflammatory edema in the early phases of OA has been described.
sonographic features similar to those observed in patients with RA but without the invasive properties of the rheumatoid synovitis.

US-detected synovial inflammation and effusion are common in painful knee OA and correlate significantly with knee synovitis, effusion, and clinical parameters suggestive of an inflammatory “flare.” As in RA, US detects more joints with inflammation than can be detected by clinical assessment.

**Diagnosis, Monitoring, and Prognostication**

**Diagnosis.** Classification criteria for OA are based on clinical and conventional radiographic findings, not on US. However, the use of US in the diagnosis of OA is favored by the fact that US can visualize the pathologic changes previously described.

**Monitoring.** US can demonstrate changes in synovitis thickness, effusion size, and popliteal cyst size. With respect to structural changes, including cartilage damage, the sensitivity to change remains to be determined. Reproducibility data are scarce, and no consensus on scoring systems has been reached. Quantitative assessment of cartilage is restricted to thickness because total volumes cannot be measured. More work is required to develop standardized definitions of pathology and to demonstrate the validity of US in OA.

**Prognostication.** The significance of ultrasonographically detected pathologic conditions in OA with regard to symptoms, prognosis, outcome, and response to therapy remains to be determined.

**OTHER IMAGING MODALITIES**

**Digital X-Ray Radiogrammetry (DXR)**

**KEY POINTS**

- DXR performs automated assessment of conventional hand x-rays and provides an approximated bone mineral density (BMD).
- Bone loss as measured by DXR-BMD may be useful for monitoring and predicting bone damage in RA.

Three types of bone loss are early features of RA: focal articular bone erosion, periarticular osteopenia, and systemic osteoporosis. Periarticular bone loss is frequently the earliest radiographic feature of RA, is associated with disease activity, and is known to precede bone erosions. However, radiographic osteopenia is detected only when bone loss is greater than 30%.

Digital x-ray radiogrammetry (DXR) is a fully automated technique used to perform radiogrammetry from standard hand radiographs to bridge the gap between radiogrammetry and bone densitometry. DXR technology is based on combined computerized radiogrammetric and textural analyses of the narrowest parts of the second, third, and fourth metacarpal bones. Based on cortical thickness (cm), porosity index, and an assumption of a constant bone density and elliptical bone, DXR calculates an approximated bone mineral density (BMD; g/cm²). Short-term precision of DXR-BMD (coefficient of variation [CV]) and long-term precision have been reported as 0.28% and 0.25%, respectively, and reproducibility as 0.05% to 0.27%. DXR is highly correlated with periarticular BMD measurements and therefore can reflect periarticular bone loss, even though it measures midshaft cortical bone. DXR-BMD shows greater loss in RA patients with high levels of inflammation, positive rheumatoid factor, or positive ACPA. Further, measurements of DXR-BMD changes are significantly associated with progression of structural damage in early RA subjects. Losses seen over 1 year with DXR-BMD are predictive of subsequent erosive development. However, so far only changes in DXR-BMD have been shown to be of predictive value, limiting its clinical applicability. DXR-BMD changes have also been shown to be responsive to treatment interventions. During disease-modifying treatment, DXR-BMD losses are less severe in patients responding to treatment than in those who are not responding. This suggests that DXR-BMD can become a useful outcome measure and predictor of joint damage in RA clinical trials. However, its clinical usefulness and applicability have yet to be established.

**Capillaroscopy**

**KEY POINTS**

- Capillaroscopy allows “in vivo” assessment of the microcirculation.
- A wide spectrum of abnormalities, including architectural disorganization, enlarged loops, and avascular areas, can be detected in patients with systemic sclerosis and related conditions.
- Capillary loss and avascular areas may be related to more extensive skin and/or visceral involvement and a poor prognosis.

Capillaroscopy has proved to be a valuable technique for “in vivo” assessment of microcirculation. Distinctive nailfold capillary abnormalities have been described in a wide spectrum of conditions, including systemic sclerosis, mixed connective tissue disease, dermatomyositis, Sjögren’s syndrome, vinyl chloride disease, Raynaud’s phenomenon, vibrating tools disease, and acrocyanosis.

Capillaroscopy can be performed using various optical instruments such as ophthalmoscope, stereomicroscope, photomacrography system, and videomicroscopy equipment. In vivo assessment of skin capillaries is generally performed at the nailfold, where the major axis of the capillaries is parallel to the skin surface. The general configuration of the cutaneous capillaries at the nailfold is “hairpin-like” or “U-shaped.” However, a wide range of variability has been noted among healthy subjects and between the fingers of the same individual. The best visibility of the nailfold capillaries is generally found at the fourth and fifth fingers of the nondominant hand.

The main capillaroscopic parameters to be evaluated include skin transparency, visibility of the subpapillary venular plexus, architectural structure of the microvascular network, density and spatial distribution of capillaries,
Sjögren’s syndrome, especially when associated with Raynaud’s phenomenon.

Tortuosity, loop elongation, enlarged and/or branching loops, and improved visibility of the subpapillary venous plexus have been reported in about 50% of patients with systemic lupus erythematosus. Even if these abnormalities appear to be nonspecific, some of them, especially enlarged loops and capillary loss, have been correlated with lung involvement. Morphea and eosinophilic fasciitis are usually characterized by a capillaroscopic pattern within the normal range.

Growing interest has been expressed in the use of capillaroscopy to assess the efficacy of treatment in patients with systemic sclerosis. Reduced capillary loss was observed after treatment with cyclosporin A and intravenous iloprost.

In patients with clinically isolated Raynaud’s phenomenon, even circumscribed abnormalities such as megacapillaries and irregularly enlarged loops should alert the physician to the possibility of an underlying scleroderma spectrum disorder.

Capillaroscopic changes in patients with undifferentiated connective tissue disease and dermatomyositis may be quite similar to those seen in patients with systemic sclerosis. However, relevant features of angiogenesis can be the dominant capillaroscopic abnormality in dermatomyositis. A scleroderma-like pattern can also be detectable in primary Sjögren’s syndrome, especially when associated with Raynaud’s phenomenon.

Nuclear Medicine

**KEY POINTS**

Nuclear imaging provides not only morphologic but also physiologic information regarding the metabolic state of tissues.

Positron emission tomography (PET) can be useful in the diagnosis and management of vasculitis fever of unknown origin, and malignancies.

Other imaging modalities are generally better suited than nuclear medicine for investigating inflammatory and degenerative rheumatic diseases in clinical practice.

**Scintigraphy (Planar)**

Nuclear imaging provides not only morphologic but also physiologic information regarding the metabolic state of tissues.
tissues. Following injection of a radionuclide, gamma radiation emitted from the patient is captured by a gamma camera to generate a two-dimensional image. A wide range of radionuclides are available. The classical “bone scan” uses $^{99m}$technetium chelated to diphosphonates (specifically, methylene diphosphonate [MDP]). Other compounds, such as $^{67}$Ga citrate and $^{111}$In, are more specific for detecting inflammatory pathology, but their use must be dictated by a specific clinical setting in terms of relative costs and technical constraints. High sensitivity for active bone and joint pathology, high negative predictive value, and the ease of performing a whole body assessment (e.g., imaging Paget’s disease and metastatic disease) are the main advantages of scintigraphy. Low specificity, limited anatomic and spatial resolution (compared with plain radiography), use of ionizing radiation, and cost are the main weaknesses of this imaging technique.

In the evaluation of patients with arthritides, scintigraphy can be a useful adjunct to clinical assessment but should not be used routinely. Bone scans are more sensitive than conventional radiographs for detecting active bone pathology. In turn, MRI is usually more sensitive and specific than bone scintigraphy, with the added advantage of not using ionizing radiation. However, bone scintigraphy can provide a useful complementary role to MRI and other modalities for several clinical situations.

Single-Photon Emission Computed Tomography (SPECT)

SPECT is the generation of cross-sectional scintigraphic images produced by processing of multiple planar images obtained over a large range of camera positioning relative to the subject. The relationship of SPECT to planar scintigraphic imaging is analogous to the relationship between CT and plain radiography.

The use of SPECT has resulted in increased sensitivity and specificity compared with planar bone scintigraphy, especially in the assessment of spine pathology. Compared with planar scintigraphy, SPECT allows cross-sectional anatomic localization of active pathology (Figure 58-29). Tomographic reconstructions provide a more precise display of tracer accumulation, which helps to differentiate structures that overlap on planar images. This feature improves the specific diagnosis of conditions such as facet joint syndrome.

Poor anatomic and spatial resolution is the main limitation of conventional planar and SPECT bone scan images. However, newer SPECT/CT technology allows concurrent acquisition of a SPECT and a conventional CT study on the same machine, followed by fusion of their images (see Figure 58-29). This new modality combines the sensitivity of functional radionuclide images with the high spatial resolution and anatomic specificity of CT. A SPECT/CT study “maps” the functional abnormality demonstrated on a bone scan to an exact anatomic site on a CT scan. This improved anatomic specificity enhances the accuracy of cases in which the specific anatomic location of the lesion strongly influences the diagnosis. One common example is a small discrete lesion at a vertebral level, which is not visible on radiography or on CT but is clearly visible (active) on a bone scan. Depending on other findings, the most likely diagnosis depends on its specific location: facet joint (arthritis), pars interarticularis (spondylolysis), or vertebral body (metastatic disease).

Positron Emission Tomography (PET)

Positron emission tomography (PET) is a functional imaging technique that enables metabolic mapping of the tissues in vivo with positron-emitting radionuclides. $^{18}$F-Fluorodeoxyglucose ($^{18}$F-FDG) is the most commonly used radiopharmaceutical in PET because of its availability, favorable 110-minute half-life, and high uptake in most cancers. When injected into the body, $^{18}$F-FDG demonstrates sites of increased glucose metabolism, thus identifying sites of metabolically active pathology in both bone and soft tissue. A PET scan is inherently tomographic (cross-sectional), and it has a slightly higher spatial resolution than a SPECT scan. A “PET/CT” scan acquires both a PET and a CT scan concurrently on the same machine, allowing fusion of the two sets of images. Similar to SPECT/CT, the CT component of PET/CT provides a precise anatomic map superimposed on the (metabolic) PET images. $^{18}$F-FDG-PET (and/or PET/CT) is used mainly in oncology, where it has a well-established role in the diagnosis and clinical management of several malignancies, but it is also useful for a variety of clinical problems in cardiology, neurology, and infectious diseases. $^{18}$F-FDG-PET is now the test of choice for the assessment of fever of unknown origin.

PET can be useful for the diagnosis and management of osteomyelitis, Takayasu’s arteritis, and large vessel vasculitis occurring in patients with granulomatosis with polyangiitis (formerly Wegener’s granulomatosis), polyarteritis nodosa, giant cell arteritis, and polymyalgia rheumatic. FDG uptake is strongly correlated with MRI synovitis in RA, but its role, if any, for quantifying disease activity in inflammatory arthritides remains to be determined.

CONCLUSION

Imaging is an integral part of management of patients with rheumatic disease. This chapter has outlined the status of imaging in rheumatic disease, with a particular focus on inflammatory joint disease. Please see the Key Points for main messages. This chapter has explained the important role of conventional radiography, as well as the exciting new opportunities attainable with newer imaging techniques. The last decade has brought a vast amount of new knowledge that has markedly changed the way we manage our patients with rheumatic disease. It is exciting that with continued dedicated research and rapid technical development, it is likely that even larger improvements may occur in the decade to come, for the benefit of our patients.

Acknowledgments

The authors thank Drs. Anne Duer Jensen, Susanne Juhl Pedersen, Rene Foggenborg (all Copenhagen), Ali Guermazi (Boston), Ida Haugen (Oslo), and Fiona McQueen (Auckland) and Ho Jen (Edmonton) for providing images, and Trine Jensen David (Copenhagen), Ho Jen (Edmonton), and Pernille Bøyesen (Oslo) for critical review of parts of the text.

The authors thank Drs. Anne Duer Jensen, Susanne Juhl Pedersen, Rene Foggenborg (all Copenhagen), Ali Guermazi (Boston), Ida Haugen (Oslo), and Fiona McQueen (Auckland) and Ho Jen (Edmonton) for providing images, and Trine Jensen David (Copenhagen), Ho Jen (Edmonton), and Pernille Bøyesen (Oslo) for critical review of parts of the text.
PART 7 | DIAGNOSTIC TESTS AND PROCEDURES IN RHEUMATIC DISEASES

866

Selected References


CHAPTER 58 | IMAGING MODALITIES IN RHEUMATIC DISEASES


Full references for this chapter can be found on www.expertconsult.com.
References

40. Steinbach LS, Resnick D: Calcium pyrophosphate dihydrate crystal deposition disease revisited, Radiology 203:1–9, 1996.


84. Østergaard M: Clarification of the role of ultrasonography, magnetic resonance imaging and conventional radiography in the ACR/EULAR 2010 rheumatoid arthritis classification criteria: comment to the article by Aletha et al, Ann Rheum Dis 2010.e-letter. Published online December 2, 2010.


functions while COX-2 expression is increased during inflammation and other pathologic situations. Inhibition of COX-2 by NSAIDs blocks PG production at sites of inflammation while inhibition of COX-1 in certain other tissues, most importantly platelets and the gastroduodenal mucosa, can lead to common adverse effects of NSAIDs such as bleeding, bruising, and gastrointestinal (GI) ulceration.

In addition to their use in rheumatoid arthritis and osteoarthritis, NSAIDs are widely used in the symptomatic management of other rheumatic diseases characterized by chronic musculoskeletal pain and diverse forms of acute pain. Aspirin, which has unique properties among NSAIDs, is used by millions more for primary and secondary prevention of cardiovascular thrombosis. In light of the widespread use of these drugs for common diseases, which are likely to increase in prevalence with the aging of the population, it is critically important to appreciate the potential adverse events and drug interactions associated with NSAIDs.

This chapter analyzes aspirin and other NSAIDs on the basis of chemical structure, pharmacologic properties, and relative inhibition of COX-1 and COX-2. Particular attention to potential adverse events of specific NSAIDs in individual patients will facilitate use of these drugs in the safest possible manner. Acetaminophen (known as paracetamol outside the United States), an antipyretic and analgesic drug without anti-inflammatory activity, inhibits COX enzymes by a different mechanism than NSAIDs and is also discussed. Colchicine possesses anti-inflammatory characteristics similar to NSAIDs in some situations and is discussed in this chapter, although this drug differs in its mechanism of action and profile of adverse effects.

HISTORY

Botanicals containing salicylates have been used since antiquity to treat pain, inflammation, and fever. The Egyptian Ebers papyrus recommended use of a decoction of dried myrtle leaves to be applied to the abdomen and back for relief of rheumatic pains about 3500 years ago. A thousand years later, Hippocrates recommended poplar tree juices for eye disease treatment and willow bark to alleviate fever and the pain of childbirth. Throughout Roman times, the use of
botanical treatments including willow bark for pain and
inflammation was widespread. The medicinal use of salicylate-containing plants occurred in China and other
parts of Asia. In addition, the curative effects of other
botanicals were known to the indigenous populations of
North America. Colchicine-containing extracts of the
autumn crocus plant were used for treatment of acute gout
as early as the sixth century AD.1

The first modern report of the therapeutic application of
salicylate-containing plants was reported to the Royal
Society of London by the Reverend Edward Stone, who
provided an account of the success of the dried bark of the
willow for fever.1 In this first “clinical trial,” a pound of bark
was dried, pulverized, and put into the tea, beer, or water of
50 people with fever. He found that one dose (1 dram =
1.8 g) cured their fever. In 1763 Stone wrote, “I have no
other motives for publishing this valuable specific, than that
it may have a fair and full trial in all its variety of circum-
stances and situations, and that the world may reap the
benefits accruing from it.”

In 1860 salicylic acid was chemically synthesized, which
led to its widespread use as an external antiseptic, anti-
pyretic, and analgesic.1 The bitter taste of salicylic acid
prompted the chemist Felix Hoffman to synthesize the more
palatable acetylsalicylic acid (ASA). After demonstration
of its anti-inflammatory effects, Dr. Heinrich Dreser of
Bayer introduced this compound into medicine in 1899 as aspirin and it remains the most widely used drug in the
world.2 Salicylate was identified as the active ingredient of
willow bark in 1929.

Phenybutazone came into clinical practice in 1949 and
was followed by indomethacin, fenamates, naproxen, and
others. Despite the diversity of their chemical structures,
these drugs shared therapeutic properties with aspirin. Fur-
thermore, adverse events including gastric upset, GI ulcer-
ation and bleeding, hypertension, edema, and renal damage
were shared by all these drugs. In 1971 it was discovered
that these drugs all acted by inhibiting PG biosynthesis,
thereby providing a unifying explanation of their therapeu-
tic actions and a rationale for grouping them together as
NSAIDs.3

COX was isolated in 1976 from the endoplasmic reticu-
lum of PG-forming cells.2,3 However, several groups specu-
lated that there must be a second COX enzyme on the basis
of observed biology. In 1990 investigators demonstrated
that bacterial lipopolysaccharide (LPS) increased PG syn-
thesis in human monocytes in vitro and in mouse peritoneal
macrophages in vivo, but only the LPS-induced increase
was inhibited by dexamethasone and required the de novo
synthesis of “new” COX protein.4 This observation was the
foundation of the concept for “constitutive” and “inducible”
forms of COX. Soon thereafter a number of investigators
working in different systems reported the discovery of an
inducible second form of COX.5 Investigators went on to
close the gene and deduced its structure, and they found
the gene product was homologous to COX, but to no other
known protein. The observation that glucocorticoids inhi-
bited the expression of COX-2 following a proinflammatory
stimulus represented a link between the anti-inflammatory
actions of NSAIDs and corticosteroids.

Because of the prediction that inhibiting COX-2 would
block PG biosynthesis participating in the inflammatory

response but was not required for homeostasis, there was a
tremendous push to develop drugs that specifically inhibit
COX-2 without effect on COX-1 in the belief that these
medications would provide clinical efficacy without adverse
effects.5,6 Identification of new drugs that differentially
inhibited COX-2 over COX-1 was accomplished quickly
as existing NSAIDs were tested on the two COX isoforms
and crystal structures revealed differences in the protein
structures on which new drug development could be based.5,6

One hundred years after aspirin was introduced and 10
years after the discovery of COX-2, selective COX-2 inhibi-
tors, celecoxib (Celebrex) and rofecoxib (Vioxx), were
developed. In clinical trials, the safety and efficacy profiles
of these and related drugs showed promise and the U.S.
Food and Drug Administration (FDA) subsequently
approved these COX-2-selective NSAIDs for treatment of
arthritis and pain. After the introduction into clinical prac-
tice, however, it became clear that the most highly
COX-2-selective NSAIDs, particularly rofecoxib, were more
likely than traditional NSAIDs to be associated with adverse
cardiovascular events.7 This finding led to the voluntary
withdrawal of rofecoxib and several other COX-2-selective
NSAIDs from the market. Debate surrounding the relative
risks of different NSAIDs to specific organ systems contin-
ues to the present.

**CYCLOOXYGENASE BIOLOGY AND
BIOACTIVE LIPIDS**

Therapeutic and adverse effects of NSAIDs are best under-
stood in the context of COX biology. The COX enzymes
are the first committed step in the synthesis of PG from
arachidonic acid (AA) (Figure 59-1). AA is an omega-6
polyunsaturated fatty acid (PUFA) commonly found at the
sn-2 position of cell membrane glycerophospholipids and
cleaved from cell membranes by one of several different
phospholipase A2 enzymes.8 Once generated, AA may enter
different pathways to generate bioactive lipids. It can be
metabolized by COX to PGG2 and then PGH2 via its
COX and peroxidase enzyme activities; by lipoxygenases
(LOX) to hydroxyeicosatetraenoic acids (HETEs), hydro-
peroxyeicosatetraenoic acids (HPETEs), or leukotrienes; or
by the cytochrome p450 family of enzymes to HETEs,
HPETEs, or epoxycycotoxtriene acids (EETs).9,10 These
bioactive lipids have diverse biologic activities in normal phys-
iology and in pathologic conditions including inflammation,
pain, cardiovascular disease, and cancer.

In the COX pathway, the unstable intermediate PGH2
spontaneously rearranges or is enzymatically converted by
specific synthases to biologically active PG, of which there
are many isoforms.11 Although phospholipase A2 activity is
required to initiate PG synthesis, the overall regulation of
the type and amount of PG produced in a given cell or tissue
is determined by the expression levels of COX-1, COX-2,
and terminal synthase enzymes.

Bioactive lipids synthesized via the COX and LOX path-
ways are important mediators of inflammation, but alternate
substrates and pathways can generate anti-inflammatory
lipids and lipids important in the resolution of inflammation
(Figure 59-2).11,12 Eicosapentaenoic acid (EPA) and docosa-
hexaenoic acid (DHA), which are omega-3 PUFAs, can
also serve as substrates for COX and LOX and their metabolic pathways lead to generation of resolvins, protectins, and maresins, bioactive lipids associated with resolution of inflammation.\textsuperscript{12,13} It is hypothesized that generation of anti-inflammatory lipids and lipids that accelerate resolution of inflammation may explain why diets high in omega-3 fatty acids are associated with reduced inflammatory and cardiovascular diseases.\textsuperscript{11} In addition, ASA possesses the capacity to alter COX-2 activity such that an “ASA-triggered” alternative catalytic cascade also results in synthesis of anti-inflammatory products including lipoxins, protectins, and resolvins (see Figure 59-2).\textsuperscript{11}

**Prostaglandin Production and Action**

PGs act as autocrine and paracrine mediators with effects limited to the immediate vicinity of their synthesis. That PG actions are limited to their site of synthesis reflects their ephemeral nature and predisposition to metabolic inactivation to regulate promptly their potent biologic activities. Thus the enzymes required for production of a specific prostanoid must be co-localized and adjacent to the receptor mediating the action of prostanoids. Most cell types produce only one or a few prostanoid species; however, transcellular metabolism of unstable intermediates of COX may mediate qualitative and quantitative changes in the profile of eicosanoids produced in a given tissue.\textsuperscript{14} PGs are exported from cells via the multidrug resistance-associated protein family of efflux transporters, and the PG transporter mediates influx of PG and metabolism to stable products via hydroxyprostaglandin dehydrogenase (see Figure 59-1).\textsuperscript{10,16}

\textit{PG}E\textsubscript{2}, the most abundant PG at sites of inflammation, can be produced by many different cell types via at least three different PG synthases.\textsuperscript{11} Expression of the inducible form of PG synthase, microsomal PG synthase-1 (mPGES-1), is similar to expression of COX-2. mPGES-1 acts in concert with COX-2 to produce high levels of PG\textsubscript{E}\textsubscript{2} during inflammation. mPGES-1 expression is blocked by NSAIDs, and experiments demonstrate that PG\textsubscript{E}\textsubscript{2} itself, along with proinflammatory stimuli, is required for mPGES-1 upregulation.\textsuperscript{16} This positive regulatory loop contributes to the high levels of PG\textsubscript{E}\textsubscript{2} at sites of inflammation and likely to the efficacy of NSAIDs. Other terminal PG synthases are not known to be highly regulated, and levels of other

---

**Figure 59-1** An overview of eicosanoid synthesis pathways. Arachidonic acid (AA) is a polyunsaturated fatty acid that constitutes the phospholipid domain of most cell membranes and is liberated from the cellular membranes by cytoplasmic phospholipase A\textsubscript{2} (PLA\textsubscript{2}). Free AA can be metabolized to eicosanoids through three major pathways: the cyclooxygenase (COX), lipoxigenase (LOX), and the cytochrome P450 mono-oxygenase pathways. In the COX pathway, the key step is the enzymatic conversion of arachidonic acid to the intermediate prostaglandin G\textsubscript{2} (PG\textsubscript{G}\textsubscript{2}), which is then reduced to an intermediate PG\textsubscript{H}\textsubscript{2}, by the peroxidase activity of COX. PG\textsubscript{H}\textsubscript{2} is subsequently metabolized to prostanoids including PGs and thromboxanes (TX) by specific synthases. LOX converts AA to biologically active metabolites such as leukotrienes (LT) and hydroxyeicosatetraenoic acids (HETE). P450 metabolizes AA to epoxygenesatrienoic acids (EET). In the 5-LOX pathway, AA is converted to an intermediary 5-HETE, which is converted to LT\textsubscript{A}\textsubscript{4}, LT\textsubscript{B}\textsubscript{4} is subsequently converted to 5-HETE, LT\textsubscript{B}\textsubscript{4}, LT\textsubscript{C}\textsubscript{4} LT\textsubscript{D}\textsubscript{4}, and LTE\textsubscript{4}. Each of the PG and LT exerts its biologic effects by binding to its cognate G protein–coupled receptor. PG\textsubscript{I}\textsubscript{2} can transactivate the nuclear peroxisome proliferator-activated receptor \(\delta\) (PPAR\(\delta\)), and a PG\textsubscript{D}\textsubscript{2} dehydrogen product, 15dPG\textsubscript{J}\textsubscript{2}, is a natural ligand for PPAR\(\alpha\). The multidrug resistance–associated protein (MRP) gene family is composed of efflux transporters for both PG and LT. PG transporter (PGT) is an influx transporter for PG. Hydroxyprostaglandin dehydrogenase 15-(NAD) (15-PGDH) mainly metabolizes intracellular PG\textsubscript{E}\textsubscript{2} to stable 13,14-dihydro-15-keto-PGs. (Reprinted with permission from Wang D, Dubois RN: Eicosanoids and cancer, Nat Rev Cancer 10:181–193, 2010.)
important bioactive PGs are stably produced or their levels are increased when phospholipase activity and COX-2 levels are increased.

The actions of PG are mediated by cell surface G protein–coupled receptors (GPCRs). There are at least nine known PG receptors with additional splice variants (see Figure 59-1).17 The PG receptors belong to three clusters within a distinct subfamily of the GPCR superfamily with the lone exception of one of the PGD2 receptors (DP2), which belongs to the chemokine receptor subfamily. The relaxant receptors for prostacyclin (IP), PGD2 (DP1), and PGE2 (EP2 and EP4) signal through G-mediated increases in intracellular cyclic adenosine monophosphate (cAMP). The contractile receptors for thromboxane A2 (TP), PGF2α (FP), and PGE2 (EP3) signal through a Gα-mediated increase in intracellular calcium. An inhibitory receptor for PGE2 (EP1) couples to G<sub>i</sub> and decreases cAMP formation. Note that PGE2 has at least four different receptors with a broad range of potential actions. EP4 in particular appears to mediate many of the proinflammatory activities of PGE2.18 Given the great diversity of PG receptors expressed by different cell types, PG signaling pathways constitute an enormously complex network controlling many biologic actions. Much work remains to understand fully all of the cellular signaling mechanisms by which PGs and their receptors elicit their respective biologic actions. This is particularly true as antagonists for many of these receptors show promise as novel targets for drug development.18

**Biochemistry and Structural Biology**

COX-1 and COX-2 are bifunctional enzymes that mediate a COX reaction whereby arachidonate plus two molecules of O<sub>2</sub> are converted to the cyclic endoperoxide PGG<sub>2</sub>, followed by a hydroperoxidase reaction in which PGG<sub>2</sub> undergoes a two-electron reduction to PGH<sub>2</sub>.8 COX enzymes are integral membrane proteins that sit within the inner leaflet of the lipid bilayer of intracellular phospholipid membranes of the nuclear envelope and the endoplasmic reticulum. The crystal structures of COX-1 and COX-2 have been solved, and they have essentially identical domain structures.6 The COX enzymes are homodimers with each monomer consisting of three structural domains. The N-terminal, epidermal growth factor–like domain is involved in dimerization via hydrophobic interactions. The membrane-binding domain is composed of four amphipathic α-helices lodged into half of the lipid bilayer to form a hydrophobic channel in the center of the large catalytic domain that contains the COX and peroxidase active sites and that constitutes about 80% of the protein. The catalytic domain is globular with two distinct intertwining lobes. The interface of these lobes creates a shallow cleft on the upper surface of the enzyme where the peroxidase active site is located and where heme is bound.

COX and hydroperoxidase reactions occur at distinct but structurally and functionally interconnected sites. The COX reaction is peroxide dependent and requires that the
heme group at the peroxidase site undergo a two-electron oxidation. A tyrosine residue (tyrosine 385) located at the COX active site is involved as a reaction intermediate. The physiologic heme oxidant in vivo is not known, but it has been shown that the COX activity of COX-2 can be activated at 10-fold lower concentrations of hydroperoxide than that of COX-1.9

NSAIDs function by blocking access of AA to the COX active site within a long, narrow, dead-end, hydrophobic channel whose entrance is framed by the four amphipathic helices of the membrane-binding domain. The channel extends into the globular catalytic domain and is about 8 Å wide. Significant narrowing of the channel occurs where arginine 120 protrudes into the channel. Arginine 120 is essential for binding both AA substrate and most carboxylate-containing NSAIDs in COX-1. By virtue of other differences in the hydrophobic channel, this residue is unessential for binding AA in COX-2, whereas it remains critical for binding of most carboxylate-containing NSAIDs.19,20 Serine 530 is required for inhibition of COX by the phenylacetic acid NSAID diclofenac. Serine 530 is also the residue transacetylated by ASA and, along with valine 349, seems to govern the stereochemistry of the pocket such that after exposure to aspirin, AA is sterically blocked from a functional interaction with the catalytic domain in COX-1 and catalytic activity is completely inhibited.

A crucial structural difference between COX-1 and COX-2 is the substitution of the small amino acid valine in COX-2 for the isoleucine with a bulky side chain in COX-1 at position 523 that opens a side pocket in the hydrophobic channel in COX-2.6 Overall, COX-2 has a wider and somewhat more flexible interior channel, a structural feature that has been exploited with respect to the development of COX-2-selective NSAIDs as shown in Figure 59-3.21

Figure 59-3 Cyclooxygenase (COX)-1 and COX-2 substrate-binding channels. Schematic depiction of the structural differences between the substrate-binding channels of COX-1 and COX-2 that allowed the design of selective inhibitors. The amino acid residues, Val434, Arg513, and Val532, form a side pocket in COX-2 that is absent in COX-1. A, Nonselective inhibitors have access to the binding channels of both isoforms. B, The more voluminous residues in COX-1, Ile434, His513, and Ile532, obstruct access of the bulky side chains of the coxibs. (From Grosser T, Fries S, FitzGerald GA: Biological basis for the cardiovascular consequences of COX-2 inhibition: therapeutic challenges and opportunities. J Clin Invest 116:4–15, 2006.)
Both enzymes are homodimers, but the monomers often behave asymmetrically as conformational heterodimers during catalysis and inhibition. That is, when a fatty acid binds to one monomer, the other monomer becomes catalytically active and only one monomer is catalytically active at any one time. The specific fatty acid bound to the noncatalytic monomer can regulate catalytic activity. Different NSAIDs interact differently with respect to allosteric inhibition of COX enzymes as one facet of their pharmacology. ASA acetylation occurs in only one of the two COX monomers, which completely inhibits the activity of COX-1. However, COX-2 retains the ability to form a reduced amount of PGH₂ and alternate aspirin-triggered lipoxins from AA. The anti-inflammatory resolvins may be synthesized by ASA-acetylated COX-2 from omega-3 PUFA.

**Molecular Biology**

In addition to the differences in their structures relevant for the pharmacology of COX-1 and COX-2 inhibition, there are physiologically relevant differences with respect to expression and regulation. Generally, COX-1 is constitutively expressed in most cells and its expression is minimally altered by inflammatory stimuli. The promoter region of COX-1 has the characteristics of a gene that is continuously transcribed and stably expressed. COX-1 activity is regulated by substrate (AA) availability. When there is an increase in substrate mobilization via phospholipase A₂ activation, there is a concordant increase in PG synthesis mediated by COX-1. COX-1 is the only isoform expressed in mature platelets and is the most highly expressed COX isoform in normal gastroduodenal mucosa. Because COX-1 is inhibited by nonselective NSAIDs, these physiologic properties may explain some of the common adverse effects of these drugs such as bleeding and GI ulceration.

In contrast, the COX-2 gene has the structure of a highly regulated product with binding sites for transcription factors such as nuclear factor κB (NFκB), cAMP-responsive element, and activating protein-1 (AP-1), which rapidly increase transcription in response to inflammatory signals. COX-2 expression is highly induced by proinflammatory cytokines such as tumor necrosis factor and interleukin-1β (IL-1β), microbial products, and mitogens and is inhibited by glucocorticoids. COX-2 mRNA stability is a key regulator of COX-2 levels. The potential for instability of the COX-2 message is due to the presence of multiple AUUUA instability sequences in the 3' region that mediate a rapid degradation of mRNA, which ultimately suppresses COX-2 protein synthesis and PG production. Conversely, some stimuli including IL-1β may interfere with mRNA degradation and increase COX-2 levels and PG production. COX-1 and COX-2 can be post-translationally modified. COX-1 is glycosylated at three asparagines involved in proper protein folding of the enzyme, whereas COX-2 can be glycosylated at four asparagines.

The generalization that COX-1 expression is constitutive and COX-2 expression is inducible has its limitations, given that COX-2 is expressed constitutively in several organ systems and regulated by physiologic, as well as pathologic, stimuli. COX-2 is basally expressed in the brain, kidney, pancreas, and blood vessels and therefore plays an important role in normal reproductive, renal, cardiovascular, and skeletal physiology.

**MECHANISM OF ACTION**

**Cyclooxygenase Inhibition**

All of the NSAIDs are synthetic inhibitors of the COX active site, but subtle mechanistic differences in the manner in which individual NSAIDs interact and bind with the active site are responsible for some of the differences in their pharmacologic characteristics. ASA is the only covalent, irreversible inhibitor of COX-1 and COX-2, whereas all of the other NSAIDs are competitive inhibitors, competing with AA for binding in the active site. The competitive inhibitors are subdivided further on the basis of whether they bind to the COX active site in a time-dependent or time-independent manner.

Crystallographic studies have shown how ASA effectively acetylates serine 530 of COX-1. Similar to other NSAIDs, ASA diffuses into the COX-1 active site at the mouth of the channel and travels to the constriction created by arginine 120, where it is in the best orientation to transacetylate serine 530, leading to the complete and irreversible inhibition of COX-1. In COX-2, the channel of the active site is larger than COX-1, the orientation of ASA for serine 530 attack is not as good, and transacylation efficiency for COX-2 is 10-fold to 100-fold less than for COX-1. ASA can also “trigger” COX-2 to alter its catalytic activity to produce 15 R-HETE and lipoxins from AA and to generate anti-inflammatory lipids from omega-3-PUFA.

The time it takes for an NSAID to inhibit the COX active site relative to how long it takes for it to leave the COX channel is a crucial factor in the inhibition of COX. Drugs such as ibuprofen exhibit such rapid rates that they essentially inhibit COX instantly but can be removed from the COX active site just as quickly when drug levels decrease. Both COX monomers must be inhibited by ibuprofen to block catalytic activity. Conversely, indomethacin and diclofenac are time-dependent allosteric inhibitors that require seconds to minutes to bind to the COX active site and need only block one of the COX monomers to completely inhibit catalytic activity. These NSAIDs also need hours to exit the COX active site. Initially, most traditional time-dependent NSAIDs form a loose complex with the COX active site before a stronger interaction is established. This complex is limited by the time it takes the drug to become properly oriented within the COX channel at arginine 120, the constriction site in the COX channel. This may involve a change in conformation to the “open state” to allow the drug to access the upper part of the COX catalytic site.

Drugs such as flurbiprofen and indomethacin form a salt bridge between the carboxylate moiety of the NSAID and the guanidinium moiety of arginine 120. Hydrophobic interactions between the aromatic rings and the hydrophobic amino acids in the channel aid binding. Such interactions at the constriction point of the channel completely block the entry of substrate to the active site. Diclofenac interacts with serine 530, not arginine 120, but also blocks entry of substrate.
**COX-2 Selectivity**

NSAIDs such as meloxicam, nimesulide, and etodolac show some selectivity for inhibiting COX-2 over COX-1. After the discovery of COX-2, efforts to further enhance COX-2 selectivity led to the development of celecoxib, rofecoxib, valdecoxb, etoricoxib, and lumiracoxib. The prototypical COX-2-selective NSAIDs, celecoxib and rofecoxib, are diaryl compounds containing a sulfonamide (celecoxib) and methysulfone (rofecoxib) rather than a carboxyl group. Both drugs are weak time-independent inhibitors of COX-1 but strong time-dependent inhibitors of COX-2 that require their entry into and stabilized binding in the catalytic pocket. Because these drugs lack a carboxyl group, arginine 120 is not involved, but multiple sites of hydrogen and hydrophobic binding stabilize drugs at the catalytic site. The sulfur-containing phenyl ring of COX-2-selective NSAIDs plays a pivotal role in binding stability by occupying the hydrophobic side pocket characteristic of the COX-2 catalytic site. If this side pocket is removed by mutagenesis, all isozyme selectivity is lost.  

COX isozyme selectivity is defined most commonly using the concentration of drug required to inhibit PG production by 50% in a particular assay system (inhibitory concentration, or IC$_{50}$). Ratios using values obtained for COX-1 IC50s compared with COX-2 IC50s can be calculated and used as a standard measure for comparing the degrees of selectivity of a particular NSAID for one or the other COX isoform. PG assay systems can vary widely, however, making it difficult to compare directly results from studies using different assay systems. To circumvent such problems, most clinicians have accepted the use of the in vitro whole-blood assay to compare NSAID selectivities. In this system, COX-1 inhibition is assessed as a function of the reduction of thromboxane made by platelets after clot formation. Inhibition of COX-2 is based on the inhibition of PGE$_{2}$ production in a heparinized blood sample after LPS stimulation. A COX-2-selective NSAID lacks inhibitory effect on platelet COX-1 at concentrations at or above those that maximally inhibit COX-2.  

**Cyclooxygenase-Independent Mechanisms of Action**

At high, nonphysiologic concentrations, some NSAIDs seem to elicit effects on cellular pathways in vitro that do not involve the inhibition of COX. Because of the high doses of drug required and the use of in vitro systems, the relevance of these effects to in vivo activity is uncertain. Some NSAIDs inhibit phosphodiesterases associated with the metabolism of cAMP leading to increased intracellular cAMP levels and the subsequent general inhibition of peripheral blood lymphocyte responses to mitogen stimulation, monocyte and neutrophil migration, and neutrophil aggregation. NSAIDs scavenge free radicals, inhibit superoxide production by polymorphonuclear neutrophils, reduce mononuclear cell phospholipase C activity, and inhibit inducible nitric oxide synthase activity. Sodium salicylate and ASA inhibit NFkB activation, as do certain inactive enantiomers of flurbiprofen. Some reports indicate that other cell signaling molecules such as mitogen-activated protein kinases and the transcription factor AP-1 may also be modulated by NSAIDs. Some NSAIDs bind to and activate members of the peroxisome proliferator-activated receptor (PPAR) family and other intracellular receptors. PPAR activation is thought to mediate anti-inflammatory activities. Selective COX-2 inhibitors may have unique structural features that promote COX-independent activities such as cell-cycle regulation, apoptosis, and antiangiogenesis.  

**Mechanism of Acetaminophen and Other Analgesic Antipyretic Drugs**

Acetaminophen (paracetamol) and dipyrone relieve pain and fever, but they are not anti-inflammatory. The precise mechanisms by which these drugs elicit their effects remain unclear. In the 1970s, it was proposed that acetaminophen worked by means of a “central” action by inhibiting COX activity primarily in the brain and not in peripheral tissues because they were not acidic and could cross the blood-brain barrier. Acetaminophen does inhibit COX-1 and COX-2, but variably so and dependent on cell and tissue type. Acetaminophen does not appear to inhibit by interaction with the COX active site; rather, it serves as a reducing co-substrate for the peroxidase site. The peroxide tone of cells and tissues in vivo may be responsible for inhibitor specificity, with platelets and activated macrophages being resistant to the action of acetaminophen and vascular endothelial cells being sensitive to its inhibitory effects on COX. Additionally, the inhibitory potency of acetaminophen is determined by the concentration of the COX enzyme. This may be an additional factor for the lack of clinical anti-inflammatory effects because inflammation is associated with a markedly increased expression of COX-2 enzyme. With the discovery of a COX-1 splice variant and studies showing that it is both highly expressed in brain and more sensitive to inhibition by acetaminophen, some authors proposed that the analgesic and antipyretic actions of acetaminophen could be explained by its ability to inhibit the COX-1 splice variants (called COX-3 by some despite the fact that this variant does not arise from a unique gene). However, more recent studies have rejected this mechanism as explanatory for acetaminophen effects.  

Salicylate has analgesic, antipyretic, and anti-inflammatory activity but, similar to acetaminophen and in contrast to ASA, is a poor COX inhibitor. Salicylate has also been shown to inhibit COX activity if substrate levels are low, and it is also dependent on the oxidative state of the enzyme, suggesting that this drug may inhibit COX by redox-related mechanisms.  

**PHARMACOLOGY AND DOSING**

**Classification**

NSAIDs are generally grouped according to their chemical structures, plasma half-life, and COX-1 versus COX-2 selectivity (Table 59-1 and Figure 59-4). Table 59-1 presents a representative compilation of common NSAIDs, formulations, dosages, half-lives, and precautions. Structurally, most NSAIDs are organic acids with low pK values that lend themselves to their accumulation at sites of inflammation,
<table>
<thead>
<tr>
<th>Drug</th>
<th>Brand Name</th>
<th>Available Formulations (mg)</th>
<th>Maximal Daily Dose (mg)</th>
<th>Tmax (hr)</th>
<th>Half-life (hr)</th>
<th>Dose Adjustment or Special Precautions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Salicylic Acids</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetylsalicylic acid</td>
<td>Aspirin</td>
<td>Tablets: 81,165, 325, 500, 650 Children’s: 81 Suppository: 120, 200, 300, 600</td>
<td>3000</td>
<td>0.5</td>
<td>4-6</td>
<td>Decrease dose by 50% in renal failure patients and patients with hepatic insufficiency</td>
</tr>
<tr>
<td>Salsalate</td>
<td>Disalcid</td>
<td>Capsule: 500 Table: 500, 750</td>
<td>3000</td>
<td>1.4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Diflunisal</td>
<td>Dolobid</td>
<td>Tablets: 250, 500</td>
<td>1500</td>
<td>2-3</td>
<td>7-15</td>
<td>Decrease dose by 50% in renal failure</td>
</tr>
<tr>
<td><strong>Acetic Acids</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diclofenac</td>
<td>Voltaren</td>
<td>Tablets: 25, 50, 75 Extended release: 100</td>
<td>225</td>
<td>1-2</td>
<td>2</td>
<td>Incidence of increased transaminase levels higher than with other NSAIDs</td>
</tr>
<tr>
<td>Diclofenac + misoprostol</td>
<td>Arthrotec</td>
<td>Tablets: 50 or 75 plus misoprostol 200 µg</td>
<td>200</td>
<td>1-2</td>
<td>2</td>
<td>Incidence of increased transaminase levels higher than with other NSAIDs</td>
</tr>
<tr>
<td>Indomethacin</td>
<td>Indocin</td>
<td>Caps: 25, 50</td>
<td>200</td>
<td>1-4</td>
<td>2-13</td>
<td>Approved for treatment of patent ductus arteriosus</td>
</tr>
<tr>
<td></td>
<td>Indocin SR</td>
<td>Sustained release: 75 Oral suspension: 25 mg/5 mL Suppositories: 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clinoril</td>
<td>Tablets: 150, 200</td>
<td>400</td>
<td>2-4</td>
<td>16</td>
<td>Prodrug metabolized to active compound Decrease dose in renal disease, liver disease, and elderly patients</td>
</tr>
<tr>
<td><strong>Salicylic Acids</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ketorolac</td>
<td>Toradol</td>
<td>IM/IV: 15 or 30 mg/mL</td>
<td>120 IV/IM</td>
<td>0.3-1</td>
<td>4-6</td>
<td>Decrease dose by 50% in renal failure patients and patients &lt; 50 kg</td>
</tr>
<tr>
<td><strong>Propionic Acids</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>Motrin Motrin</td>
<td>Tablets: 200 (OTC), 300, 400, 600</td>
<td>3200</td>
<td>1-2</td>
<td>2</td>
<td>Avoid in severe hepatic disease</td>
</tr>
<tr>
<td></td>
<td>Advil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nupren</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rufen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naproxen</td>
<td>Naprosyn</td>
<td>Tablets: 125 (OTC), 250, 375, 500 Sustained release: 375, 500 Suspension: 125 mg/5 mL</td>
<td>1500</td>
<td>2-4</td>
<td>12-15</td>
<td>Decrease dose in renal disease, liver disease, and elderly patients</td>
</tr>
<tr>
<td></td>
<td>Aleve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anaprox</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EC-Naprosyn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Naprelan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nalfon</td>
<td>Caps: 200, 300, 600</td>
<td>3200</td>
<td>1-2</td>
<td>2-3</td>
<td>Idiosyncratic nephropathy more frequent than with other NSAIDs Decrease dose in severe renal disease, hepatic disease, and elderly patients</td>
</tr>
<tr>
<td>Ketoprofen</td>
<td>Orudis</td>
<td>Tablets: 12.5 (OTC) Caps: 25, 50, 75 Sustained release: 100, 150, 200</td>
<td>300</td>
<td>0.5-2</td>
<td>2-4</td>
<td>Decrease dose in severe renal disease, hepatic disease, and elderly patients</td>
</tr>
<tr>
<td></td>
<td>Oruvail</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flurbiprofen</td>
<td>Ansaid</td>
<td>Tablets: 50, 100 1800 or 26 mg/kg/day</td>
<td>300</td>
<td>1.5-2</td>
<td>3-4</td>
<td>Decrease dose in renal failure patients and patients &lt; 50 kg</td>
</tr>
<tr>
<td></td>
<td>Oxaprozin</td>
<td>Tablets: 50, 100 Table: 600</td>
<td></td>
<td>3-6</td>
<td>49-60</td>
<td></td>
</tr>
<tr>
<td><strong>Fenamic Acids</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meclofenamate</td>
<td>Meclen</td>
<td>Caps: 50, 100</td>
<td>400</td>
<td>0.5</td>
<td>2-3</td>
<td></td>
</tr>
<tr>
<td><strong>Oxicams</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piroxicam</td>
<td>Feldene</td>
<td>Caps: 10, 20</td>
<td>20</td>
<td>2-5</td>
<td>3-86</td>
<td>Decrease dose in hepatic disease and elderly patients</td>
</tr>
<tr>
<td>Meloxicam</td>
<td>Mobic</td>
<td>Tab: 7.5, 15</td>
<td>15</td>
<td>5-6</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td><strong>Nonacidic Compounds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nabumetone</td>
<td>Relafen</td>
<td>Tablets: 500, 750</td>
<td>2000</td>
<td>3-6</td>
<td>24</td>
<td>Food increases peak concentration Decrease dose in renal disease Avoid in severe liver disease Limit dose to 1 g/day in elderly patients</td>
</tr>
</tbody>
</table>
**Table 59-1** Common Nonsteroidal Anti-inflammatory Drugs (NSAIDs)—cont’d

<table>
<thead>
<tr>
<th>Drug</th>
<th>Brand Name</th>
<th>Available Formulations (mg)</th>
<th>Maximal Daily Dose (mg)</th>
<th>Tmax (hr)</th>
<th>Half-life (hr)</th>
<th>Dose Adjustment or Special Precautions</th>
</tr>
</thead>
<tbody>
<tr>
<td>COX-2 Selective Inhibitors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Celecoxib</td>
<td>Celebrex</td>
<td>Caps: 100, 200, 400</td>
<td>400 (800 mg in FAP)</td>
<td>3</td>
<td>11</td>
<td>Contraindicated with sulfonamide allergy</td>
</tr>
<tr>
<td>Etoricoxib*</td>
<td>Arcoxia</td>
<td>Tablets: 60, 90, 120</td>
<td>120</td>
<td>1-1.5</td>
<td>22</td>
<td>Contraindicated in severe renal or liver disease patients Caution in mild-to-moderate disease</td>
</tr>
</tbody>
</table>

*Not approved by U.S. Food and Drug Administration.
FAP, familial adenomatous polyposis; IM/IV, intramuscular/intravenous; OTC, over the counter; PO, Oral.

areas that often exhibit lower pHs than uninvolved sites. Most often, there is a direct relationship between low pK and short half-life, but there are exceptions such as nabumetone, which is nonacidic. Classifying NSAIDs on the basis of plasma half-life can be problematic given the fact that these drugs tend to accumulate in synovial fluid, where the concentration of drug may remain more stable than in the plasma. Short half-life NSAIDs potentially could be given less frequently than indicated by their plasma half-life. NSAIDs exhibiting longer half-lives require more time to reach steady-state plasma levels. Drugs with a half-life greater than 12 hours can be given once or twice a day. Plasma levels increase for a few days to several weeks (depending on the specific half-life) but then tend to remain constant between doses. NSAIDs with longer half-lives also enable drug concentrations to equilibrate between the plasma and the synovial fluid, though total bound and unbound drug levels are usually lower in synovial fluid because there is less albumin in synovial fluid than in plasma. However, NSAIDs with longer half-lives or extended release formulation may be associated with increased propensity to cause adverse effects.

COX-isozyme selectivity is likely to be a critically important factor in determining relative GI and cardiovascular risk, which should also be considered in addition to other pharmacologic properties for each NSAID.

**Figure 59-4** Classification and representative structures of the traditional nonsteroidal anti-inflammatory drugs (NSAIDs) and cyclooxygenase-2 (COX-2)-selective NSAIDs. NSAIDs. *Selected NSAID structure from each subclass.
NSAID Metabolism

Almost all NSAIDs are more than 90% bound to plasma proteins. If total drug concentrations are increased beyond the point at which the binding sites on albumin are saturated, biologically active free drug concentrations increase disproportionately to the increasing total drug concentration. The clearance of NSAIDs is usually by hepatic metabolism with production of inactive metabolites that are excreted in the bile and urine. Most NSAIDs are metabolized through the microsomal cytochrome P450-containing mixed-function oxidase system. NSAIDs are most often metabolized by CYP3A, CYP2C9, or both. However, some are metabolized by other cytosolic hepatic enzymes.

Salicylate Metabolism and Aspirin Resistance

Salicylates are acetylated (e.g., aspirin) or nonacetylated (e.g., sodium salicylate, choline salicylate, choline magnesium trisalicylate, salicylsalicylic acid). Although the nonacetylated salicylates are only weak inhibitors of COX in vitro, they are able to reduce inflammation in vivo. Aspirin is rapidly deacetylated to salicylate, both spontaneously and enzymatically. Differences in formulation of these agents affect the absorption properties, but not bioavailability. Buffered aspirin tablets contain antacids that increase the pH of the microenvironment, whereas enteric coating slows absorption. The bioavailability of rectal aspirin suppositories increases with retention time. Salicylates primarily bind to albumin and rapidly diffuse into most body fluids. Salicylate is metabolized principally by the liver and excreted primarily by the kidney. In the kidney, salicylate and its metabolites are freely filtered by the glomerulus, then reabsorbed and secreted by the tubules. Salicylate serum levels usually do not correlate well with dosage, however, and small increases in dosage may result in disproportionate increases in serum levels. The drug clearance rate is a function of serum concentration. The primary factors regulating serum salicylate levels are urinary pH and metabolic enzyme activity.

The term “aspirin resistance” is broadly used to describe the failure of aspirin to prevent a thrombotic event whether due to pharmacologic resistance to the antiplatelet effects of aspirin or due to the inability of aspirin to overcome thrombophilia in a given clinical setting. Factors such as sex, genetic polymorphisms, and clinical factors including smoking, obesity, and diabetes may alter aspirin effects on platelet function. Lack of adherence and drug interactions also may play a role in aspirin resistance.

Pharmacologic Variability

Different patients can respond to the same NSAID in a variety of ways, and the basis for this individual variability remains unclear. Several pharmacologic factors related to NSAIDs may influence this variability such as dose response, plasma half-life, enantiomeric conversion, urinary excretion, and pharmacodynamic variation. Other important drug factors include protein binding, the metabolic profile of the drug, and the percentage of the drug that is available as the active (S) enantiomer. Some NSAIDs exist as two enantiomers; these include the propionic acid derivatives ibuprofen, ketoprofen, and flurbiprofen, which exist as mixtures of inactive (R) and active (S) enantiomers. Naproxen is composed of the active (S) enantiomer. Conversion of the propionic acid NSAIDs from the inactive (R) enantiomer to the active (S) enantiomer occurs in vivo to various degrees, providing some basis for the variability in patient response. There is also genetic variability in the cytochrome P450 metabolic enzymes such that some individuals or ethnic groups metabolize drugs more slowly. For example, Asians are frequently slow metabolizers through the CYP2C9 pathway. Finally, the pharmacokinetics of some NSAIDs are affected by hepatic disease, renal disease, or old age.

Routes of Drug Delivery

NSAIDs are produced in a variety of dosage forms including intravenous, slow-release and sustained-release oral preparations, topical preparations in various forms including gels and patches, and suppositories. Given the desire to reduce NSAID toxicity while preserving drug delivery to a specific site, efforts continue to alter drug formulation and delivery systems. Nanoparticles, liposomes, and microspheres are under investigation to allow dose reduction and specific targeting. Intra-articular delivery is under consideration, but because joints have efficient lymphatic clearance systems, the utility of this form of targeting remains to be proved.

Topical NSAID formulations were developed to reduce systemic exposure while preserving efficacy. Diclofenac, for example, is available as a solution, gel, or patch. The systemic effects are directly proportional to the surface area, and this method of delivery results in a relatively stable systemic diclofenac level compared with oral administration.

Combination Drugs

NSAIDs have also been combined with agents having gastroprotective effects into “polypills” that are currently available on the market. This strategy may increase compliance with effective protective agents, thereby reducing adverse effects in clinical practice. Combining diclofenac with the synthetic PGE1 analogue misoprostol (Arthrotec) is shown to reduce risk of NSAID-related peptic ulcerations and mucosal injury, but utility of the combination is often limited by misoprostol-induced cramping and diarrhea. The combination of enteric-coated naproxen and the proton pump inhibitor (PPI) esomeprazole (Vimovo) into a single pill has been approved by the U.S. Food and Drug Administration. This agent was shown to reduce endoscopically detected gastric ulcers.

A different strategy is nitric oxide releasing NSAIDs (NO-NSAIDs), which are synthesized by the ester linkage of an NO-releasing moiety to conventional NSAIDs including aspirin, flurbiprofen, diclofenac, sulindac, and others. The NO moiety is slowly released by enzymatic activity in vivo, likely by esterases, resulting in slow accumulation of the parent NSAID. The lower rate of GI ulceration
associated with these drugs is likely related to NO-associated vasodilation and the relatively lower concentration of the parent NSAID.

**THERAPEUTIC EFFECTS**

**Anti-inflammatory Effects**

NSAIDs are frequently used as first-line agents for the symptomatic relief of many different inflammatory conditions. In double-blind, randomized clinical trials of inflammatory arthritis, NSAIDs have been compared with placebo, aspirin, and each other. Clinical trials of NSAID efficacy in rheumatoid arthritis (and osteoarthritis) most often employ a design whereby the current NSAID is discontinued and the patient must have an increase in symptoms or flare to enter the study. Although there is some variation in primary outcome measures, most include parameters that make up the American College of Rheumatology (ACR)-20. Efficacy superior to that of placebo is easily demonstrated for NSAIDs within 1 to 2 weeks in patients with active RA who are not receiving corticosteroids or other anti-inflammatory medications. Comparisons of adequate doses of traditional NSAIDs or COX-2-selective NSAIDs with one another almost always show comparable efficacy. Despite improvement in pain and stiffness with NSAIDs, these agents do not usually reduce acute phase reactants, nor do they modify radiographic progression. The anti-inflammatory effects of NSAIDs have also been demonstrated in rheumatic fever, juvenile rheumatoid arthritis, ankylosing spondylitis, gout, osteoarthritis, and systemic lupus erythematosus (SLE). Although not as rigorously proven, their efficacy is also accepted in treatment of reactive arthritis, psoriatic arthritis, acute and chronic bursitis, and tendinitis.

**Analgesic Effects**

Virtually all NSAIDs relieve pain when used in doses substantially lower than those required to suppress inflammation. The analgesic action of NSAIDs is due to inhibition of PG production in peripheral tissues and in the central nervous system (CNS). In the periphery, PGs do not induce pain per se but rather sensitize peripheral nociceptors to the effects of mediators such as bradykinin or histamine. PGs released during inflammation or other lower the activation threshold of tetrodotoxin-resistant sodium channels on sensory neurons. In the CNS, where NSAIDs and acetaminophen exert analgesic effects, PGs also play an important role in neuronal sensitization. COX-2 is constitutively expressed in the dorsal horn of the spinal cord, and its expression is increased during inflammation. Centrally generated PGs activates spinal neurons and also microglia that contribute to neuropathic pain. Both COX-1 and COX-2 play a role in nociception as demonstrated by reductions in experimental pain in mice deficient in either COX-1 or COX-2.

**Antipyretic Effects**

The NSAIDs and acetaminophen effectively suppress fever in humans and experimental animals. Fever results from the production of PG, primarily PGE2, from vascular endothelial cells via COX-2 and mPGES-1. These PGs generate neuronal signals that activate the thermoregulatory center in the preoptic area of the anterior hypothalamus. PGE2 synthesis is stimulated by endogenous (e.g., interleukin-1) or exogenous (e.g., lipopolysaccharide) pyrogens. Mice with a targeted deletion of either the COX-2 or mPGES-1 genes fail to develop fever in response to inflammatory stimuli.

Little evidence suggests that any NSAID has superior efficacy as an antipyretic. However, in fever associated with viral illnesses, aspirin should be avoided due to the association with hepatocellular failure (Reye’s syndrome).

**Other Therapeutic Effects**

**Antiplatelet Effects**

Aspirin and traditional NSAIDs inhibit platelet COX-1 to variable degrees. Except for aspirin, inhibition of platelet aggregation is reversible and depends on the concentration of drug in the platelet. Aspirin acetylates platelet COX-1, which cannot be resynthesized. The antiaggregatory effect of as little as 80 mg of aspirin can last for up to 4 to 6 days, until the bone marrow can synthesize new platelets.

On the basis of accumulated data showing its benefits, the FDA has approved ASA for use in the secondary prevention of cardiovascular disease. Major trials have shown that meaningful decreases in nonfatal myocardial infarction (MI), nonfatal stroke, and death can be realized by daily administration of ASA of 75 to 325 mg. Major vascular events can be reduced by 10 to 20 events for every 1000 patients treated, at a cost of one to two major GI bleeds.

There was no reduction in rates of MI observed with the use of ASA, 100 mg every other day, in the Nurses Health Study of primary prevention of major vascular events, whereas rates of GI bleeding were increased. However, stroke rates were significantly reduced on this regimen. The U.S. Preventative Health Task Force has updated its recommendations to encourage use of low-dose ASA in men aged 45 to 79 and women aged 55 to 79. ASA used for primary prevention of cardiovascular events appears to lower the risk for MI in men and for stroke in women.

**Cancer Chemoprevention**

A large body of epidemiologic and animal studies provides evidence that a high-fat diet can be associated with a risk for cancer. AA, one of the major ingredients of animal fats, and the eicosanoids derived from AA are shown to be important contributors to cancer development. Large-scale epidemiologic studies have long indicated that long-term NSAID use reduces the incidence of a variety of cancers including colon, intestinal, gastric, breast, and bladder, 40% to 50%. Given the ability of the NSAIDs to inhibit COX and PG production, the COX pathway immediately becomes implicated as playing an important role in the pathogenic process. It is well recognized that growth factors, tumor promoters, and oncogenes stimulate PG production via the induction of COX-2 and that human tumorigenic tissues exhibit increased COX activity compared with their normal, nontumorigenic counterparts. COX-2 is overexpressed in 80% of colorectal cancer tissues.
the PGs, PGE$_2$ is most abundant in human neoplasms. The inducible mPGES-1 enzyme is highly expressed in tumors, and its absence suppresses intestinal tumorigenesis in animal models. Furthermore, the enzyme that metabolizes intracellular PGE$_2$, 15-hydroxyprostaglandin dehydrogenase, is ubiquitously lacking in tumors, and mice with a genetic deletion of this enzyme have accelerated tumorigenesis. Many natural products including resveratrol (red wine), catechins (green tea), and curcumin (safron) also inhibit COX, which may be an important mechanism underlying their putative cancer-preventing effects.

A retrospective cohort study shows that aspirin and traditional NSAIDs specifically reduce cancer risk in the subgroup of patients whose colon tumors express higher levels of COX-2. In a meta-analysis of ASA (75 mg daily and upward without dose dependence) effects on cancer, allocation to aspirin reduced death due to cancer by more than 20%. On analysis of individual patient data, cancer death benefit was apparent only after 5 years' follow-up, and benefit increased with scheduled duration of trial treatment. ASA effects appear greater on adenomatous cancer than other cancer types. Other studies demonstrated a reduction in both incidence of colorectal cancer and death from colorectal cancer, particularly for cancers of the proximal colon. Long-term low-dose ASA use also appears to reduce the risk of prostate cancer.

Clinical trials also demonstrated that traditional and COX-2-selective NSAIDs could cause regression of polyps in patients with familial adenomatous polyposis (FAP). Celecoxib was subsequently approved by the FDA for reduction of polyps in patients with FAP. Although NSAIDs are still among the most promising chemopreventative agents for cancer, cardiovascular and GI side effects have dampened enthusiasm. Alternative strategies for blocking effects of PG and other eicosanoids need evaluation to determine their effectiveness as cancer prevention and adjunct therapy.

**ADVERSE EFFECTS**

The NSAIDs share a common spectrum of clinical toxicities, although the frequency of particular side effects varies with the compound (Table 59-2). Hazard of individual NSAIDs is related to the pharmacologic characteristics such as bioavailability and half-life, as well as potency for inhibition of COX-1 and COX-2.

**Gastrointestinal Tract Effects**

NSAIDs cause GI injury through both topical and systemic effects. Following mucosal injury, NSAIDs inhibit the early events necessary to repair superficial injury, as well as later events of cell proliferation and angiogenesis leading to delayed ulcer healing. Topical injury initiates the initial mucosal erosions by disrupting the gastric epithelial cell barrier, but PG depletion is essential for the development of clinically significant gastric and duodenal ulcers. The propensity of specific NSAIDs to cause acute topical injury depends on the pKa, with most NSAIDs being weak organic acids. Aspirin is particularly prone to cause mucosal injury when administered orally. Nonacidic NSAIDs such as nabumetone, etodolac, and celecoxib do not cause acute mucosal lesions. Esterification of acidic NSAIDs such as the case for NO-NSAIDs suppresses mucosal injury.

Acidic NSAIDs accumulate in gastric epithelial cells and lead to generation of reactive oxygen species. NSAIDs may also cause apoptosis, which appears independent of COX inhibition. They also associate with extracellular phospholipids resulting in attenuation of the hydrophobic surface barrier of the stomach. Phospholipids associated with cell membranes are disrupted, leading to increased permeability and acid or bile back diffusion.

The integrity of mucosal defense depends on generation of PGE$_2$ and PGI$_2$ from COX enzymes. COX-1 is abundantly expressed under basal conditions in gastric mucosa, whereas COX-2 is almost undetectable. However, both COX-1 and COX-2 are rapidly upregulated following injury or when there are pre-existing ulcers. This may explain the observation that concurrent *Helicobacter pylori* infection increases the risk of developing peptic ulcers and increases bleeding in NSAID users. It appears that concurrent COX-1 and COX-2 inhibition is associated with the highest propensity to develop gastric ulcers. This clinical observation is consistent with animal studies whereby mice deficient for a single COX enzyme or treated with drugs that specifically inhibit either COX-1 or COX-2 do not develop ulceration. Severe gastric lesions are seen when both enzymes are simultaneously blocked. However, when gastric mucosa is damaged, inhibition of either COX-1 or COX-2 is associated with development of ulcers. Traditional and COX-2-selective NSAIDs delay ulcer healing, with nonselective NSAIDs doing so to a

---

### Table 59-2  
**Shared Toxicities of Nonsteroidal Anti-inflammatory Drugs**

<table>
<thead>
<tr>
<th>Organ System</th>
<th>Toxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastrointestinal</td>
<td>Dyspepsia, Esophagitis, Gastrroduodenal ulcers, Ulcer complications, Acute renal failure, Cardiovascular death, Myocardial infarction, Stroke, Renal</td>
</tr>
<tr>
<td></td>
<td>(bleeding, perforation obstruction)</td>
</tr>
<tr>
<td></td>
<td>Small bowel erosions and strictures, Colitis, Ulcer complications, Type IV renal tubular acidosis and hyperkalemia, Elevated transaminases, Cardiovascular death</td>
</tr>
<tr>
<td></td>
<td>Aspirin-exacerbated respiratory disease* (susceptible patients)</td>
</tr>
<tr>
<td>Hematologic</td>
<td>Rash, Cytopenias, Dizziness/confusion/drowsiness, Seizures, Aseptic meningitis</td>
</tr>
<tr>
<td>Hepatic</td>
<td>Renal toxicity, Weight gain and edema, Hypertension, Acute renal failure, Papillary necrosis, Acute interstitial nephritis, Hypertension, Myocardial infarction, Stroke, Cardiovascular death</td>
</tr>
<tr>
<td>Nervous</td>
<td>Dizziness/confusion/drowsiness, Seizures, Aseptic meningitis</td>
</tr>
<tr>
<td>Bone</td>
<td>Delayed healing</td>
</tr>
</tbody>
</table>

*Reduced risk in COX-2-selective nonsteroidal anti-inflammatory drugs.
greater degree.\textsuperscript{22} It should also be noted that GI bleeding may be related to the combination of injury and inhibition of platelet aggregation. This is an additional factor in the propensity of aspirin and other nonselective NSAIDs to cause clinically apparent ulcers.\textsuperscript{24,68}

**Dyspepsia**

Nonulcer dyspepsia is the most common adverse event (10% to 20%) associated with use of NSAIDs and may account for poor tolerability.\textsuperscript{53} Dyspepsia is more often reported in younger as compared with older patients.\textsuperscript{74} Though expected to reduce dyspepsia, COX-2-selective NSAIDs are also associated with a substantial level of adverse GI symptoms.\textsuperscript{23} PPIs have been shown to reduce dyspepsia in controlled trials.\textsuperscript{73} Studies have shown that histamine-2-receptor antagonists (H\textsubscript{2}RAs) are also effective for reducing dyspepsia.\textsuperscript{75} Crucially from a clinical perspective, subjective symptoms of dyspepsia, fecal blood loss, and endoscopic findings correlate poorly. Furthermore, only a minority of patients with serious GI events report antecedent dyspepsia.\textsuperscript{30}

**Gastritis and Gastroduodenal Ulcer**

Up to 25% of chronic NSAID users will develop ulcer disease, and 2% to 4% will bleed or perforate. These GI events result in more than 100,000 hospital admissions annually in the United States and between 7000 and 10,000 deaths, especially in those patients with highest risk.\textsuperscript{45} The risk for ulcer complications appears highest within the first 3 months of use but remains present with longer-term therapy. A recent meta-analysis of observational studies on NSAIDs and upper GI bleeding or perforation of studies published between 2000 and 2008 demonstrated a relative risk of 4.50 (95% confidence interval [95% CI], 3.82 to 5.31) for traditional NSAIDs and 1.88 (95% CI, 0.96 to 3.71) for selective COX-2 inhibitors.\textsuperscript{81} For traditional NSAIDs, the risk of low or medium doses was associated with a lower risk than higher dose. Several NSAIDs had far higher than average risk including ketorolac 14.54 (95% CI, 5.87 to 36.04) and piroxicam 9.94 (95% CI, 5.99 to 16.50). The risks for celecoxib 1.42 (95% CI, 0.85 to 2.37) and ibuprofen 2.69 (95% CI, 2.17 to 3.33) were lower than NSAIDs overall. Relative risk for other commonly used NSAIDs including naproxen, indomethacin, meloxicam, and diclofenac ranged from 3.98 to 5.63. Drugs with a long half-life or slow-release formulation were associated with higher risk even accounting for dose.\textsuperscript{41} Profound and coincident inhibition of both COX-1 and COX-2 using whole blood assay, as seen for ketorolac, piroxicam, naproxen, ketoprofen, and indomethacin, was associated with a relative risk of greater than 5 for GI bleeding and perforation.\textsuperscript{41} The use of low-dose aspirin, even in the absence of other risk factors, increases risk for bleeding and death. Many patients taking low-dose aspirin may do so without the knowledge of their physician; it is essential to query patients specifically on this point.

Other patient-specific factors influence the overall risk for GI ulcers and ulcer complications (Figure 59.5).\textsuperscript{45,57} A previous history of ulcer or ulcer complications is an important risk, especially if combined with other risks. Infection with *H. pylori* is likely to be associated with additive effect.\textsuperscript{71} It remains unclear if eradication of *H. pylori* would be useful in primary prevention of NSAID-induced ulcers, but this may be advantageous in those patients requiring long-term NSAIDs.\textsuperscript{45} Eradication alone is insufficient as a single strategy for secondary prevention of ulcer complications. This strategy appears to be most effective in reducing the bleeding risk of patients on low-dose aspirin but is less useful than use of PPIs in patients taking NSAIDs.\textsuperscript{35}

Table 59-3 provides recommendations for patients who need NSAIDs and have GI risks.\textsuperscript{45} Misoprostol has consistently been shown to be effective in reduction of gastroduodenal ulcers. Meta-analysis showed a reduction of 74% in gastric ulcers and 53% in duodenal ulcers when compared with placebo.\textsuperscript{78} The effectiveness of misoprostol is comparable with the PPI lansoprazole.\textsuperscript{79} However, the high prevalence of abdominal cramping and diarrhea limit misoprostol use at full doses. For those who do not tolerate full doses (200 mcg four times daily), lower doses of 400 to 600 mcg/day may be useful and comparable with PPIs.

PPIs have been used extensively for prevention of NSAID-induced ulcers and are also used for ulcer healing. Their excellent tolerability and availability over-the-counter have led to their dominance as pharmacologic agents for preventing NSAID-induced gastroduodenal ulcers. Studies have shown a reduction in endoscopic ulcer rate from 17% in patients taking traditional or COX-2-selective NSAIDs plus placebo to 5.2% and 4.6% in patients taking NSAIDs plus esomeprazole 20 mg or 40 mg, respectively.\textsuperscript{80} As noted previously, a combination pill containing naproxen and

---


---

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSAIDs (including aspirin)</td>
<td>8.9 (4.6–17.0)</td>
</tr>
<tr>
<td>High-dose NSAID</td>
<td>6.4 (2.4–17.0)</td>
</tr>
<tr>
<td>Anticoagulant treatment</td>
<td>5.6 (2.3–13.5)</td>
</tr>
<tr>
<td>Previous uncomplicated ulcer</td>
<td>3.5 (1.5–8.0)</td>
</tr>
<tr>
<td>Age 70–80 years</td>
<td></td>
</tr>
<tr>
<td>Helicobacter pylori infection</td>
<td></td>
</tr>
<tr>
<td>Oral corticosteroids</td>
<td>2.2 (1.0–4.5)</td>
</tr>
</tbody>
</table>
esomeprazole has been approved for use. It may reduce non-compliance but will be associated with higher cost.

High-dose, twice-daily doses of H$_2$RA reduce the risk of NSAID-induced endoscopic ulcers and are the least costly alternative. However, these agents are inferior to PPIs and, like PPIs, there are no randomized clinical outcome trials that evaluate the efficacy of H$_2$RAs in chronic NSAID users.\textsuperscript{14}

**Esophageal Injury**

Aspirin and NSAIDs are associated with esophagitis related to mechanisms similar to those in the gastric mucosa.\textsuperscript{81,82} Esophageal emptying may be slowed in the elderly, resulting in a prolonged exposure of the mucosa to the irritant action of aspirin and NSAIDs. Gastroesophageal reflux may be an aggravating factor and lead to stricture formation. Bleeding may also complicate esophagitis. NSAIDs should be prescribed with caution in the presence of gastroesophageal reflux disease.

**Small Bowel Injury**

The availability of video capsule endoscopy (VCE) and balloon enteroscopy has advanced the ability to detect small intestinal lesions in patients taking NSAIDs. NSAIDs can cause a concentric “diaphragm-like” stricture in the small bowel in addition to causing mucosal injury and bleeding. Two recent studies of patients on NSAIDs for at least 3 months using VCE demonstrated a prevalence of 70% to 80% for small bowel injuries.\textsuperscript{83} Furthermore, NSAID-induced small bowel injury is likely a common cause of obscure GI bleeding. NSAIDs that undergo enterohepatic circulation are likely to be associated with higher risk. Small bowel injury may be detected by anemia or symptoms of obstruction related to stricture.\textsuperscript{83} Strategies effective for gastroduodenal ulcers such as misoprostol or certain PPIs may also reduce the risk for small bowel mucosal injury. Strictures may require balloon endoscopy or surgical intervention.\textsuperscript{83}

### Colitis

NSAIDs cause erosions, ulcers, hemorrhage, perforations, strictures, and complications of diverticulosis in the large bowel.\textsuperscript{84} NSAID-induced injury is more common in the right colon (80%) but can occur in the transverse and left colon. NSAID-containing suppositories can cause erosions, ulcers, and stenoses in the rectum. NSAID colonopathy is in the differential diagnosis of inflammatory bowel disease. Patients with NSAID-induced colonopathy are typically older, and the erosions are more likely to be transverse or circular.\textsuperscript{85} There is also a concern that treatment with traditional and COX-2-selective NSAIDs may exacerbate inflammatory bowel disease.\textsuperscript{86} NSAIDs are also implicated in the development of collagenous colitis.\textsuperscript{87}

### Renal Effects

PGs play a vital role in solute and renovascular homeostasis.\textsuperscript{88} It is becoming quite clear that PGs are produced by both COX-1 and COX-2, generally in different locations within the kidney, and that these PGs may play different physiologic roles in renal function.\textsuperscript{89,90} COX-1 is highly expressed in the renal vasculature, glomerular mesangial cells, and collecting duct. COX-2 expression is restricted to the vasculature, cortical thick ascending limb (specifically in cells associated with the macula densa), and in medullary interstitial cells. COX-2 expression in the macula densa increases in high-renin states (e.g., salt restriction, angiotensin-converting enzyme inhibition, renovascular hypertension), and selective COX-2 inhibitors significantly decrease plasma renin levels and renal renin activity. COX-2 expression in the macula densa is reduced by angiotensin II and mineralocorticoids. Dehydration or hypertonicity appears to regulate COX-2 expression in the medullary interstitium. COX-2 is also necessary for normal renal development.

### Sodium Excretion

PGs are known to regulate renal sodium resorption by their ability to inhibit active transport of sodium in both the thick ascending limb and the collecting duct and to increase renal water excretion by blunting the actions of vasopressin.\textsuperscript{91} The cellular source of COX-2-derived prostanoids that promote natriuresis remains uncertain, but it is possible that they may in large part be derived from the medullary interstitial cells. Sodium retention has been reported to occur in up to 25% of NSAID-treated patients and may be particularly apparent in patients who have an existing avidity for sodium, such as those with mild heart failure or liver disease.\textsuperscript{91} Decreased sodium excretion in NSAID-treated patients can lead to weight gain and peripheral edema. This effect may be sufficiently important to cause clinically important exacerbations of congestive heart failure.

### Hypertension

NSAIDs may cause altered blood pressure, with average increases of mean arterial pressure of between 5 and 10 mm Hg. In addition, using NSAIDs may increase the risk of

### Table 59-3 Strategies for Gastrointestinal Risk Reduction\textsuperscript{85}

<table>
<thead>
<tr>
<th>Gastrointestinal Risk</th>
<th>Potential Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low risk</td>
<td>Intermittent NSAID use</td>
</tr>
<tr>
<td>Moderate risk (1-2 risk factors)</td>
<td>Low-dose NSAID</td>
</tr>
<tr>
<td>Age &gt; 65</td>
<td>Intermittent NSAID use</td>
</tr>
<tr>
<td>High-dose NSAID</td>
<td>NSAID + PPI</td>
</tr>
<tr>
<td>Previous history of uncomplicated ulcer</td>
<td>NSAID + misoprostol</td>
</tr>
<tr>
<td>Concurrent use of aspirin, corticosteroids, or anticoagulants</td>
<td>NSAID + high-dose H$_2$RA*</td>
</tr>
<tr>
<td>High risk&gt;2 Risk factors</td>
<td>Alternative treatment</td>
</tr>
<tr>
<td>History of previous complicated ulcer, especially recent</td>
<td>COX-2-selective NSAID + PPI</td>
</tr>
<tr>
<td>Helicobacter pylori positive</td>
<td>Consider eradication in moderate- to high-risk patients</td>
</tr>
</tbody>
</table>

*Less effective than PPI or misoprostol. COX-2, cyclooxygenase-2; H$_2$RA, histamine-2-receptor antagonist; NSAID, nonsteroidal anti-inflammatory drug; PPI, proton pump inhibitor.
initiating antihypertensive therapy in older patients, with the magnitude of increased risk being proportional to the NSAID dose. Furthermore, in a large (n = 51,630) prospective cohort of women aged 44 to 69 without hypertension in 1990, incident hypertension over the following 8 years was significantly more likely in frequent users of aspirin, acetaminophen, and NSAIDs. NSAIDs can attenuate the effects of antihypertensive agents including diuretics, angiotensin-converting enzyme inhibitors, and β-blockers, interfering with blood pressure control.

PGs stimulate renin release which, in turn, increases secretion of aldosterone and, subsequently, potassium secretion by the distal nephron. For this reason, NSAID-treated patients may develop hyporeninemic hypoaldosteronism that manifests as type IV renal tubular acidosis and hyperkalemia. The degree of hyperkalemia is generally mild; however, patients with renal insufficiency or those that may otherwise be prone to hyperkalemia (e.g., patients with diabetes mellitus and those on angiotensin-converting enzyme inhibitors or potassium-sparing diuretics) may be at greater risk.

**Acute Renal Failure and Papillary Necrosis**

Acute renal failure is an uncommon consequence of NSAID treatment. This is due to the vasoconstrictive effects of NSAIDs and is reversible. In most cases, renal failure occurs in patients who have a depleted actual or effective intravascular volume (e.g., congestive heart failure, cirrhosis, renal insufficiency).

Marked reduction in medullary blood flow may result in papillary necrosis that may arise from apoptosis of medullary interstitial cells. Inhibition of COX-2 may be a predisposing factor.

**Interstitial Nephritis**

Another adverse renal effect resulting from NSAIDs involves an idiosyncratic reaction accompanied by massive proteinuria and acute interstitial nephritis. Hypersensitivity phenomena such as fever, rash, and eosinophilia may occur. This syndrome has been observed with most NSAIDs.

**Chronic Kidney Disease**

Use of analgesics, particularly acetaminophen and aspirin, has been associated with nephropathy leading to chronic renal failure. In one large case-control study, the regular use of aspirin or acetaminophen was associated with a risk of chronic renal failure 2.5 times as high as that for nonuse, and the risk increased significantly with an increasing cumulative lifetime dose. In subjects regularly using both acetaminophen and aspirin, the risk was also significantly increased compared with users of either agent alone. No association between the use of nonaspirin NSAIDs and chronic renal failure could be detected after adjusting for acetaminophen and aspirin use. Pre-existing renal or systemic disease was a necessary precursor to analgesic-associated renal failure and those without pre-existing renal disease had only a small risk of end-stage renal disease.

**Cardiovascular Effects**

The risk of adverse cardiovascular effects associated with NSAID use was not widely appreciated until COX-2-selective NSAIDs were introduced into clinical practice. Rofecoxib, a potent, highly specific COX-2 inhibitor with a long half-life, was shown to have a substantially increased risk of MI and stroke and removed from the market because of this adverse effect. The mechanisms for cardiovascular risks associated with all NSAIDs are likely related to an imbalance between complete inhibition of COX-1 and COX-2 across the dosing interval. The COX-1 isoenzyme is responsible for generation of platelet TXA in platelets and aggregates and thrombus formation. In order to inhibit this activity, COX-1 must be inhibited by 95% or greater. Antithrombotic PGI synthesized by endothelial COX-2 is inhibited almost completely by both traditional and COX-2-selective NSAIDs. It is proposed that the relationship between excess cardiovascular risk for all NSAIDs, not only COX-2-selective NSAIDs, is related to the degree of COX-2 inhibition absent complete inhibition of COX-1.

Investigators have shown that drugs that inhibit COX-2 less than 90% at therapeutic concentrations in the whole blood assay present a relative risk for MI of 1.18 (95% CI, 1.02 to 1.38), whereas drugs that inhibit COX-2 to a greater degree present a relative risk of 1.60 (95% CI, 1.41 to 1.81). Relative inhibition of the COX isoforms is not the only mechanism that contributes to cardiovascular hazard. Other actions of NSAIDs including effects on blood pressure, endothelial function and nitric oxide production, and other renal effects may play a role in cardiovascular risks. Multiple analyses have demonstrated that the risk for cardiovascular hazard is significantly higher in those with pre-existing coronary artery disease. Some NSAIDs, notably ibuprofen and naproxen, may interfere with the irreversible inhibition of platelet COX-1 by aspirin, thereby increasing cardiovascular hazard in aspirin users.

A number of large-scale randomized controlled trials comparing NSAIDs with placebo or with each other have been performed and analyzed to determine the risk of MI, stroke, cardiovascular death, death from any cause, and Antiplatelet Trialists' Collaboration (APTC) composite outcomes. Because event rates in most of these studies were low, uncertainty regarding absolute and relative risk remains. For example, there were only 554 MIs in aggregate across all trials included in the most comprehensive analysis to date. Nevertheless, it appears from analyses of these aggregated clinical trials that all traditional and COX-2-selective NSAIDs except naproxen carry an excess risk of more than 30% compared with placebo. Pairwise comparisons of the most commonly used traditional and COX-2-selective NSAIDs studied in clinical trials also suggest that naproxen may have lower cardiovascular risk. One meta-analysis explored the effects of dose and dosing regimen in a pooled analysis of six randomized placebo-controlled trials of celecoxib. Lower doses and once-daily regimens were associated with lower relative risks for the APTC outcomes. This finding confirms data from other studies that suggest avoiding continuous interference with PG biosynthesis is associated with lower cardiovascular risk.
Because clinical trials have been underpowered to specifically address the relative cardiovascular risk of NSAIDs, investigators have turned to observational datasets. Using a large observational database with 8852 cases of nonfatal MI, a recent case-control study also identified a 35% increase in the risk of MI while using NSAIDs. This type of study also identifies naproxen as potentially having a lower risk. In this analysis, long half-life was an independent predictor of MI hazard. The effect of dose and slow-release formulation demonstrated that risk was a direct consequence of prolonged drug exposure. It appears that the risk associated with these pharmacologic factors may be even more important than COX-2 specificity for most NSAIDs.

A number of strategies have been suggested to mitigate cardiovascular risks associated with NSAID use (Table 59-4). These recommendations take into account a patient’s underlying risk, aspirin use, and the interaction between NSAIDs. In addition, the specific choice of NSAID should consider its pharmacologic properties.

### Table 59-4 Strategies for Reducing Cardiovascular Risk

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>If using aspirin, take aspirin dose ≥ 2 hr before NSAID dose*</td>
<td>Especially ibuprofen. Celecoxib does not appear to interfere with aspirin actions.</td>
</tr>
<tr>
<td>Do not use NSAIDs within 3-6 mo of an acute cardiovascular event or procedure</td>
<td>NSAID, nonsteroidal anti-inflammatory drug.</td>
</tr>
<tr>
<td>Carefully monitor and control blood pressure</td>
<td></td>
</tr>
<tr>
<td>Use low-dose, short half-life NSAIDs and avoid extended-release formulations</td>
<td></td>
</tr>
</tbody>
</table>

**Heart Failure**

NSAIDs are associated with reduced sodium excretion, volume expansion, increased preload, and hypertension. As a result of these properties, patients with pre-existing heart failure are at risk of decompensation with a relative risk of 3.8 (95% CI, 1.1 to 12.7). After adjusting for age, sex, and concomitant medication, the relative risk was 9.9 (95% CI, 1.7 to 57.0). Studies disagree as to whether NSAIDs are a risk for new heart failure, although the elderly may be at particular risk.

**Closure of the Ductus Arteriosus**

The maintenance of an open ductus arteriosus and its closure during the postnatal period are regulated by PG. COX-1, COX-2, and EP2-deficient mice die from neonatal circulatory failure because the ductus arteriosus remains open. It is inadvisable for pregnant women to take NSAIDs during the last trimester of pregnancy because of the risk of a persistently patent ductus arteriosus.

**Hepatic Effects**

Small elevations of one or more liver tests may occur in up to 15% of patients taking NSAIDs, and notable elevations of alanine aminotransferase or aspartate aminotransferase (=three or more times the upper limit of normal) have been reported in approximately 1% of patients in clinical trials of NSAIDs. Patients usually have no symptoms, and discontinuation or dose reduction generally results in normalization of the transaminase values, although rare, fatal outcomes have been reported with almost all NSAIDs. Those NSAIDs that appear most likely to be associated with hepatic adverse events are diclofenac and sulindac.

In clinical trial reports to the FDA, 5.4% of patients with rheumatoid arthritis who were treated with aspirin experienced persistent elevations of results in more than one liver function test. In children with viral illnesses, hepatocellular failure and fatty degeneration (Reye’s syndrome) are associated with aspirin ingestion.

### Asthma and Allergic Reactions

#### Asthma

Up to 10% to 20% of the general asthmatic population, especially those with the triad of vasomotor rhinitis, nasal polyposis, and asthma, are hypersensitive to aspirin. In these patients, ingestion of aspirin and nonspecific NSAIDs leads to severe exacerbations of asthma with naso-ocular reactions. Formerly termed aspirin-sensitive asthma, these patients are now characterized as having aspirin-exacerbated respiratory disease (AERD) because they have chronic upper and lower respiratory mucosal inflammation, sinusitis, nasal polyposis, and asthma independent of their hypersensitivity reactions. It is now thought that production of protective PGs in the setting of AERD is derived from a COX-1. A number of studies have been reported demonstrating that the COX-2-specific NSAIDs, rofecoxib and celecoxib, fail to trigger asthma exacerbation or naso-ocular symptoms in patients with AERD. Nevertheless, these studies were performed as challenge tests rather than long-term placebo-controlled trials, and caution is advised. The fact that specific COX-2 inhibitors appear safe in AERD does not imply that other hypersensitivity reactions do not occur.

#### Allergic Reactions

A wide variety of cutaneous reactions have been associated with NSAIDs. Almost all the NSAIDs have been associated with cutaneous vasculitis, erythema multiforme, Stevens-Johnson syndrome, or toxic epidermal necrolysis. NSAIDs are also associated with urticaria/angioedema and anaphylactoid or anaphylactic reactions. Special note should be made that celecoxib and valdecoxib contain a sulfonamide group and should not be given to patients who report allergy to sulfa-containing drugs.

#### Hematologic Effects

Aplastic anemia, agranulocytosis, and thrombocytopenia are rarely associated with NSAIDs, but they are prominent among the causes of deaths attributed to these drugs. Because of the risk of hematologic effects, phenylbutazone is no longer recommended for use in any condition in the United States and has been taken off the market.

### Effects on the Immune System

Virtually all cell types composing the immune system produce and respond to PGs. PGE2 is a potent inhibitor of
chemotaxis, aggregation, superoxide production, lysosomal enzyme release, and LTB4 generation of activated neutrophils.109 Antigen-presenting cells play a pivotal role in the immune response, bridging local innate immune responses and the developing acquired immune response, which takes place in more specialized lymphoid organs. Bone marrow cultured in the presence of indomethacin yields increased numbers of dendritic cells, whereas bone marrow cultured in the presence of exogenous PGE2 yields lower numbers of dendritic cells.109 Dendritic cell cytokines are one of the many important factors that polarize the T cell response toward T-helper (TH)1, TH2, or TH17 cytokine profiles. Failure of dendritic cell maturation and/or their secretion of inhibitory cytokines leads to development of T regulatory cells that modulate T helper activity and suppress some of their functions.110 Eicosanoids are likely to be one of the tissue factors that determine dendritic cell phenotype during interaction with specific antigens/pathogens via pattern recognition receptors.110 Exogenous PGE2 was reported to induce IL-23 production, important for generation of TH17 cells.111,112 A recent study using human peripheral blood mononuclear cells also supports a role for PGE2 in promoting a TH17 response.111 Although a TH17 immune response is associated with development of autoimmune diseases, the clinical relevance of these observations with respect to NSAID use remains unclear.

PGE2 also plays a role in B and T lymphocytes. PGE2, through increased cAMP, inhibits many T cell functions.113 Treatment with COX-2-selective NSAIDs severely diminishes proliferation and expression of IL-2, TNF, and IFN-γ. B lymphocytes express COX-2 and produce PGE2.114 Treatment with traditional and COX-2-selective NSAIDs profoundly reduced production of IgG and IgM after stimulation with minimal effects on proliferation. COX-2 null mice had 64% less IgM and 35% less IgG than normal littermates following in vitro treatment with LPS. COX-2-deficient mice also produce markedly reduced IgG and 10-fold reduced neutralizing antibody to HPV-like particles compared with WT mice without evidence of differences in B cell precursors.116 Similar findings of deficient humoral immune responses have been demonstrated for mice deficient in mPGES-1.117 Human memory B cell antibody production was also diminished in the presence of a specific COX-2 inhibitor.116 Again, the clinical relevance of these observations remains unclear with respect to NSAID use.

Central Nervous System Effects

Elderly patients may be particularly susceptible to developing cognitive dysfunction and other CNS effects including headache, dizziness, depression, hallucination, and seizures related to NSAIDs. Acute aseptic meningitis has been reported in patients with SLE or mixed connective tissue disease treated with ibuprofen, sulindac, tolmetin, or naproxen.

Effects on Bone

The complex effects of prostanoids on bone formation and remodeling have been appreciated for many years. It is now clear that COX-2 is required for many functions of both osteoblasts and osteoclasts.27 COX-2 is rapidly inducible and highly expressed and regulated in osteoblasts. Parathyroid hormone (PTH) is a strong inducer of COX-2. The production of PG by osteoblasts is an important mechanism for the regulation of bone turnover.27 The major effect of PGE2 is considered to occur indirectly via upregulation of receptor activator of NFκB ligand (RANKL) expression and by inhibition of osteoprotegerin (OPG) expression in osteoblastic cells, which facilitated osteoclastogenesis. Genetic deletion of PTGS2 or COX-2-selective NSAIDs partially block the PTH- or 1,25-OH vitamin D-induced formation of osteoclasts in organ cultures. Recently, a familial disorder, primary idiopathic hypertrophic osteoarthropathy, was found to be associated with a mutation in the enzyme 15-hydroxyprostaglandin dehydrogenase, the enzyme that inactivates PGE2.118 These patients have chronically elevated PGE2 levels and digital clubbing with evidence of increased bone formation and resorption in the phalanges.

The role of endogenous PG and NSAIDs in skeletal pathology remains complex. LPS-induced bone loss can be ameliorated in mice lacking mPGES-1.119 Inflammatory bone loss likely reflects increased bone resorption and decreased bone formation.21 It has long been appreciated that NSAIDs can inhibit experimental fracture healing and reduce formation of heterotopic bone in patients.120 Furthermore, fracture healing is impaired in rats treated with specific COX-2 inhibitors and in mice with genetic deletion of the COX-2 gene.121 Given the effectiveness of NSAIDs as analgesics, it is important to understand the clinical concern regarding impaired fracture healing and NSAIDs. A recent meta-analysis found a pooled odds ratio for nonunion with NSAID exposure of 3 (95% CI, 1.6 to 5.6).122 However, there was a significant association between lower-quality studies, and higher reported odds ratios for nonunion was observed. When only higher-quality studies were considered, no statistically significant association between NSAID exposure and nonunion was identified.

The impact of NSAIDs on bone mineral density (BMD) also remains unclear.123 In older men, daily use of COX-2-selective NSAIDs was associated with lower hip and spine BMD compared with nonusers, but in postmenopausal women not taking hormone replacement therapy there was a higher BMD.124 It is speculated that the beneficial effects of mechanical loading may be reduced by COX-2 inhibition but that the proinflammatory state and increased bone turnover associated with estrogen withdrawal may be suppressed by COX-2 inhibition. Still, a causal role for endogenous PGs in bone loss resulting from estrogen deficiency has not been confirmed.

Effects on Ovarian and Uterine Function

PGs derived from COX-2 have been implicated as mediators in multiple stages of the female reproductive cycle. Induction of COX-2 immediately after the luteinizing hormone surge was the first observation involving the iso-enzyme during a normal physiologic event. It has been suggested that COX-2-derived PGs may signal the time of ovulation in mammals.124,125 Studies using COX-2 null mice show reproductive failure at ovulation, fertilization, implantation, and decidualization.126 COX-2-dependent prostanoid production probably leads to the generation of
proteolytic enzymes that rupture the follicles. After fertilization, COX-2 also plays a role in embryo implantation in the myometrium. PGs are important for inducing uterine contractions during labor. Murine studies have shown that the mechanism of uterine contraction involves fetal release of PGF2α, a compound that induces luteolysis. This pathway leads to reduced maternal progesterone levels, induction of oxytocin receptors in the myometrium, and parturition.

Based on these observations in animals, one could hypothesize that NSAIDs may have an influence on fertility. In fact, studies suggest that luteinized unruptured follicle syndrome as a cause of reversible infertility can be related to ingestion of NSAIDs. For this reason, although it remains unproved in controlled or observational studies, women should be cautioned that chronic NSAID use may impair fertility.

Salicylate Intoxication and NSAID Overdose

The new appearance of tachypnea, confusion, ataxia, oliguria, or a rising blood urea nitrogen (BUN)/creatinine in a patient, particularly an elderly patient, taking aspirin or salicylates should suggest the possibility of salicylate intoxication. In adults, metabolic acidosis is masked by hyperventilation due to stimulation of respiratory centers, which is a direct effect of salicylates. Sudden increases in salicylate levels can occur even if there is no change in dose. This is particularly common in patients who develop acidosis from any cause, suffer from dehydration, or ingest other drugs that displace salicylate from protein-binding sites. Therapy consists of removing residual drug from the GI tract, forced diuresis while maintaining the urinary pH in the alkaline range and with potassium replacement, or hemodialysis if diuresis is unsatisfactory. Vitamin K is recommended because large doses of salicylate may interfere with the synthesis of the vitamin K–dependent clotting factors.

Acute overdoses of NSAIDs are much less toxic than are overdoses of aspirin or salicylates. This subject has been most carefully evaluated for ibuprofen, prompted by its approval for over-the-counter sale to the general public. Symptoms with overdoses ranging up to 40 g include CNS depression, seizures, apnea, nystagmus, blurred vision, diplopia, headache, tinnitus, bradycardia, hypotension, abdominal pain, nausea, vomiting, hematuria, abnormal renal function, coma, and cardiac arrest. Treatment includes prompt evacuation of the stomach contents, observation, and administration of fluids.

Adverse Effects of Acetaminophen

Acetaminophen is used widely as the first-line treatment of pain, chiefly because it is viewed as effective and safer than NSAIDs. Used in doses below 2 g daily, there is little evidence of toxicity. Acetaminophen-induced acute liver failure is due to direct injury from the toxic metabolite, N-acetyl-p-benzoquinoneimine, a highly reactive electrophilic compound that depletes glutathione and subsequently accumulated in hepatocytes. Acetaminophen is a highly predictable hepatotoxin with a threshold dose of 10 to 15 g in adults and 150 mg/kg in children. In the United States, acetaminophen overdoses are usually unintentional, with most taking acetaminophen preparations for chronic pain. Intentional self-poisoning with acetaminophen also remains an important problem. Treatment of acetaminophen overdose includes gastric lavage, activated charcoal, or induction of vomiting within the first 3 hours of injection. In addition, intensive support measures and early treatment with N-acetylcysteine, which replenished glutathione, have reduced mortality associated with acute acetaminophen toxicity. With high doses of acetaminophen, other toxicities may occur including GI ulcers and bleeding. Regular use of acetaminophen has also been associated with an increased risk for chronic renal failure.

**EFFECTS OF CONCOMITANT DRUGS, DISEASES, AND AGING**

Because of the widespread use of prescription and nonprescription NSAIDs, there are ample opportunities for interaction with other drugs and for interactions with patient-specific factors. Specific drug interactions are listed on the package inserts of individual agents.

**Drug-Drug Interactions**

Because most NSAIDs are extensively bound to plasma proteins, they may displace other drugs from binding sites or may themselves be displaced by other agents. Aspirin and other NSAIDs may increase the activity or toxicity of sulfonamide, hypoglycemic agents, oral anticoagulants, phenytoin, sulfonamides, and methotrexate by displacing these drugs from their protein-binding sites and increasing the free fraction of the drug in plasma. NSAIDs may blunt the antihypertensive effects of β-blockers, angiotensin-converting enzyme inhibitors, and thiazides leading to destabilization of blood pressure control. There is an increased risk of GI toxicity when NSAIDs and selective serotonin reuptake inhibitors are taken concomitantly compared with taking either agent alone and more than an additive risk.

There are interactions between aspirin and NSAIDs, particularly ibuprofen, related to blocking the ability of aspirin to access the COX active site. This may be important when aspirin is used for prevention of cardiovascular disease. It is prudent to recommend that aspirin be taken 2 hours before ibuprofen dosing.

**Drug-Disease Interactions**

Rheumatoid arthritis and other diseases (e.g., hepatic and renal disease) that decrease serum albumin concentrations are associated with increased concentrations of free NSAIDs. Hepatic and renal diseases may also impair drug metabolism or excretion and thereby increase the toxicity of a given dose of NSAID to an individual patient. Renal insufficiency may be accompanied by accumulated endogenous organic acids that may displace NSAIDs from protein-binding sites.

**Drug Reactions in the Elderly**

Aging is accompanied by changes in physiology resulting in altered pharmacokinetics and pharmacodynamics. Decreased
drug clearance may be the consequence of reductions in hepatic mass, enzymatic activity, blood flow, renal plasma flow, glomerular filtration rate, and tubular function associated with aging. The elderly are more likely to experience adverse GI and renal effects related to NSAIDs. The increased risk of cardiovascular disease in elderly patients raises concerns of accelerated MI or stroke. The use of aspirin for prevention of cardiovascular disease increases the toxicity of NSAIDs, and conversely the concomitant use of NSAIDs may increase aspirin resistance. Use of PPI for gastroprotection may interfere with the efficacy of antplatelet agents such as clopidogrel.\textsuperscript{133} The elderly have more illnesses than younger patients and therefore take more medications, increasing the possibility of drug-drug interactions. Older patients may also be more likely to self-medicate or make errors in drug dosing. For these reasons, frequent monitoring for compliance and toxicity should be a part of the use of NSAIDs in this population.

**COLCHICINE**

Colchicine is frequently used in the context of NSAID replacement therapy when the latter are contraindicated. Colchicine exhibits excellent anti-inflammatory activity in acute gouty arthritis. A recent study has clarified that a low-dose colchicine regimen of 1.2 mg followed in 1 hour by 0.6 mg provides equal efficacy to higher-dose regimens with a marked reduction in adverse events.\textsuperscript{136} Colchicine is also used to prevent acute gouty attacks. Prophylactic treatment appears to reduce the frequency of attacks by 75% to 85% and mitigates the severity of attacks that occur.\textsuperscript{137} However, there is concern that prophylactic therapy should be initiated only if hyperuricemia is controlled because tophi may develop without the usual warning signs of acute gouty attacks. Albeit with less efficacy, colchicine treatment may also benefit patients with acute episodes of pseudogout and arthritis due to other crystals.

Daily colchicine (1.2 to 1.8 mg) is the mainstay of treatment for familial Mediterranean fever (FMF). It is effective in preventing acute attacks and amyloidosis. Colchicine has been used empirically in other rheumatologic disorders, where neutrophils play an important role such as Behçet’s disease, recurring pericarditis, and cutaneous neutrophilic vasculitis.\textsuperscript{138}

**Mechanism of Colchicine Action**

Colchicine appears to interfere with steps of the inflammatory response in which neutrophils play a central role by interfering with the organization of the fibrillar microtubules involved in cell morphology and movement. This leads to disaggregation of microtubules and to decreased neutrophil motility and chemotaxis. Furthermore, colchicine inhibits release of chemotactic factors (e.g., leukotriene B\textsubscript{3}), formation of digestive vacuoles, and lysosomal degranulation. This results in an inhibition of neutrophil migration into an area of inflammation and a reduction of the metabolic and phagocytic activity of the neutrophils already present. Clinically, this results in an interruption in the inflammatory process of gout and other neutrophildominated acute inflammatory diseases.\textsuperscript{136}

**Adverse Effects of Colchicine**

Because the mechanism of colchicine actions is different from NSAIDs, the adverse event profile is also different. More than 80% of patients who take a traditional high-dose oral therapeutic dose of colchicine for an acute gouty attack experience cramping, abdominal pain, diarrhea, nausea, or vomiting, and these symptoms usually limit the dose. For this reason, and because of equal efficacy, high-dose colchicine is not recommended for acute gout flare.\textsuperscript{136}

Bone marrow depression, hair loss, amenorrhea, dysmenorrhea, oligospermia, and azoospermia have been reported with chronic colchicine treatment.\textsuperscript{138} There may be an increase in trisomy 21 in the offspring of FMF patients taking colchicine at the time of conception. Colchicine can cause subacute-onset muscle and peripheral nerve toxicity, particularly in patients with chronic renal failure. For this reason, colchicine should be used cautiously in patients with chronic renal failure and dose adjustment should be considered. Patients with colchicine-induced neuromyopathy present with proximal muscle weakness, elevated serum creatine kinase (CK) levels, and neuropathy and/or myopathy on electromyography (EMG).\textsuperscript{139}

Death has occurred with as little as 8 mg of colchicine but is inevitable after the ingestion of more than 40 mg. Treatment includes aspiration of the stomach, intensive support measures, and hemodialysis, although there is no specific evidence that colchicine can be removed by dialysis.

**CHOOSING ANTI-INFLAMMATORY ANALGESIC THERAPY**

In choosing an NSAID for a particular patient, the clinician must consider efficacy, potential toxicity related to concomitant drugs and patient factors, and cost. Furthermore, patient preference for factors such as dosing regimen may be taken into account. In addition to choices from the perspective of the individual patient and physician, it may be important to take a broader view. Choice of anti-inflammatory analgesic therapy can also be considered from the perspective of health care institutions and payers. The symptoms and conditions for which NSAIDs are used are extraordinarily common. Consequently, the cost of NSAIDs as a proportion of total drug costs can be high when drugs are expensive. The increased cost of branded NSAIDs has an important pharmacoeconomic impact. On the other hand, adverse events can have important economic consequences, and improved safety may be cost-effective.

Choosing anti-inflammatory analgesic therapy has become increasingly complex with the increased understanding of associated toxicities. Prospectively considering the presence of GI and cardiovascular risk factors is essential when considering treatment options (Table 59.5). GI risks are well known, and strategies to prevent ulceration and bleeding are available. Many questions regarding the risk for cardiovascular events in patients using NSAIDs exist. In general, the data suggest that physicians should be cautious in using NSAIDs in patients with known cardiovascular disease. In those patients with risks for
Table 59-5  Choosing Analgesic Anti-inflammatory Therapy

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>Treatment Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low:</td>
<td></td>
</tr>
<tr>
<td>&lt;65 yr old</td>
<td>Traditional NSAID</td>
</tr>
<tr>
<td>No cardiovascular risk factors</td>
<td>Shortest duration and lowest dose possible</td>
</tr>
<tr>
<td>No requirement for high-dose or chronic therapy</td>
<td></td>
</tr>
<tr>
<td>No concomitant aspirin, corticosteroids, or anticoagulants</td>
<td></td>
</tr>
<tr>
<td>Intermediate:</td>
<td></td>
</tr>
<tr>
<td>≥65 yr old</td>
<td>Traditional NSAID + PPI, misoprostol, or high-dose</td>
</tr>
<tr>
<td>No history of previous complicated GI ulceration, may be using aspirin for primary prevention</td>
<td>H,RA</td>
</tr>
<tr>
<td>Requirement for chronic therapy and/or high-dose therapy</td>
<td>Once-daily celecoxib + PPI, misoprostol, or high-dose</td>
</tr>
<tr>
<td>High:</td>
<td></td>
</tr>
<tr>
<td>Elderly, especially if frail or if hypertension, renal, or liver disease present</td>
<td>Use acetaminophen &lt;2 g/day</td>
</tr>
<tr>
<td>History of previous complicated ulcer or multiple GI risk factors</td>
<td>Avoid chronic NSAIDs if at all possible:</td>
</tr>
<tr>
<td>History of cardiovascular disease and on aspirin or other antplatelet agent for secondary prevention</td>
<td>Use intermittent NSAID dosing</td>
</tr>
<tr>
<td>History of heart failure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use low-dose, short half-life NSAIDs</td>
</tr>
<tr>
<td></td>
<td>Do not use extended-release NSAID formulation</td>
</tr>
<tr>
<td></td>
<td>If chronic NSAID required, consider:</td>
</tr>
<tr>
<td></td>
<td>Once daily celecoxib + PPI/ misoprostol (GI &gt; CV risk)</td>
</tr>
<tr>
<td></td>
<td>Naproxen + PPI/misoprostol (CV &gt; GI risk)</td>
</tr>
<tr>
<td></td>
<td>Avoid PPI if using antplatelet agent such as clopidogrel</td>
</tr>
<tr>
<td></td>
<td>Monitor and treat blood pressure</td>
</tr>
<tr>
<td></td>
<td>Monitor creatinine and electrolytes</td>
</tr>
</tbody>
</table>

CV, cardiovascular; GI, gastrointestinal; H,RA, histamine-2-receptor antagonist; NSAID, nonsteroidal anti-inflammatory drug; PPI, proton pump inhibitor.

NSAID toxicity, avoiding potent drugs with a long half-life or extended-release formulations is prudent. Intermittent dosing rather than continuous daily use reduces toxicity.

Absence of anti-inflammatory activity reduces the effectiveness of acetaminophen for diseases accompanied by a significant component of inflammation (e.g., rheumatoid arthritis, gout). However, acetaminophen is a safe and effective alternative for milder pain conditions including osteoarthritis. With respect to patient preference, a survey study demonstrated that only 14% of a large group of rheumatic disease patients (n = 1799) with rheumatoid arthritis, osteoarthritis, or fibromyalgia preferred acetaminophen over NSAIDs, whereas 60% preferred NSAIDs. In a head-to-head clinical trial of acetaminophen versus diclofenac plus misoprostol, there was significantly greater improvement in pain scores in patients in the diclofenac group. This finding was magnified in those patients with more severe disease at baseline.

Acetaminophen should be tried as the initial therapy in patients with mild to moderate pain for reasons of safety and cost. However, if patients have moderate to severe symptoms or if evidence of inflammation is present, moving to treatment with NSAIDs may provide more rapid and effective relief.

Future Directions

The strategy of blocking PG production by inhibiting the COX enzymes has provided relief from pain and inflammation for centuries. Given the proven importance of PG in this pathway and the advances in understanding the molecules involved, pharmacologic targeting of enzymes involved in biosynthesis, transport, or degradation may provide therapeutic efficacy. Similar to COX-2, mPGES-1 is induced during inflammation and in other pathologic states. Strategies to inhibit mPGES-1 have been proposed as potential alternatives to inhibiting COX. Drug development has been somewhat hampered because of the species specificity of mPGES-1 binding. However, this may be a particularly appealing strategy because preclinical data suggest that inhibition of mPGES-1 is associated with a reduced propensity to develop hypertension, thrombosis, atherosclerotic plaques, aortic aneurysm, and neointimal hyperplasia after vascular injury. Additionally, receptor antagonists could also be useful. Indeed, a number of EP receptor antagonists have proved useful in animal models of rheumatoid arthritis, osteoarthritis, and pain.  

Selected References


Full references for this chapter can be found on [www.expertconsult.com](http://www.expertconsult.com).
References


Glucocorticoids Therapy

JOHANNES W.G. JACOBS • JOHANNES W.J. BIJLSMA

Glucocorticoids are widely used for the treatment of patients with rheumatic disease. The first to be isolated, in 1935, was the naturally occurring glucocorticoid hormone, cortisol. It was synthesized in 1944 and subsequently became available for clinical use. In 1948, cortisol (then called compound E) was administered by the American physician Philip S. Hench to a 29-year-old woman with active rheumatoid arthritis (RA) of longer than 4 years’ duration. Her joints were so painful she could “hardly get out of bed.” After 2 days of treatment with 100 mg of intramuscular compound E daily, “She rolled over in bed with ease, and noted much less muscular soreness.” The next day, she was able to walk with “only a slight limp.” Hench published this case of dramatic improvement in 1949 and won the 1950 Nobel Prize in Physiology or Medicine for his research, which he shared with two colleagues at the Mayo Clinic. Later, by chemical modification of natural steroids, different synthetic glucocorticoids were produced, some of which have proved to be very effective anti-inflammatory and immunosuppressive substances with rapid, sometimes instant, effects.

Initially, there was considerable enthusiasm about glucocorticoid therapy because of the striking relief of symptoms seen in patients treated with supraphysiologic dosages. When the wide array of potentially serious adverse side effects became apparent, however, the use of glucocorticoids decreased. Nevertheless, because glucocorticoids can be considered the most effective anti-inflammatory and immunosuppressive substances currently known, they have become a cornerstone of therapy for many rheumatic disorders, including systemic lupus erythematosus (SLE), vasculitis, polymyalgia rheumatica, and myositis. The use of glucocorticoids in therapeutic strategies for patients with RA has become accepted.

During past decades, knowledge about glucocorticoids has increased, but much remains to be learned about the modes of actions of these drugs in rheumatic autoimmune disorders. It is hoped that the unraveling of these mechanisms eventually may lead to new applications of glucocorticoids or novel classes of therapy.

**CHARACTERISTICS OF GLUCOCORTICOIDS**

**Structure and Classification**

The precursor molecule of all steroid hormones is cholesterol, which is also a building block for vitamin D and cell membranes and organelles (Figure 60-1). Steroid hormones and cholesterol are characterized by a sterol skeleton, formed by three six-carbon hexane rings and one five-carbon pentane ring. The carbon atoms of this sterol nucleus are numbered in a specific sequence; the term steroid refers to this basic sterol nucleus (Figure 60-2).

Steroid hormones can be classified on the basis of their main function into sex hormones (male and female), mineralocorticoids, and glucocorticoids. Sex hormones are synthesized mainly in the gonads, but also in the adrenal cortex. Mineralocorticoids and glucocorticoids are synthesized only in the adrenal cortex; the terms corticosteroid and corticoid for these hormones refer to the adrenal cortex. Some glucocorticoids also have a mineralocorticoid effect and vice versa. The main natural mineralocorticoid is aldosterone, and the main natural glucocorticoid is cortisol (hydrocortisone). Although separation of corticoids into the classes mineralocorticoids and glucocorticoids is not absolute (see later), it is better (more precise) to use the term glucocorticoid than the term corticosteroid when referring to one of these compounds. The importance of standardized nomenclature is illustrated by the fact that an electronic literature search can be complicated by multiple synonyms.

In the 1950s, chemical modification of natural steroids revealed numerous structural features essential for specific biologic activities. Synthetic steroid hormones more potent than natural steroid hormones and steroid hormones with...
The introduction of a double bond between the 1 and 2 positions of cortisol yields prednisolone, which has about four times more glucocorticoid activity than cortisol (Table 60-1). Addition of a six-methyl group to prednisolone yields methylprednisolone, which is about five times more potent than cortisol. All the aforementioned glucocorticoids also have a mineralocorticoid effect. The synthetic glucocorticoids triamcinolone and dexamethasone have negligible mineralocorticoid activity, however.

Biologic Characteristics and Therapeutic Consequences

Apart from the steroid configuration, biologic characteristics of glucocorticoids also depend on whether they are in free form (as alcohol) or are chemically bound (as ester or salt). In their free form, glucocorticoids are virtually insoluble in water, so they can be used in tablets but not in parenteral preparations. For this reason, synthetic glucocorticoids are formulated as organic esters or as salts. Esters, such as (di)acetate and (hex)acetonide, are lipid soluble but have limited water solubility and are suitable for oral use and intramuscular, intralesional, and intra-articular injection. Salts, such as sodium phosphate and sodium succinate, are generally more water soluble and thus are also suitable for intravenous use. Dexamethasone sodium phosphate can be used intravenously, whereas dexamethasone acetate cannot. When given intramuscularly, dexamethasone sodium phosphate is absorbed much faster from the injection site than dexamethasone acetate. If an immediate effect is required, dexamethasone sodium phosphate given intravenously is more rapidly effective than the same preparation given intramuscularly; the least rapidly active is that of intramuscular dexamethasone acetate. For local use, less solubility means longer duration of the local effect, which generally is beneficial.
Pharmacokinetics and Pharmacology

Water insolubility does not impair absorption from the digestive tract. Most orally administered glucocorticoids, whether in free form or as an ester or salt, are absorbed readily, probably within about 30 minutes. Bioavailability of prednisone and prednisolone is high. Commercially available oral and rectal prednisone and prednisolone preparations are considered approximately bioequivalent.

The affinity of the different glucocorticoids for various plasma proteins varies (see Table 60-1). Of cortisol in plasma, 90% to 95% is bound to plasma proteins, primarily
transcortin (also called corticosteroid-binding globulin) and, to a lesser degree, albumin. Protein-bound cortisol is not biologically active, but the remaining 5% to 10% of free cortisol is. Prednisolone has—in contrast to methylprednisolone, dexamethasone, and triamcinolone—a high affinity for transcortin and competes with cortisol for this binding protein. The other synthetic glucocorticoids with little or no affinity for transcortin are two-thirds (weakly) bound to albumin, and about one-third circulate as free glucocorticoid.

Because only unbound glucocorticoids are pharmacologically active, patients with low levels of plasma protein, such as albumin (e.g., because of liver diseases or chronic active inflammatory diseases), are more susceptible to effects and side effects of glucocorticoids. Dosage adjustment should be considered in these patients. In liver disease, an additional argument for dosage adjustment is reduced clearance of glucocorticoids (see later).

Glucocorticoids have biologic half-lives 2 to 36 times longer than their plasma half-lives (see Table 60-1). With a plasma half-life of about 3 hours, prednisolone can be dosed once daily for most diseases. Maximal effects of glucocorticoids lag behind peak serum concentrations. Transcortin binds these compounds more strongly than does albumin. The plasma elimination of glucocorticoids bound to transcortin is slower than that of glucocorticoids that do not bind. Transcortin binding is not a major determinant of biologic half-lives of glucocorticoids, however, in contrast to distribution to different compartments of the body and binding to the cytosolic glucocorticoid receptor. Synthetic glucocorticoids have lower affinity for transcortin but higher affinity for the cytosolic glucocorticoid receptor than does cortisol (see later). The affinity of prednisolone and triamcinolone for the glucocorticoid receptor is approximately two times higher, and for dexamethasone it is seven times higher. Prednisone and cortisol have had negligible glucocorticoid bioactivity before they have been chemically reduced because of their very low affinity for the glucocorticoid receptor.

Another important factor determining biologic half-lives of glucocorticoids is the rate of metabolism. Synthetic glucocorticoids are subject to the same reduction, oxidation, hydroxylation, and conjugation reactions as cortisol. Pharmacologically active glucocorticoids are metabolized primarily in the liver into inactive metabolites and are excreted by the kidneys; only small amounts of unmetabolized drug are also excreted in the urine. An inverse correlation has been noted between prednisolone clearance and age, which means that a given dose may have a greater effect in older individuals.6 Prednisolone clearance also is slower in African-Americans compared with that in whites.7 The serum half-life of prednisolone is 2.5 to 5 hours, but it is increased in patients with renal disease and liver cirrhosis, and in the elderly. Prednisolone can be removed by hemodialysis, but overall, the amount removed does not require dosage adjustment in patients on hemodialysis. In patients with cirrhosis of the liver, clearance of unbound steroid is about two-thirds of normal—a difference that should be taken into account with dosing.

### Drug Interactions

Cytochrome P450 (CYP) is a family of isozymes responsible for the biotransformation of several drugs. Drug interactions can be based on induction or on inhibition of these enzymes. Certain drugs (e.g., barbiturates, phenytoin, rifampin) by inducing CYP isoenzymes (e.g., CYP3A4) increase the metabolism (breakdown) of synthetic and natural glucocorticoids, particularly by enhancing hepatic hydroxylase activity, thus reducing glucocorticoid concentrations (Figure 60-3). Rifampin-induced nonresponsiveness to prednisone in inflammatory diseases indeed has been described,8,9 as has rifampin-induced adrenal crisis in patients on glucocorticoid replacement therapy.10Clinicians should consider increasing the dosage of glucocorticoids in patients who are concomitantly treated with these medications.

Conversely, concomitant use of glucocorticoids with inhibitors of CYP3A4 (e.g., ketoconazole, itraconazole, diltiazem, mibebradil and grapefruit juice) decreases glucocorticoid clearance and leads to higher concentrations and prolonged biologic half-lives of glucocorticoid drugs, thus increasing the risk of adverse effects.11 Antifungal therapies, especially ketoconazole, on the other hand are known to interfere with endogenous glucocorticoid synthesis and therefore are also used, in doses of 400 to 800 mg per day, to treat hypercortisolism.11 Etomidate, a

### Table 60-1 Pharmacodynamics of Commonly Used Glucocorticoids

<table>
<thead>
<tr>
<th>Glucocorticoid</th>
<th>Equivalent Dose (mg)</th>
<th>Relative Glucocorticoid Activity</th>
<th>Relative Mineralocorticoid Activity</th>
<th>Protein Binding</th>
<th>Plasma Half-Life (hr)</th>
<th>Biologic Half-Life (hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-Acting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cortisone</td>
<td>25</td>
<td>0.8</td>
<td>0.8</td>
<td>−</td>
<td>0.5</td>
<td>8-12</td>
</tr>
<tr>
<td>Cortisol</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>+++</td>
<td>1.5-2</td>
<td>8-12</td>
</tr>
<tr>
<td>Intermediate-Acting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylprednisolone</td>
<td>4</td>
<td>5</td>
<td>0.5</td>
<td>−</td>
<td>&gt;3.5</td>
<td>18-36</td>
</tr>
<tr>
<td>Prednisolone</td>
<td>5</td>
<td>4</td>
<td>0.6</td>
<td>++</td>
<td>2.1-3.5</td>
<td>18-36</td>
</tr>
<tr>
<td>Prednisone</td>
<td>5</td>
<td>4</td>
<td>0.6</td>
<td>+++</td>
<td>3.4-3.8</td>
<td>18-36</td>
</tr>
<tr>
<td>Triamcinolone</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>++</td>
<td>2-5</td>
<td>18-36</td>
</tr>
<tr>
<td>Long-Acting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dexamethasone</td>
<td>0.75</td>
<td>20-30</td>
<td>0</td>
<td>++</td>
<td>3-4.5</td>
<td>36-54</td>
</tr>
<tr>
<td>Betamethasone</td>
<td>0.6</td>
<td>20-30</td>
<td>0</td>
<td>++</td>
<td>3-5</td>
<td>36-54</td>
</tr>
</tbody>
</table>

*Clinically; sodium and water retention, potassium depletion.
−, None; ++, high; ++++, high to very high; ++++, very high.
is poorly metabolized by 11β-HSD, and methasone has little or no affinity for transport proteins and is about 10:1, owing to these mechanisms. In contrast, dexamethasone, prednisone, prednisolone, and methylprednisolone would be good choices; if the unborn child has to be treated, fluorinated glucocorticoids, such as betamethasone or dexamethasone, would be indicated. Fear of physical (e.g., reduced growth) and neurocognitive adverse effects in children exposed to antenatal repeat doses of 12 mg betamethasone has not been substantiated, in contrast to postnatal glucocorticoid exposure. However, because of a small but increased risk of an oral cleft, it is advised to avoid high doses (1 to 2 mg/kg prednisone equivalent) in the first trimester of pregnancy, whereas low to moderate doses of prednisone seem to be safe.

Prednisolone and prednisone are excreted in small quantities in breast milk. Breastfeeding is generally considered safe for an infant whose mother is taking these drugs. Because curves of milk and serum concentrations for prednisolone are virtually parallel in time, exposure of the infant is minimized if breastfeeding is avoided during the first 4 hours after the intake of prednisolone.

**Figure 60-3** Serum prednisolone concentration in time in one patient, after 0.9 mg/kg prednisone orally daily, in the presence and absence of therapy with rifampin. Curve without rifampin, after a washout of rifampin of 4 weeks. Rifampin induces a reduced area under the curve of prednisolone, indicating reduced bioavailability.

short-acting intravenous anesthetic agent used for the induction of general anesthesia and for sedation, can also lower cortisol levels, which could be clinically relevant in critically ill patients.

Concomitant administration of prednisolone and cyclosporine may result in increased plasma concentrations of the former drug; concomitant administration of methylprednisolone and cyclosporine may result in increased plasma concentrations of the latter drug. The mechanism of this probably is competitive inhibition of microsomal liver enzymes. Antibiotics such as erythromycin may increase plasma concentrations of glucocorticoids. Synthetic estrogens in oral contraceptives increase the level of transcortin and thus total (sum of bound and unbound) glucocorticoid levels. Therefore, in women taking oral contraceptives, care is required in the interpretation of cortisol measurements, especially because adrenal insufficiency may be present even when total cortisol levels are within the normal range. Next to glucocorticoids, other steroid drugs such as megestrol acetate and medroxyprogesterone inhibit the hypothalamic-pituitary-adrenal axis; this risk may be increased when they are used concomitantly with glucocorticoids. Sulfasalazine has been reported to increase the sensitivity of immune cells for glucocorticoids, which could be beneficial.

**Pregnancy and Lactation**

In pregnancy, two mechanisms protect the fetus from exogenous glucocorticoids. First, glucocorticoids bound to transport proteins cannot pass the placenta, in contrast to unbound glucocorticoids. Second, the enzyme 11β-HSD in the placenta, which catalyzes the conversion of active cortisol, corticosterone, and prednisolone into the inactive 11-dehydro-prohormones (cortisone, 11-dehydrocorticosterone, and prednisone), protects the fetus from glucocorticoids in the blood of the mother. The maternal-to-fetal prednisolone blood concentration ratio is about 10:1, owing to these mechanisms. In contrast, dexamethasone has little or no affinity for transport proteins and is poorly metabolized by 11β-HSD in the placenta; the maternal-to-fetal dexamethasone blood concentration ratio is about 1:1.

If a pregnant woman has to be treated with glucocorticoids, prednisone, prednisolone, and methylprednisolone would be good choices; if the unborn child has to be treated, fluorinated glucocorticoids, such as betamethasone or dexamethasone, would be indicated. Fear of physical (e.g., reduced growth) and neurocognitive adverse effects in children exposed to antenatal repeat doses of 12 mg betamethasone has not been substantiated, in contrast to postnatal glucocorticoid exposure. However, because of a small but increased risk of an oral cleft, it is advised to avoid high doses (1 to 2 mg/kg prednisone equivalent) in the first trimester of pregnancy, whereas low to moderate doses of prednisone seem to be safe.

Prednisolone and prednisone are excreted in small quantities in breast milk. Breastfeeding is generally considered safe for an infant whose mother is taking these drugs. Because curves of milk and serum concentrations for prednisolone are virtually parallel in time, exposure of the infant is minimized if breastfeeding is avoided during the first 4 hours after the intake of prednisolone.

**BASIC MECHANISMS OF GLUCOCORTICOIDS**

**Genomic and Nongenomic Effects**

Glucocorticoids at any therapeutically relevant dosage exhibit pharmacologic effects via classic genomic mechanisms. The lipophilic glucocorticoid passes across the cell membrane, attaches to the cytosolic glucocorticoid receptor and heat shock protein, and binds to glucocorticoid-responsive elements on genomic DNA; it interacts with nuclear transcription factors. This process takes time. When acting through genomic mechanisms, it takes at least 30 minutes before the clinical effect of a glucocorticoid begins to show. Only when high doses are given, as in pulse therapy, can glucocorticoids act within minutes by nongenomic mechanisms; this occurs via specific receptor-mediated activity or via nonspecific membrane-associated physicochemical activity. The response to high-dose pulse methylprednisolone therapy may be biphasic, consisting of an early, rapid, nongenomic effect and a delayed and more sustained classic genomic effect. Clinically, genomic and nongenomic effects cannot be separated, however.

**Genomic Mechanisms**

Most of the effects of glucocorticoids are exerted via genomic mechanisms by binding to the glucocorticoid receptor located in the cytoplasm of the target cells; glucocorticoids are lipophilic and have a low molecular mass; thus they can pass through the cell membrane easily. Next to the tissue-specific intracellular density of glucocorticoid receptors, the balance of intracellular 11β-HSDs (see earlier) probably determines the sensitivity of specific tissues for glucocorticoids. Of the isoforms α and β of the glucocorticoid receptor, only the α isofrom, common in all target tissues, binds to glucocorticoids. This is a 94-kD protein to which several heat shock proteins (chaperones) are bound. Binding of the glucocorticoid to this complex causes shedding of heat shock proteins (chaperones) are bound. Binding of the glucocorticoid to this complex causes shedding of heat shock proteins (chaperones).
the chaperones. The resulting activated glucocorticoid receptor–glucocorticoid complex is rapidly translocated into the nucleus, where it binds (as a dimer) to specific consensus sites in the DNA (glucocorticoid-responsive elements), regulating the transcription of a large variety of target genes. This process is termed transactivation. Binding to glucocorticoid-responsive elements results in stimulation or suppression of transcription of these target genes. Suppression of genes also may be mediated by mechanisms involving interaction of the glucocorticoid receptor–glucocorticoid complex (as a monomer) with transcriptional factors, such as activator protein-1 and nuclear factor κB. This process is termed transrepression (Figure 60-4).

The nature and availability of these transcription factors may be pivotal in determining the differential sensitivity of different tissues to glucocorticoids because they play a crucial role in regulating the expression of a wide variety of proinflammatory genes induced by cytokines. The binding of transcriptional factors to DNA is inhibited by glucocorticoids, resulting in depressed expression of these genes and inhibition of their amplifying role in inflammation. Activated glucocorticoid receptors also may inhibit protein synthesis by decreasing the stability of mRNA through the induction of ribonucleases. This mechanism has been proposed to mediate glucocorticoid-induced inhibition of the synthesis of interleukin (IL)-1, IL-6, granulocyte-macrophage colony-stimulating factor, and inducible cyclooxygenase (COX)-2.

There is increasing acceptance of the hypothesis that side effects of glucocorticoids, such as diabetes mellitus, osteoporosis, skin atrophy, growth retardation, and cushingoid appearance, may be based predominantly on transactivation of genes after binding of glucocorticoid receptor–glucocorticoid to DNA, whereas the anti-inflammatory effects may be due mostly to the binding of a single glucocorticoid receptor–glucocorticoid complex to transcription factors or co-activators, resulting in gene repression (transrepression). Understanding of these molecular mechanisms may lead to development of novel glucocorticoids, such as selective glucocorticoid receptor agonists, with a more favorable balance of transactivation and transrepression and, clinically, to a more favorable balance of metabolic and endocrine side effects and therapeutic effects (see later).

Expression of multiple target genes at the post-transcriptional level, also those influenced by glucocorticoids, is modulated by microRNAs (miRNAs), short noncoding RNA molecules that are implicated in a wide array of cellular and immune processes. Abnormal expression of miRNAs has been found in patients with rheumatoid arthritis. This identifies miRNAs as targets for immunomodulatory drug development.

**Glucocorticoid Effects on the Immune System**

Glucocorticoids reduce activation, proliferation, differentiation, and survival of a variety of inflammatory cells, including macrophages and T lymphocytes, and promote apoptosis, especially in immature and activated T cells (Figure 60-5). This activity is mediated mainly by changes in cytokine production and secretion. In contrast, B lymphocytes and neutrophils are less sensitive to glucocorticoids, and their survival may be increased by glucocorticoid treatment. The main effect of glucocorticoids on neutrophils seems to be inhibition of adhesion to endothelial cells. Glucocorticoids inhibit not only the expression of adhesion molecules, but also the secretion of complement pathway proteins and prostaglandins. At supraphysiologic concentrations, glucocorticoids suppress fibroblast proliferation and IL-1 and tumor necrosis factor (TNF)-induced metalloproteinase synthesis. By these effects, glucocorticoids may retard bone and cartilage destruction in the inflamed joint.

![Figure 60-4: Genomic action of glucocorticoids. Glucocorticoid binds to the glucocorticoid receptor in the cytoplasm. This complex migrates into the nucleus. Activation of transcription (transactivation) by binding of glucocorticoid receptor–glucocorticoid dimers to glucocorticoid-responsive elements of DNA up-regulates synthesis of regulatory proteins, thought to be responsible for metabolic effects and also some anti-inflammatory/immunosuppressive effects. Interference of glucocorticoid receptor–glucocorticoid monomers with proinflammatory transcription factors, such as nuclear factor κB (NFκB), inhibits their binding to NFκB-responsive elements of DNA and transcription. This is called transrepression and down-regulates synthesis of predominantly inflammatory/immunosuppressive proteins.](image)
**Leukocytes and Fibroblasts**

Administration of glucocorticoids leads to an increase in the total leukocyte count caused by an increase in circulating neutrophil granulocytes in the blood, although the numbers of other leukocyte subsets in blood such as eosinophil and basophil granulocytes, monocytes/macrophages (decreased myelopoiesis and bone marrow release), and T cells (redistribution effect) are decreased. Table 60-2 summarizes the effects of glucocorticoids on leukocyte subsets. The redistribution of lymphocytes, which is maximal 4 to 6 hours after administration of a single high dose of prednisone and returns to normal within 24 hours, has no clinical consequences. B cell function and immunoglobulin production are hardly affected. The effects of glucocorticoids on monocytes and macrophages, including decreased expression of major histocompatibility complex (MHC) class II molecules and Fc receptors, may increase susceptibility to infection, however.

Effects of glucocorticoids on fibroblasts include decreased proliferation and decreased production of fibronectin and prostaglandins.

**Cytokines**

The influence of glucocorticoids on cytokine production and action represents one of the major mechanisms of glucocorticoid action in chronic inflammatory diseases such as RA. Glucocorticoids exert potent inhibitory effects on the transcription and action of a large variety of cytokines with pivotal importance in the pathogenesis of RA. Most T helper type 1 (Th1) proinflammatory cytokines are inhibited by glucocorticoids, including IL-1β, IL-2, IL-3, IL-6,

---

**Table 60-2  Anti-inflammatory Effects of Glucocorticoids on Immune Cells**

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrophils</td>
<td>Increased blood count, decreased trafficking, relatively unaltered functioning</td>
</tr>
<tr>
<td>Macrophages and monocytes</td>
<td>Decreased blood count, decreased trafficking, decreased phagocytosis and bactericidal effects, inhibited antigen presentation, decreased cytokine and eicosanoid release</td>
</tr>
<tr>
<td>Lymphocytes</td>
<td>Decreased blood count, decreased trafficking, decreased cytokine production, decreased proliferation and impaired activation, little effect on immunoglobulin synthesis</td>
</tr>
<tr>
<td>Eosinophils</td>
<td>Decreased blood count, increased apoptosis</td>
</tr>
<tr>
<td>Basophils</td>
<td>Decreased blood count, decreased release of mediators of inflammation</td>
</tr>
</tbody>
</table>
TNF, interferon-γ (indicative of Th1 helper cells), IL-17 (indicative of Th17 helper cells), and granulocyte-macrophage colony-stimulating factor (see Figure 60-5). In RA, these cytokines are considered responsible for synovitis, cartilage degradation, and bone erosion. Conversely, the production of Th2 cytokines, such as IL-4, IL-10, and IL-13, may be stimulated or not affected by glucocorticoids (see Figure 60-5). These cytokines have been related to the extra-articular features of erosive RA associated with B cell overactivity, such as immune complex formation and vasculitis. Activation of Th2 cells can suppress rheumatoid synovitis and joint destruction through release of the anti-inflammatory cytokines IL-4 and IL-10, which inhibit Th1 activity and downregulate monocyte and macrophage functions.27

Inflammatory Enzymes

An important part of the inflammatory cascade is arachidonic acid metabolism, which leads to the production of prostaglandins and leukotrienes, most of which are strongly proinflammatory. Through the induction of lipocortin (an inhibitor of phospholipase A2), glucocorticoids inhibit the formation of arachidonic acid metabolites. Glucocorticoids also have been shown to inhibit the production of COX-2 and phospholipase A2 induced by cytokines in monocytes/macrophages, fibroblasts, and endothelial cells. In addition, glucocorticoids are potent inhibitors of the production of metalloproteinases in vitro and in vivo, especially collagenase and stromelysin, which are the main effectors of cartilage degradation induced by IL-1 and TNF.28

Adhesion Molecules and Permeability Factors

Pharmacologic doses of glucocorticoids dramatically inhibit exudation of plasma and migration of leukocytes into inflammatory sites. Adhesion molecules play a central role in chronic inflammatory diseases by controlling the trafficking of inflammatory cells into sites of inflammation. Glucocorticoids reduce the expression of adhesion molecules through inhibition of proinflammatory cytokines and by direct inhibitory effects on the expression of adhesion molecules, such as intercellular adhesion molecule-1 and E-selectin.29 Chemotactic cytokines attracting immune cells to the inflammatory site, such as IL-8 and macrophage chemoattractant proteins, also are inhibited by glucocorticoids. Nitric oxide production in inflammatory sites is increased by proinflammatory cytokines, resulting in increased blood flow, exudation, and probably amplification of the inflammatory response. The inducible form of nitric oxide synthase by cytokines is potently inhibited by glucocorticoids.30

Hypothalamic-Pituitary-Adrenal Axis

Pathophysiology

Proinflammatory cytokines, such as IL-1 and IL-6, and eicosanoids, such as prostaglandin E2, and endotoxins all activate corticotropin-releasing hormone (CRH) at the hypothalamic level (Figure 60-6). This activation stimulates the secretion of adrenocorticotropic hormone (ACTH) by the pituitary gland and of glucocorticoids by the adrenal glands. In otherwise healthy individuals with severe infection or other major physical stress, cortisol production may increase to six times the normal amount.31 In patients with active RA (or other chronic inflammatory diseases), the increase in cortisol driven by elevated cytokines might be inappropriately low,32 meaning that cortisol levels—as although normal or elevated in the absolute sense—are insufficient to control the inflammatory response. This is the concept of relative adrenal insufficiency.31,33 Endogenous and exogenous glucocorticoids exert negative feedback control on the hypothalamic-pituitary-adrenal axis directly by suppressing secretion of ACTH and CRH, and indirectly by suppressing release from inflammatory tissues of proinflammatory cytokines, which stimulate secretion of ACTH and CRH (see Figure 60-6). Sensitivity of the hypothalamic-pituitary-adrenal axis for proinflammatory cytokines is probably decreased in RA.34

ACTH is secreted in brief, episodic bursts, resulting in sharp increases in plasma concentrations of ACTH and cortisol, followed by slower declines in cortisol levels—the normal diurnal rhythm in cortisol secretion. Secretory ACTH episodic bursts increase in amplitude but not in frequency after 3 to 5 hours of sleep, reach a maximum during the hours before and the hour after awakening, decline throughout the morning, and are minimal in the evening. Cortisol levels are highest at about the time of awakening in the morning, are low in the late afternoon and evening, and reach their lowest level some hours after falling asleep (see Figure 60-6). Glucocorticoids are not stored in the adrenal glands in significant quantities. Continuing synthesis and release are required to maintain basal secretion or to increase blood levels during stress. The total daily basal or physiologic secretion of cortisol in humans has been estimated to range from 5.7 to 10 mg/m²/day.35,36 This would be covered in primary adrenal insufficiency by oral administration of 15 to 25 mg cortisol,35 equivalent to about 4 to 6 mg prednisone. This low daily cortisol production rate may explain the cushingoid symptoms and other adverse effects that are sometimes observed in patients with adrenal insufficiency who are using glucocorticoids at doses previously regarded to be replacement doses (based on estimates of physiologic secretion of cortisol of 12 to 15 mg/m²/day), which are in fact supraphysiologic doses.

Effects of Glucocorticoids on the Hypothalamic-Pituitary-Adrenal Axis

Chronic suppression of the hypothalamic-pituitary-adrenal axis by administration of exogenous glucocorticoids leads by negative feedback loops on CRH and ACTH (see Figure 60-6) to failure in pituitary ACTH release, and thus to partial functional adrenal atrophy with loss of cortisol secretory capability in the fasciculata-reticularis zone. This inner cortical zone is the site of cortisol and adrenal androgen synthesis and is dependent on ACTH for structure and function. The outer cortical (glomerulosa) zone is involved in mineralocorticoid (aldosterone) biosynthesis and is functionally independent of ACTH. It stays functionally intact. Patients have failure of pituitary ACTH release and adrenal
Figure 60-6  Upper part, Stimulation (plus signs) and inhibition (minus signs) of the hypothalamic-hypopituitary-adrenal axis. Lower part, On the x-axis hours, plasma cortisol levels (blue line) in rheumatoid arthritis (RA) show an earlier and higher circadian rise compared with those in healthy controls, possibly caused by the rise in the proinflammatory cytokine interleukin-6 (IL-6); this rise is absent in healthy controls. IL-6 stimulates the hypothalamus and thus the release of cortisol, but probably also contributes to early morning stiffness and other inflammatory symptoms in (rheumatoid) arthritis. ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone.
responsiveness to ACTH. Serum cortisol, ACTH levels, and adrenal responsiveness to ACTH are low, but other pituitary axes function normally, in contrast to the situation in most primary pituitary disorders. The time required to achieve suppression depends on the dosage and the serum half-life of the glucocorticoid used, but it also varies among patients, probably because of individual differences in glucocorticoid sensitivity and rates of glucocorticoid metabolism. Prediction with certainty of chronic suppression of the hypothalamic-pituitary-adrenal axis and adrenal insufficiency is impossible. This risk may be increased when glucocorticoids are used concomitantly with other steroid drugs such as mesterol acetate and medroxyprogesterone, inhibiting the hypothalamic-pituitary-adrenal axis.11

The duration of the anti-inflammatory effect of one dose of a glucocorticoid approximates the duration of hypothalamic-pituitary-adrenal suppression. After a single oral dose of 250 mg of hydrocortisone or cortisol, 50 mg of prednisone or prednisolone, or 40 mg of methylprednisolone, suppression for 1.25 to 1.5 days has been described. Duration of suppression after 40 mg of triamcinolone and 5 mg of dexamethasone was 2.25 and 2.75 days.32 After intramuscular administration of a single dose of 40 to 80 mg of triamcinolone acetonide, the duration of hypothalamic-pituitary-adrenal suppression is 2 to 4 weeks, and after 40 to 80 mg of methylprednisolone, suppression lasts 4 to 8 days.32

In the case of long-term therapy, for patients who have had less than 10 mg of prednisone or its equivalent per day in one dose in the morning, the risk of clinical (symptomatic) adrenal insufficiency is not high, but neither is it negligible. A review of adrenal insufficiency stated that if the daily dose is 7.5 mg of prednisolone or equivalent or more for at least 3 weeks, adrenal hypofunction should be anticipated, and acute cessation of glucocorticoid in this situation could lead to problems.12 Patients who have received glucocorticoids for less than 3 weeks or have been treated with alternate-day prednisolone therapy do not have zero risk of suppression of the hypothalamic-pituitary-adrenal axis, depending on the dose,13 but the risk is low. After 5 to 30 days of at least 25 mg of prednisone or equivalent daily, suppression of adrenal response (measured by a low-dose corticotropin test) was present in 34 of 75 patients studied (45%).40 In these patients, a basal plasma cortisol concentration less than 100 nmol/L was highly suggestive of adrenal suppression, whereas levels of basal cortisol greater than 220 nmol/L predicted a normal adrenal response in most, but not all, patients. When in doubt, it seems prudent to treat patients as having secondary adrenal insufficiency. Secondary adrenal insufficiency generally has a less dramatic presentation than primary adrenal insufficiency because aldosterone levels, which are controlled predominantly by the renin-angiotensin system, are preserved; mineralocorticoid therapy is not necessary.

### TREATMENT WITH GLUCOCORTICOIDS

Glucocorticoids are widely used in various dosages for several rheumatic diseases. Often it is unclear what is meant by the semi-quantitative terms used for dosages, such as low or high. Based on pathophysiologic and pharmacokinetic data, standardization has been proposed to minimize problems in interpretation of these generally used terms (Table 60-3).2

<table>
<thead>
<tr>
<th>Terminology of Dosages of Glucocorticoids for Use in Rheumatology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low dose</td>
</tr>
<tr>
<td>Medium dose</td>
</tr>
<tr>
<td>High dose</td>
</tr>
<tr>
<td>Very high dose</td>
</tr>
<tr>
<td>Pulse therapy</td>
</tr>
</tbody>
</table>

### Indications

For each disease, indications for glucocorticoid therapy are discussed in the specific chapters. An overview is given here (Table 60-4), which summarizes only the general uses and dosages of glucocorticoids. Without detailed description, some of the indications could be considered questionable at first glance. In systemic sclerosis, glucocorticoids, especially in high doses, are contraindicated because of the risk of scleroderma renal crisis, but they may be useful for myositis or interstitial lung disease. Glucocorticoids are a basic part of the therapeutic strategy in myositis, polymyalgia rheumatica, and systemic vasculitis. For other diseases, glucocorticoids serve as adjunctive therapy or are not used at all. For instance in RA, glucocorticoids are almost exclusively used as adjunctive therapy in combination with other disease-modifying antirheumatic drugs (DMARDs) (see later). In osteoarthritis, glucocorticoids are not given except for intra-articular injection if signs of synovitis of the osteoarthritic joint are present.41 For generalized soft tissue disorders, glucocorticoids are not indicated, and for localized soft tissue disorders, they should be used only for intralesional injection.42

### Glucocorticoid Therapy in Rheumatoid Arthritis

Glucocorticoids are a frequently applied medication in RA. In the past, more patients with RA seemed to be given concomitant glucocorticoids in the United States than in Europe—54% versus 27%43,44—whereas more recent data suggest that 38% of RA patients in the United States use glucocorticoids45 versus up to 55% of German RA patients.46 Aims of this therapy include reduction of signs and symptoms and inhibition of joint damage.

### Signs and Symptoms

As can be seen in Table 60-4, RA is the only disease in which glucocorticoid therapy is often started and maintained at a low dose as additional therapy. The rationale for this therapy is a probable, relative insufficiency of the adrenal gland in patients with active RA.31 Glucocorticoids are highly effective for relieving symptoms in patients with active RA in doses of less than 10 mg/day. Many
Table 60-4 General Use of Glucocorticoids in Rheumatology

<table>
<thead>
<tr>
<th>Arthritis</th>
<th>Initial Oral Dose*</th>
<th>Intravenous, Very High Dose† or Pulse</th>
<th>Intra-articular Injection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low†</td>
<td>Medium†</td>
<td>High†</td>
</tr>
<tr>
<td>Gout</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Juvenile idiopathic arthritis</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Pseudogout</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
<td>–</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Reactive arthritis</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>–</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Collagen Disorders</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dermatomyositis, polymyositis</td>
<td>–</td>
<td>–</td>
<td>3</td>
</tr>
<tr>
<td>Mixed connective tissue disease</td>
<td>–</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Polymyalgia rheumatica</td>
<td>–</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>Sjögren’s syndrome, primary</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>–</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Systemic sclerosis</td>
<td>–</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Systemic Vasculitis in General</td>
<td>–</td>
<td>–</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Initial dose is the dose at the start of therapy and often is decreased in time depending on disease activity.
†Dose in prednisone equivalents per day: low, ≤7.5 mg; medium, >7.5 but ≤30 mg; high, >30 but ≤100 mg; very high, >100 mg.
–, Rare use.
1. Infrequent use or use for therapy-resistant disease, complications, severe flare, and major exacerbation.
2. Frequently added to the basic therapeutic strategy.
3. Basic part of therapeutic strategy.

Patients become functionally dependent on this therapy, however, and continue it over the long term. A review of seven studies (253 patients) concluded that glucocorticoids, when administered for approximately 6 months, are effective for the treatment of RA. After 6 months of therapy, the beneficial effects of glucocorticoids seem to diminish. If this therapy then is tapered off and stopped, however, patients often—over some months—experience aggravation of symptoms.

**Radiologic Joint Damage: Glucocorticoids as DMARDs**

In 1995, joint-preserving effects of 7.5 mg of prednisolone daily for 2 years were described in patients with RA of short and intermediate duration who also were treated with DMARDs. The group of RA patients participating in this randomized, placebo-controlled trial was heterogeneous, not only with respect to disease duration, but also with respect to stages of the disease and types and dosages of DMARDs. In another trial published in 1997, patients with early RA were randomly assigned to step-down therapy with two DMARDs (sulfasalazine and methotrexate) and prednisolone (start 60 mg/day, tapered in six weekly steps to 7.5 mg/day and stopped at 34 weeks) or to sulfasalazine alone. In the combined drug strategy group, a statistically significant and clinically relevant effect in retarding joint damage was shown compared with the effect of sulfasalazine alone. In an extension of this study, long-term (4 to 5 years) beneficial benefits were shown regarding radiologic damage after the combination strategy. It has been hypothesized that the superior effect of the combination therapy in this trial can be ascribed to prednisolone because in three double-blind, randomized trials, the effect of the combination of methotrexate and sulfasalazine was not superior to that of either drug alone.

In a German study, 200 patients with early RA were treated with methotrexate or intramuscular gold and were randomly assigned to additional treatment with 5 mg of prednisolone or placebo. After 2 years, progression of radiologic damage proved to be less in the prednisolone-treated patients than in those treated with placebo. In 2002, results of the Utrecht study on the effects of prednisolone in DMARD-naïve patients with early RA were published. This is the only placebo-controlled trial in which prednisolone was applied as monotherapy as the first step. The progression of radiologic joint damage was inhibited by 10 mg of prednisolone daily in these patients (who received DMARD therapy only as rescue).

The Utrecht study reported a 40% decreased need for intra-articular glucocorticoid injections, a 49% decreased need for acetaminophen use, and a 55% decreased need for nonsteroidal anti-inflammatory drugs (NSAIDs) in the prednisolone group compared with the placebo group. This indicates that in clinical trials evaluating the clinical effects of DMARDs or glucocorticoids, additional therapies should be taken into account. In an extension of this study, at 3 years after the end of the study and 2 years after tapering off and stopping the prednisolone therapy, beneficial radiologic benefits of prednisolone were still present (Figure 60-7). In another 2-year study in 250 patients with early RA, 7.5 mg/day of prednisolone added to DMARD therapy retarded joint damage and increased the remission rate compared with placebo added to DMARDs. Even in an intensive treat-to-target methotrexate-based strategy in early RA, prednisone enhanced clinical efficiency and reduced erosive joint damage.
Negative studies on the effects of glucocorticoids on radiologic damage have also been published, but in early RA, evidence of glucocorticoid joint-sparing effects, which persist after therapy is stopped, seems convincing, thus classifying glucocorticoids as DMARDs. A meta-analysis on radiographic outcome analyzed 15 studies (two with negative results) that included a total of 1414 patients. Because different methods had been used in the individual trials, radiographic scores were expressed as a percentage of the maximum possible score for the specific radiographic method used. The standardized mean difference in progression was 0.40 in favor of strategies using glucocorticoids (95% confidence interval, 0.27 to 0.54). This was considered a conservative estimate because the most conservative estimate of the difference in each study had been chosen.

It is still unknown, however, whether glucocorticoids can also inhibit progression of erosion in RA of longer duration than 2 years. A so-called window of opportunity may exist in the treatment of RA. If this window is present, effective treatment of early RA with glucocorticoids and DMARDs may result in an effect that lasts for a long time and in disease that is easier to control, whereas if effective treatment starts later, this opportunity may be lost, resulting in more difficult control of the disease with inflammation fueled by joint damage. Most studies on glucocorticoids and radiologic damage employed a dose of 5 to 10 mg/day of prednisone equivalent during 2 years, but a scheme starting with 60 mg/day tapered off and stopped within 34 weeks also was effective. In addition, because glucocorticoid-induced osteoporosis and peptic ulcer complications (if glucocorticoids are combined with nonsteroidal anti-inflammatory drugs [NSAIDs]) can be prevented much more effectively now than some decades ago, the joint-protective effect of prednisolone in RA during the first 2 years of the disease in a dose of 5 to 10 mg daily is a relevant finding.

The joint-sparing effect of glucocorticoids probably is based on inhibition of proinflammatory cytokines such as IL-1 and TNF, which stimulate osteoblasts and T cells to produce receptor activator of nuclear factor κB (RANK) ligand. This binds to RANK on osteoclast precursor cells and on mature osteoclasts, leading to activation of osteoclasts, which are responsible for bone resorption, periarticular osteopenia, and formation of bone erosions in RA.

A toxicity index score for DMARDs was published (based on symptoms, laboratory abnormalities, and hospitalization data) after evaluation of 3000 patients with more than 7300 patient-years from the Arthritis, Rheumatism, and Aging Medical Information System (ARAMIS) database. Although this score has not been validated and is influenced by confounding-by-indication, it gives an impression of the relative toxicity of glucocorticoids. It is comparable with that of other immunosuppressive medications used in RA, such as methotrexate and azathioprine. A review also showed that the incidence, severity, and impact of adverse effects of low-dose glucocorticoid therapy in RA trials were modest and suggested that probably many of the well-known adverse effects of glucocorticoids are predominantly associated with high-dose treatment.

Because many questions remain to be answered, such as how the effects of glucocorticoids compare with those of high dosages of methotrexate or of TNF blockers, and for how long glucocorticoids should be prescribed and in what dosages, the final place of glucocorticoid therapy in RA has to be clearly determined. Nevertheless in early RA the use of glucocorticoids generally has been accepted. Guidelines on how to use (low-dose) glucocorticoids and how to monitor this therapy have been developed.

Prevention of Early (Rheumatoid) Arthritis Development with Glucocorticoids

Recently, trials have been done to try to prevent arthralgia or early arthritis from progressing to chronic arthritis. In patients with (very) early arthritis or individuals with arthralgia and antibodies to citrullinated proteins or rheumatoid factor, intramuscular glucocorticoid injections did not prevent arthritis development in two placebo-controlled trials, but in another placebo-controlled double-blind trial these injections postponed the need for DMARDs and prevented 1 in 10 patients from progressing into RA at assessment at 12 months. These results are preliminary and nonconclusive, and it is clear that further research is needed.

Chronobiology

The rheumatoid inflammatory process and symptoms have a diurnal rhythm. Early in the morning, patients experience the most extensive joint stiffness and other symptoms and signs; this is due to the long rest period during the night that facilitates edema formation around inflamed joints and the circadian rhythm of cortisol (see Figure 60-6). In patients with RA with low or medium disease activity, serum cortisol maximum and minimum shift to earlier times of the day and night, whereas in patients with high
disease activity, the circadian rhythm is markedly reduced or even lost.

The timing of glucocorticoid administration may be important for efficacy and side effects. Older data in the literature on this topic are ambiguous. Recently, a trial was performed with a newly developed modified-release prednisone tablet that releases prednisone about 4 hours after ingestion. When it was taken in the evening, thus adapting its release to circadian increases in proinflammatory cytokine concentrations, symptoms of RA early in the morning were lessened compared with those reported when the same dose of prednisone was taken early in the morning. This 3-month double-blind trial included RA patients with a duration of morning stiffness of 45 minutes or longer, a pain score of 30 mm or less on a 100-mm visual analog scale, three or more painful joints, one or more swollen joints, and an erythrocyte sedimentation rate (ESR) of 28 mm or greater or a C-reactive protein concentration 1.5 times or more the upper limit of normal, who were on glucocorticoids at least 3 months with a stable daily dose of 2 to 10 mg prednisone equivalent for at least 1 month. Patients were randomized to continue their prednisone or to switch to modified-release prednisone in a double-dummy way. At the end of the trial, the difference in duration of morning stiffness was about 30 minutes, in favor of the modified-release prednisone group. However, no differences were noted in all other variables of disease activity between the two groups. The safety profile did not differ between treatments. Long-term benefits and risks of this preparation and application in other inflammatory rheumatic diseases have yet to be investigated.

Other Developments to Improve the Therapeutic Ratio of Glucocorticoids

In addition to guidelines put forth to improve the clinical use of existing glucocorticoids, other formulations have been and are being developed. Deflazacort, an oxazoline derivative of prednisolone introduced in 1969, was initially thought to be as effective as prednisone while inducing fewer adverse events, but there was the issue of the real equivalence ratio compared with prednisone; this drug has not represented a major breakthrough. Knowledge about the mechanisms of glucocorticoids (transrepression and transactivation leading, respectively, to predominantly beneficial effects and adverse effects; see earlier) led to the development of selective glucocorticoid receptor agonists or dissociating glucocorticoids, but as yet they have not entered the market. Glucocorticoid preparations releasing nitric oxide, the so-called nitrosteroids, could induce stronger anti-inflammatory effects because nitric oxide has anti-inflammatory effects too. These drugs have to be tested in patients yet. The drug combination prednisolone and diprydiamole has been reported to boost and extend the net glucocorticoid effect in laboratory models. The next required step will be to demonstrate the improved therapeutic ratio in patients in adequate comparative clinical trials by assessing predefined beneficial effects and adverse effects in a standardized way. Liposomes containing glucocorticoids and targeted to integrins expressed on endothelial cells at sites of inflammation have been studied; these deliver their glucocorticoids specifically at sites of inflammation. Their selective biodistribution might allow for less frequent and lower dosing, which could result in an improved therapeutic ratio. The safety of liposomal prednisolone has been evaluated in a small group of RA patients, and the results (up until now published only as an abstract) seem promising. All of these new applications have to be tested further before they can be used in daily clinical practice.

Alternate-Day Regimens

For oral, long-term use of glucocorticoid therapy, alternate-day regimens have been devised in an attempt to alleviate the undesirable side effects, such as hypothalamic-pituitary-adrenal axis suppression. Alternate-day therapy uses a single dose administered every other morning, which is usually equivalent to, or higher than, twice the usual or pre-established daily dose. The rationale for this regimen is that the body, including the hypothalamic-pituitary-adrenal axis, is exposed to exogenous glucocorticoid only on alternate days. This rationale makes sense only for usage of a class and dosage of a glucocorticoid that suppresses the hypothalamic-pituitary-adrenal axis activity for less than 36 hours after a single dose. Another prerequisite is that the patient should have a responsive hypothalamic-pituitary-adrenal axis that is not chronically suppressed by previous glucocorticoid regimens. The alternate-day schedule does not work in patients on long-term medium- or high-dose glucocorticoids suppressing hypothalamic-pituitary-adrenal axis activity for longer than 36 hours.

Alternate-day therapy is unsuccessful in most patients who require glucocorticoids. Patients with RA often experience exacerbation of symptoms on the second day. This experience is in line with the clinical impression that a single dose of glucocorticoids daily is less effective in RA than half that dose, given twice daily. In giant cell arteritis, alternate-day glucocorticoid therapy also is less effective than daily administration. Generally, alternate-day regimens are used rarely in rheumatology today, except in patients with juvenile idiopathic arthritis, in whom alternate-day glucocorticoid usage results in less inhibition of body growth than is associated with daily usage. If treatment has been initiated with daily administration, the change to alternate-day therapy preferably should be made after the disease has stabilized.

Glucocorticoid Sensitivity and Resistance

A small proportion of patients does not react favorably to glucocorticoids or even fails to respond to high doses. Also, susceptibility to adverse effects of glucocorticoids varies widely. Several different factors are involved in the variability of glucocorticoid sensitivity in patients with rheumatic diseases, and an understanding of the mechanisms involved might eventually allow their modulation. Potential mechanisms of glucocorticoid resistance in inflammatory diseases have been reviewed extensively.

Hereditary glucocorticoid resistance (rare) and increased susceptibility to glucocorticoids have been related to specific polymorphisms of the glucocorticoid receptor gene. The glucocorticoid receptor exists as α and β isoforms, but only the α isoform binds glucocorticoids. The β isoform
functions as an endogenous inhibitor of glucocorticoids and is expressed in several tissues. Glucocorticoid resistance has been associated with enhanced expression of this β receptor, but this is unlikely to be an important mechanism for glucocorticoid resistance because in most cells, apart from neutrophilic granulocytes, expression of the β receptor is much less than that of the α receptor.\(^5\) The protein lipocortin-1 (or annexin-1) inhibits eicosanoid synthesis. Glucocorticoids are thought to stimulate lipocortin-1. In patients with RA, autoantibodies to lipocortin-1 have been described. The titers in these patients correlate with the height of maintenance doses of glucocorticoids, suggesting that these antibodies may lead to glucocorticoid resistance.

Although glucocorticoids exert most of their immunosuppressive actions through inhibition of cytokine production, high concentrations of cytokines, especially IL-2, antagonize the suppressive effects of glucocorticoids in a dose-dependent manner.\(^7\) The balance is usually in favor of glucocorticoids, but high local concentrations of cytokines may result in localized glucocorticoid resistance that cannot be overridden by exogenous glucocorticoids. Also, the macrophage migration inhibitory factor may play a role in steroid resistance in RA. This proinflammatory cytokine is involved in TNF synthesis and T cell activation, suggesting a role in the pathogenesis of RA. Macrophage migration inhibitory factor is suppressed by higher concentrations of glucocorticoids, but it is induced by low concentrations, leading to stimulation of inflammation.\(^7\) Other possible mechanisms of glucocorticoid resistance include activation of mitogen-activated protein kinase pathways by certain cytokines, excessive activation of the transcription factor activator protein-1, reduced histone deacetylase-2 expression, and increased P-glycoprotein–mediated drug efflux.\(^8\)

Also, drugs may play a role in glucocorticoid sensitivity and resistance (see also the section on drug interactions). Sulfasalazine increases the sensitivity of immune cells for glucocorticoids and thus might be a future option for preventing or treating glucocorticoid resistance.\(^1\) Mifepristone is an antiprogestrone drug and glucocorticoid receptor antagonist; chlorpromazine inhibits glucocorticoid receptor–mediated gene transcription.\(^9\)

### Glucocorticoid Withdrawal Regimens

Because of potential side effects, glucocorticoids usually are tapered off as soon as the disease being treated is under control. Tapering must be done carefully to avoid recurrent activity of the disease and, infrequently, cortisol deficiency resulting from chronic hypothalamic-pituitary-adrenal axis suppression. Gradual tapering permits recovery of adrenal function. There is no best scheme based on controlled, comparative studies for tapering glucocorticoids. Tapering depends on the individual disease, the disease activity, doses and duration of therapy, and clinical response, which also depends on each individual's glucocorticoid sensitivity. Only generic guidelines can be offered. To taper the dose of prednisone, decrements of 5 to 10 mg every 1 to 2 weeks can be used when the prednisone dose is more than 40 mg/day, followed by 5-mg decrements every 1 to 2 weeks at a dose between 40 and 20 mg/day, and finally 1 to 2.5 mg/day decrements every 2 to 3 weeks at a prednisone dose of less than 20 mg/day. Another scheme is to taper 5 to 10 mg every 1 to 2 weeks down to 30 mg/day of prednisone, and when the dose is less than 20 mg/day, taper 2.5 to 5 mg every 2 to 4 weeks down to 10 mg/day; thereafter, the dose is tapered 1 mg each month or 2.5 mg (half a 5-mg tablet of prednisolone) each 7 weeks. For tapering every 7 weeks or over longer periods, a printed schedule can be given to the patient, such as the one shown in Table 60-5.

### Adaptations of Glucocorticoid Doses, Stress Regimens, and Perioperative Care

Patients on long-term low-dose glucocorticoid medication have suppressed adrenal activity and should be advised to double their daily glucocorticoid dose or to increase the dose to 15 mg prednisolone or equivalent if they develop fever attributed to infection, and to seek medical help. In case of major surgery, given the unreliable prediction of adrenal suppression on the basis of duration and dose of glucocorticoid therapy (see the section on effects of glucocorticoids on the hypothalamic-pituitary-adrenal axis), many physicians recommend “stress doses” of glucocorticoids for patients with low risk of adrenal suppression. The scheme of 100 mg of hydrocortisone intravenously just before surgery, followed by an additional 100 mg every 6 hours for 3 days, is based on anecdotal information and is not always necessary.\(^9,10\) A scheme with a lower dose, possibly reducing the risk of postoperative bacterial infectious complications, is to infuse continuously 100 mg of hydrocortisone intravenously on the day of surgery, followed by 25 to 50 mg of hydrocortisone every 8 hours for 2 or 3 days. Another option is to administer the usual dose of oral glucocorticoid orally (or the equivalent) parenterally on the day of surgery, followed by 25 to 50 mg of hydrocortisone every 8 hours for 2 or 3 days.

In cases of minor surgery, it is probably sufficient to double the oral dose or to increase the dose to 15 mg of

---

**Table 60-5  Glucocorticoid Tapering Scheme to Hand Out to Patients***

<table>
<thead>
<tr>
<th>Period 1</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
<th>Sunday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period 2</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Period 3</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Period 4</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Period 5</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Period 6</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Period 7</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

*At each consecutive period (e.g., 1 week or some weeks), the number of days during which a low dose should be taken increases by 1. After completion of period 7, the next step in tapering can be taken; the dose called “low” during the previous 7 periods now is “high,” and so on. In case of aggravation of symptoms, the patient should not diminish the dose and should contact the specialist.

---


prednisolone or equivalent for 1 to 3 days. No comparative randomized studies on different perioperative glucocorticoid stress schemes have been published, however. Because in glucocorticoid-induced secondary adrenal insufficiency, aldosterone secretion is preserved, mineralocorticoid therapy is unnecessary, in contrast to in primary adrenal insufficiency.

**Glucocorticoid-Sparing Agents**

For most inflammatory rheumatic diseases, including SLE, vasculitis, RA, and myositis, other immunomodulatory drugs are often added to therapy with glucocorticoids, such as azathioprine and methotrexate, and especially in case of systemic vasculitis, cyclophosphamide. For these indications, biologic agents are increasingly used. An exception is polymyalgia rheumatica, which is managed primarily with glucocorticoids alone. Combination therapy is applied early in the disease when the disease is one for which it is known that the effect of the combination is better than that of glucocorticoids alone (e.g., in the case of systemic vasculitis), or if the disease (e.g., inflammatory myositis) seems resistant to high initial doses of glucocorticoids.

If at a later stage of the disease, immunomodulatory drugs are added to therapy with glucocorticoids to enable further reduction of the dose to decrease the risk of side effects, these immunomodulatory drugs are termed glucocorticoid-sparing agents. For this purpose, azathioprine and methotrexate are often used, although any drug that has an additive or synergistic effect in suppressing the disease, enabling reduction of the glucocorticoid dose, could be used as a glucocorticoid-sparing agent.

**Glucocorticoid Pulse Therapy**

Glucocorticoid pulse therapy is used in rheumatology, especially for remission induction or treatment of flares of inflammatory rheumatic disorders and vasculitides (see Table 60-4). In RA, pulse therapy is applied to treat serious complications of the disease and to induce remission in active disease, often during the initiation phase of a (new) DMARD strategy. In the latter patients, pulse therapy with schemes of 1000 mg of methylprednisolone given intravenously has been proven effective in many studies. The beneficial effect generally lasts about 6 weeks, with large variation in the duration of the effect. If it does not seem sensible to apply pulse therapy in active RA, unless a change in the therapeutic strategy (i.e., in second-line antirheumatic treatment) aims to stabilize over the long term any remission induced by the pulse therapy. Short-term effects of pulse therapy in patients with established, active disease, often during the initiation phase of a (new) DMARD therapy, such as methotrexate, in patients with early RA.

In 144 patients with biopsy-confirmed giant cell arteritis, of whom 91 were seen initially with visual loss and 53 without visual loss, no evidence was found that intravenous glucocorticoid pulse therapy (usually 150 mg dexamethasone sodium phosphate every 8 hours for 1 to 3 days) was more effective than high daily doses (80 to 120 mg) of oral prednisone in preventing visual deterioration.

The risk of adverse effects of pulse therapy is not the same for all rheumatic disorders. In patients with SLE, osteonecrosis and psychosis seem to be more frequent side effects of pulse therapy compared with those seen in patients with RA. Osteonecrosis and psychosis also can be complications of SLE itself, however. Contraindications for pulse therapy include pregnancy and lactation, infection, current peptic ulcer disease, glaucoma, badly controlled hypertension, and diabetes mellitus. In cases with a family history of glaucoma or well-controlled hypertension or diabetes mellitus, pulse therapy can be applied with checks, respectively, of eye and blood pressure and of blood glucose values.

**Intraleisonal and Intra-articular Glucocorticoid Injections**

Injections with glucocorticoids are widely used for arthritis (see Table 60-4), tenosynovitis, bursitis, enthesitis, and compression neuropathies such as carpal tunnel syndrome. Generally, the effect occurs within days; it can be long-lasting, but if the underlying disease is active, the effect is of short duration. Administration of a local anesthetic concurrently with intra-articular or soft tissue injection of a glucocorticoid may provide immediate pain relief.

Soluble glucocorticoids (e.g., phosphate salts) have a more rapid onset of action with probably less risk of subcutaneous tissue atrophy and depigmentation of the skin when given intraleisonally. Insoluble glucocorticoids are longer acting and might further decrease the soft tissue fibrous matrix, so they should be used with caution in places with thin skin, especially in elderly patients and in those with peripheral vascular disease. Insoluble glucocorticoids are more safely given into deep sites. Short-acting soluble glucocorticoids can be mixed with long-acting insoluble glucocorticoids to combine rapid onset with long-acting effect.

The effect of intra-articular glucocorticoid injection probably depends on several factors: the underlying disease (e.g., RA, osteoarthritis), the treated joint (size, weight bearing, or non–weight bearing), the activity of arthritis, the volume of synovial fluid in the joint to be treated, the application of arthrocentesis (synovial fluid aspiration) before injection, the choice and dose of the glucocorticoid preparation, application of rest to the injected joint, and the injection technique used. The effects of injections seem to be less favorable in osteoarthritis than in RA. Arthrocentesis before injection of the glucocorticoid preparation reduces the risk for relapse of arthritis. Triamcinolone hexacetonide, which, among the injectable glucocorticoids, is the least soluble preparation, shows the longest effect.

Theoretically, rest of the injected joint minimizes leakage of the injected glucocorticoid preparation into the systemic circulation (via the hyperemic, inflamed synovium by enhanced pressure in the joint during activity), minimizes the risk of cartilage damage, and enhances repair of inflammatory tissue damage. Advice and procedures for the postinjection period in terms of activity vary from no restrictions, to minimal activity of the injected joint for a couple of days, to bed rest for 24 hours after injection of a knee joint or splinting of injected joints. Based on the literature, no definite evidence-based recommendations can be made, but it
seems prudent to rest and to not overuse the injected joint for several days, even if pain is relieved.

It is recommended that intra-articular glucocorticoid injections be repeated no more often than once every 3 weeks, and that they be given no more frequently than three times a year in a weight-bearing joint (e.g., the knee) to minimize glucocorticoid-induced joint damage. This recommendation seems sensible, but no definitive clinical evidence is available to support it. As one would expect, accuracy of steroid placement influences the clinical outcome of glucocorticoid injections into the shoulder and probably into other joints as well. This is important because it is estimated that a few more than half of shoulder injections are inaccurately placed. Introduction of disposable needles and syringes has helped reduce the risk. In a 3-year prospective study in an urban area of 1 million people in the Netherlands, bacterial infections were detected in 214 joints (including 58 joints with a prosthesis or osteosynthetic material) of 186 patients; only 3 of these joint infections were attributed to an intra-articular injection.

Other adverse effects of local glucocorticoid injections include systemic adverse effects of the glucocorticoid, such as disturbance in the menstrual pattern, hot flush–like symptoms the day of or the day after injection, and hyperglycemia in diabetes mellitus. Local complications include subcutaneous fat tissue atrophy (especially after improper local injection), local depigmentation of the skin, tendon slip and rupture, and lesions to local nerves. This is important because it is estimated that a few more than half of shoulder injections are inaccurately placed. The reported infection rate of joints after local injection with glucocorticoids is low, ranging from 1 case in 13,900 to 1 in 77,300 injections. Introduction of disposable needles and syringes has helped reduce the risk. In a 3-year prospective study in an urban area of 1 million people in the Netherlands, bacterial infections were detected in 214 joints (including 58 joints with a prosthesis or osteosynthetic material) of 186 patients; only 3 of these joint infections were attributed to an intra-articular injection.

Table 60-6  Adverse Effects of Glucocorticoids

<table>
<thead>
<tr>
<th>System</th>
<th>Adverse Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skeletal</td>
<td>Osteoporosis, osteonecrosis, myopathy</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>Peptic ulcer disease (in combination with nonsteroidal anti-inflammatory drugs), fatty liver</td>
</tr>
<tr>
<td>Immunologic</td>
<td>Predisposition to infection, suppressed delayed hypersensitivity (Mantoux test)</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>Fluid retention, hypertension, accelerated arteriosclerosis, arrhythmias</td>
</tr>
<tr>
<td>Ocular</td>
<td>Glaucoma, cataract</td>
</tr>
<tr>
<td>Cutaneous</td>
<td>Skin atrophy, striae, ecchymoses, impaired wound healing, acne, buffalo hump, hirsutism</td>
</tr>
<tr>
<td>Endocrine</td>
<td>Cushingoid appearance, diabetes mellitus, changes in lipid metabolism, enhanced appetite and weight gain, electrolyte abnormalities, hypothalamic-pituitary-adrenal axis suppression, suppression of gonadal hormones</td>
</tr>
<tr>
<td>Behavioral</td>
<td>Insomnia, psychosis, emotional instability, cognitive effects</td>
</tr>
</tbody>
</table>

ADVERSE EFFECTS AND MONITORING

Given the diversity of their mechanisms and sites of action, it is not surprising that glucocorticoids can cause a wide array of adverse effects (Table 60-6 and Figure 60-8). Most of these adverse effects cannot be avoided. However, the risk of most complications is dosage and time dependent; minimizing the quantity of glucocorticoid minimizes the risk of complications. Dose-related patterns of adverse effects are shown in Figure 60-8. GC treatment can cause a wide array of adverse effects, including pain, swelling, stiffness, physical disability, vasculitis, serositis, endothelial dysfunction, permeability, increased CV risk, cataract, glaucoma, hirsutism, skin thinning, gastric ulcer, and weight gain/obesity. These effects are due to the immunosuppressive and anti-inflammatory actions of glucocorticoids on cells, tissue, and organs, as shown in Figure 60-8. (Adapted from Buttgereit F, Burmester GR, Lipworth BJ: Optimised glucocorticoids therapy: the sharpening of an old spear, Lancet 365:801–803, 2005.)
effects of glucocorticoids have been described. Low-dose glucocorticoid therapy is safer than is commonly thought, and medium- to long-term glucocorticoid therapy in RA is associated with limited toxicity compared with use of placebo, but sensitivity for adverse effects differs among individuals. It is a clinical observation that some patients develop adverse effects after small doses of glucocorticoids, whereas other patients receive high doses without serious adverse effects. Apparent individual susceptibility to adverse effects does not seem to always parallel individual susceptibility to beneficial effects. Osteoporosis, diabetes, and cardiovascular disease are ranked by both patients and rheumatologists among the most worrisome adverse effects of glucocorticoids. However, the frequency and the severity of glucocorticoid-related adverse effects have seldom been studied systematically. A problem for nonrandomized studies looking at glucocorticoid-related adverse effects is bias by indication: patients with severe disease tend to take glucocorticoids more frequently than those with less severe disease, and the disease as well as the glucocorticoids can cause unfavorable signs and symptoms, on the other hand, glucocorticoids decrease disease activity and therewith influence the frequency and severity of disease-associated signs and symptoms (Figure 60-9).

Skeletal Adverse Effects

Osteoporosis

Osteoporosis is a well-known adverse effect of glucocorticoids that can be prevented to a large degree. International and national guidelines to minimize the occurrence of glucocorticoid-induced osteoporosis have been developed and are updated periodically. Preventive and therapeutic management of glucocorticoid-induced osteoporosis is discussed in detail in Chapter 99. In short, following the actual guideline consists of providing calcium and vitamin D supplementation and prescribing a bisphosphonate on indication.

Osteonecrosis

High-dose glucocorticoids given over longer periods are implicated as a cause of osteonecrosis, especially in children and patients with SLE. Vascular mechanisms seem to be involved. Ischemia possibly may be caused by microscopic fat emboli or impingement of the sinusoidal vascular bed by increased intraosseous pressure caused by fat accumulation. An early symptom is diffuse pain, which becomes persistent and increases with activity. Most frequently, hip or knee joints are involved; ankle and shoulder joints are involved less frequently. For early assessment, magnetic resonance imaging is the most sensitive investigative tool. Radionuclide bone scans provide less specific information. Plain radiographs are adequate only for follow-up. Treatment in the early stage includes immobilization and decreased weight bearing. Surgical decompression, joint replacement, or both follow this if needed. No preventive measures are known; awareness is the most important factor in early detection.

Myopathy

Weakness in proximal muscles, especially of the lower extremities, occurring within weeks to months after initiation of treatment with glucocorticoids, or after an increase in the dosage, may indicate steroid myopathy. It is often suspected but is infrequently found; it occurs almost exclusively in patients treated with high dosages (>30 mg/day prednisone or equivalent). Diagnosis is clinical and can be confirmed by a muscle biopsy specimen that reveals atrophy of type II fibers and lack of inflammation; no elevation of serum muscle enzymes is noted. Treatment consists of withdrawal of the glucocorticoid; if this is possible, a prompt decrease in symptoms may ensue. A rare syndrome of rapid-onset, acute myopathy, occurring within days after the start of high-dose glucocorticoids or pulse therapy, has been described; muscle biopsy specimens show atrophy and necrosis of muscle fibers.

Gastrointestinal Adverse Effects

Peptic Ulcer Disease

Data from the literature on upper gastrointestinal safety of oral glucocorticoids are inconclusive. The fact that glucocorticoids inhibit the production of COX-2 without hampering the production of COX-1 supports studies that found no increased risk. In other studies, a relative risk of serious upper gastrointestinal peptic complications of about 2 was found. When glucocorticoids are used in combination with NSAIDs, the relative risk of peptic ulcer disease and associated complications is about 4. Therefore in cases of co-medication with NSAIDs, consider co-treatment with a proton pump inhibitor, or prescribe a COX-1-sparing NSAID. In patients treated with glucocorticoids without concomitant use of NSAIDs, no indication for gastrointestinal protective agents exists, unless other risk factors for peptic complications are present.
**Other Gastrointestinal Adverse Effects**

Although many glucocorticoids usually are listed as one of the many potential causes of pancreatitis, evidence for such an association is weak and is difficult to separate from the underlying disease, such as vasculitis or SLE. Asymptomatic and symptomatic colonization of the upper gastrointestinal tract with *Candida albicans* is increased in patients treated with glucocorticoids, especially when other risk factors are present, such as advanced age, diabetes mellitus, and concomitant use of other immunosuppressive agents. Glucocorticoids may mask symptoms and signs usually associated with the occurrence of intra-abdominal complications, such as perforation of the intestine and peritonitis (e.g., as a complication of diverticulitis), and can lead to a delay in diagnosis with increased morbidity and mortality.

**Immunologic Adverse Effects**

At high doses, glucocorticoids diminish neutrophil phagocytosis and bacterial killing in vitro, whereas in vivo, normal bactericidal and phagocytic activities are found. Monocytes are more susceptible; during treatment with medium to high doses of glucocorticoids, bactericidal and fungicidal activity in vivo and in vitro is reduced. These factors may influence the risk of infection. From epidemiologic studies, treatment with a daily dose of less than 10 mg of prednisone or equivalent seems to lead to no or an only slightly increased risk of infection; however, if doses of 20 to 40 mg daily are used, the risk of infection is increased (relative risk of 1.3 to 3.6). This risk increases with increased dose and duration of treatment.

In a meta-analysis of 71 trials involving more than 2000 patients with different diseases and different doses of glucocorticoids, an increased relative risk of infection of 2 was found. The risk varied according to the type of disease being treated. Five of these trials involved patients with rheumatic diseases and showed no increased risk (relative risk of 1). The same was found in a double-blind, placebo-controlled, 2-year trial in patients with early RA, in which the effect of 10 mg of prednisone daily was compared with that of placebo. In one study, after adjustments were made for covariates, prednisone use dose dependently increased the risk of hospitalization for pneumonia. In patients treated with glucocorticoids, especially at high doses, clinicians should anticipate infections with usual and unusual organisms, realizing that glucocorticoids may blunt classic clinical features, thus delaying diagnosis.

**Cardiovascular Adverse Effects**

**Mineralocorticoid Effects**

Some glucocorticoids have mineralocorticoid actions (see Table 60-1), including reduced renal excretion of sodium and chloride and increased excretion of potassium, calcium, and phosphate. This activity may lead to edema, weight gain, increased blood pressure, and heart failure (caused by reduced excretion of sodium and chloride); cardiac arrhythmia (resulting from increased excretion of potassium); or tetany and electrocardiographic changes (related to hypocalcemia).

Low doses of glucocorticoids are not a cause of hypertension, in contrast to higher doses. No formal studies addressing the effects of glucocorticoids in previously hypertensive patients have been reported. Two randomized, controlled studies in patients with myocarditis and idiopathic cardiomyopathy showed no differences between placebo-treated or glucocorticoid-treated groups after 1 year or in survival at 2 and 4 years.

**Atherosclerosis**

Accelerated atherosclerosis and elevated cardiovascular risk have been reported in patients with SLE and in patients with RA. Glucocorticoids may enhance cardiovascular risk via their potentially deleterious effects on lipids, glucose tolerance, insulin production and resistance, blood pressure, and obesity. However, these conditions seem not to be adverse effects of low-dose glucocorticoids. Furthermore, atherosclerosis itself has been recognized as an inflammatory disease of arterial walls, for which glucocorticoids may be beneficial; glucocorticoids have been found to inhibit macrophage accumulation in injured arterial walls in vitro, possibly resulting in attenuation of the local inflammatory response. Low-dose glucocorticoids might also improve dyslipidemia associated with inflammatory disease.

However, the effects on lipids and other cardiovascular risk factors of low-dose glucocorticoids in inflammatory diseases probably are different from those of medium and high doses of glucocorticoids, or those of glucocorticoid therapy in noninflammatory diseases. This, along with the interplay of disease activity, glucocorticoids, and adverse effects (see Figure 60-9), makes it difficult to judge the net adverse effects of glucocorticoids on cardiovascular risk and lipids. The finding that a common haplotype of the glucocorticoid receptor gene is associated with heart failure, and that this association is mediated in part by low-grade inflammation, complicates this issue even further.

**Ocular Adverse Effects**

**Cataract**

Glucocorticoids tend to stimulate the formation of posterior subcapsular cataract especially, but the risk of cortical cataract also seems increased, with an odds ratio of 2.6. To some extent, the likelihood or severity of this adverse effect depends on dose and duration of treatment. In patients treated long term with glucocorticoids at a dosage of 15 mg or more of prednisone daily for 1 year, cataract is observed frequently; in patients receiving long-term therapy with less than 10 mg of prednisone daily, the percentage of cataract is less, but cataract may develop at dosages greater than 5 mg/day of prednisone equivalent. These cataracts are usually bilateral but progress slowly. They may cause glare disturbance but usually cause little visual impairment, except at early stages.

**Glaucoma**

By increasing intraocular pressure, glucocorticoids may cause or aggravate glaucoma. Patients with a family history
of open-angle glaucoma and patients with high myopia are probably prone to develop this adverse effect, especially when receiving high doses of glucocorticoids; checks of intraocular pressure are then warranted. If increased, patients need to be treated with medications that reduce intraocular pressure, often for a prolonged period after stopping the glucocorticoid. Topical application of a glucocorticoid in the eye has a more pronounced effect on intraocular pressure compared with systemic glucocorticoid therapy.

Dermal Adverse Effects

Clinically relevant adverse effects of glucocorticoids on skin include cushingoid appearance, easy bruising, ecchymoses, skin atrophy, striae, disturbed wound healing, acne, perioral dermatitis, hyperpigmentation, facial redness, mild hirsutism, and thinning of scalp hair. The physician often considers these changes to be of minor clinical importance, but they may be disturbing to the patient. No reliable data on the exact frequency of these adverse effects are available, but these adverse effects are dependent on duration of therapy and dose. Many physicians recognize immediately the skin of a patient who has been taking glucocorticoids on a long-term basis.

Endocrine Adverse Effects

Glucose Intolerance and Diabetes Mellitus

Glucocorticoids increase hepatic glucose production and induce insulin resistance by inhibiting insulin-stimulated glucose uptake and metabolism by peripheral tissues. Glucocorticoids probably also have a direct effect on beta cells of the pancreas, resulting in enhanced insulin secretion during glucocorticoid therapy. It may take only a few weeks before glucocorticoid-induced hyperglycemia occurs with low and medium glucocorticoid doses. One case-control, population-based study in previously nondiabetic subjects suggested an odds ratio of 1.8 for the need to initiate anti-hyperglycemic drugs during glucocorticoid therapy with doses of 10 mg or less of prednisone or equivalent per day. This risk increased with higher daily doses of glucocorticoids. The odds ratio was 3 for 10 to 20 mg, 5.8 for 20 to 30 mg, and 10.3 for 30 mg or more of prednisone or equivalent per day.

It is likely that risk is increased further in patients with other risk factors for diabetes mellitus, such as a family history of the disease, advanced age, obesity, and previous gestational diabetes. Postprandial hyperglycemia and only mildly elevated fasting glucose concentrations are characteristic of glucocorticoid-induced diabetes mellitus. Worsening of glycemic control can be expected in patients with established glucose intolerance or diabetes mellitus. Usually, glucocorticoid-induced diabetes is reversible when the drug is discontinued, unless clear glucose intolerance was pre-existing.

Fat Redistribution and Body Weight

One of the most notable effects of long-term endogenous or exogenous glucocorticoid excess is the redistribution of body fat. Centripetal fat accumulation with thin extremities is a characteristic feature of patients exposed to long-term high-dose glucocorticoids. Potential mechanisms include increased conversion of cortisone to cortisol in visceral adipocytes, hyperinsulinemia, and changes in expression and activity of adipocyte-derived hormones and cytokines, such as leptin and TNF. Protein loss resulting in muscle atrophy also contributes to the change in body appearance. Increased appetite influences body weight during glucocorticoid therapy, but patients with active inflammatory disease tend to lose weight, which can be prevented with disease control by drugs, including glucocorticoids. Trials in patients with RA given low-dose glucocorticoids for a prolonged period showed only minor effects on fat redistribution and body weight.

Dyslipidemia

See the earlier section, “Atherosclerosis.”

Suppression of the Hypothalamic-Pituitary-Adrenal Axis

In the section on effects of glucocorticoids on the hypothalamic-pituitary-adrenal axis, mechanisms of chronic suppression of the hypothalamic-pituitary-adrenal axis by administration of exogenous glucocorticoids are described. In such a situation, acute discontinuation of glucocorticoid therapy may lead to acute adrenal insufficiency with possible circulatory collapse and death. About 10 years after glucocorticoid therapy became available, the first well-documented case of adrenal insufficiency after withdrawal of exogenous glucocorticoid was reported. Acute cessation of glucocorticoid therapy without tapering is indicated for corneal ulceration by herpes virus, which can lead rapidly to perforation of the cornea, and glucocorticoid-induced acute psychosis. In these patients, assessment of adrenal responsiveness on a corticotropin test seems prudent. Not all patients with a blunted cortisol response have signs or symptoms of adrenal insufficiency, however.

Clinical signs and symptoms of chronic adrenal hypo-functioning are nonspecific and include fatigue and weakness, lethargy, orthostatic hypotension, nausea, loss of appetite, vomiting, diarrhea, arthralgia, and myalgia. These symptoms partially overlap glucocorticoid withdrawal symptoms, such as fatigue, arthralgia, and myalgia. When in doubt, measurements of serum cortisol levels and the corticotropin stimulation test are indicated. Glucocorticoid withdrawal symptoms are sometimes difficult to discriminate from symptoms of the primary disease, such as polymyalgia rheumatica. Because mineralocorticoid secretion remains intact via the renin-angiotensin-aldosterone axis, serious electrolyte disturbances are uncommon.

Adverse Behavioral Effects

Glucocorticoid treatment is associated with a variety of behavioral symptoms. Although most attention has been directed toward specific dramatic disturbances collectively described under the term glucocorticoid psychosis, less florid effects also occur that may cause distress to a patient and
warrant medical attention. Minor behavioral manifestations may also occur on withdrawal of glucocorticoids.

**Steroid Psychosis**

Overt psychosis is rare and usually is associated with high-dose glucocorticoids or glucocorticoid pulse therapy, but psychosis may also be a complication of the disease itself, especially SLE. This makes it difficult to distinguish in an individual SLE patient with psychosis whether the condition is a complication of the disease, of the therapy, or of both.

Isolated psychosis is seen in about 10% of glucocorticoid-related cases, and in most patients, affective disorders are present as well. Around 40% of cases of glucocorticoid-induced psychosis manifest as depression, whereas mania, often dominated by irritability, is predominant in 30% of cases. Psychotic symptoms usually start just after initiation of treatment (60% within the first 2 weeks, 90% within the first 6 weeks), and remission after drug dose reduction or withdrawal follows the same pattern. Although the data are largely anecdotal, individuals developing steroid psychosis frequently have had prior evidence of some dissociative symptoms. Occasionally, remission occurs without dose reduction.

**Minor Mood Disturbances**

Glucocorticoids have been associated with a wide variety of low-grade disturbances, such as depressed or elated mood (euphoria), insomnia, irritability, emotional instability, anxiety, memory failure, and other cognition impairments. Although the symptoms may not become severe enough for a specific diagnosis, they warrant attention—not only because they cause distress to the patient, but also because they may interfere with evaluation and treatment of the underlying disease. Most physicians recognize the occurrence of such symptoms in many glucocorticoid-treated patients; these symptoms may occur in varying degrees in up to 50% of treated patients within the first week. The exact incidence in rheumatic patients exposed to the usual doses of glucocorticoids is unknown; most series dedicated to mood disturbances studied high doses. It is important to inform patients about these minor mood disturbances before starting glucocorticoid therapy.

**Monitoring**

Glucocorticoid-related adverse effects have seldom been studied systematically. Mostly based on expert and patient opinion, recommendations have been formulated for monitoring low-dose glucocorticoid therapy. The conclusion is that in daily practice, standard care monitoring for serious diseases warranting glucocorticoid therapy need not be extended for patients on low-dose glucocorticoid therapy, except for monitoring for osteoporosis (follow national guidelines) and baseline assessments of fasting blood glucose and of risk factors for glaucoma, as well as a baseline check for ankle edema. Of course, for medium and high dosages, monitoring should be extended, not only to monitor for adverse effects of glucocorticoid therapy, but also to check for adverse effects of the concomitant medication and complications of the severe disease; for these glucocorticoid dosages, monitoring guidelines are being developed. In these situations, next to good clinical practice monitoring, including for instance blood pressure measurements, checks of ocular pressure and urine glucose specifically seem indicated. For clinical trials on glucocorticoids, it is advised to monitor and report more comprehensively and to sample more data on the spectrum, incidence, and severity of adverse events of glucocorticoids. If applied prudently, glucocorticoids are still one of the most relevant therapeutic tools in clinical medicine of the 21st century.

**Future Directions**

Although glucocorticoids have been used in clinical practice for many years, they still are the anchor drugs in autoimmune and inflammatory diseases and vasculitides. In contrast with their established use, there is a paucity of data on the spectrum, incidence, and severity of adverse effects of glucocorticoids at different dosages and in different diseases. To develop evidence-based guidelines and to evaluate the adverse effects of new compounds with glucocorticoid actions that are being developed, additional research into molecular mechanisms and continued collection of data are needed.

**References**


58. Svensson B, Boonen A, Albertsson K, et al: Low-dose prednisolone in addition to the initial disease-modifying antirheumatic drug in...


The references for this chapter can also be found on www.expertconsult.com.
References


Traditional DMARDs: Methotrexate, Leflunomide, Sulfasalazine, Hydroxychloroquine, and Combination Therapies

AMY C. CANNELLA • JAMES R. O’DELL

more than 50 years ago to treat cancer. Over the last quarter century, it has become the disease-modifying antirheumatic drug (DMARD) of choice in the treatment of RA and is used in many other rheumatic diseases as well.

Chemical Structure

MTX is a structural analogue of folic acid and has substitutions in the pteridine group and para-aminobenzoic acid structure (Figure 61-1). The structure of folic acid (pteroylglutamic acid) consists of three elements: a multi-ring pteridine group, linked to a para-aminobenzoic acid, which is connected to a terminal glutamic acid residue.

Actions of Methotrexate

Because MTX is a folate analogue, it enters cells via a reduced folate carrier (RFC). Leucovorin competes with MTX for uptake using the same RFC; however, folic acid enters cells via another group of transmembrane receptors called folate receptors (FRs).

FRs may be upregulated in cells with increased metabolic activity, including synovial macrophages, and serve as a second conduit for MTX influx. MTX efflux occurs via members of the adenosine triphosphate (ATP)-binding cassette (ABC) family of transporters, specifically ABCC1-4 and ABCG2.

Genetic polymorphisms may affect MTX transporter proteins (influx and efflux) and can result in a variable MTX response and toxicity profile. Furthermore, multidrug resistance proteins have been identified that transport MTX, folic acid, and leucovorin out of cells, leading to MTX resistance.

Once inside the cell, naturally occurring folates as well as MTX undergo polyglutamation by the enzyme folylpolyglutamyl synthetase (FPGS). Polyglutamation of MTX (MTX-PG) is essential to prevent efflux of MTX, which easily occurs in the monoglutaminated state. MTX-PG has several key inhibitory effects on intracellular enzymes, which result in its postulated anti-inflammatory and anti-proliferative (immunosuppressive) mechanisms: (1) Inhibition of aminoimidazole carboxamide ribonucleotide (AICAR) transformylase (ATIC) results in increased intracellular and extracellular adenosine, (2) inhibition of thymidylate synthetase (TYMS) results in decreased pyrimidine

KEY POINTS

Methotrexate is one of the most durable and frequently used disease-modifying antirheumatic drugs (DMARDs) in monotherapy as well as the cornerstone of combination therapy for rheumatoid arthritis (RA).

Leflunomide, sulfasalazine, and hydroxychloroquine are effective therapies in RA and are commonly employed in combination therapy.

Although the precise mechanisms of action of the traditional DMARDs are incompletely understood, most have both anti-inflammatory and immunomodulatory actions.

Choice of DMARD therapy should be tailored to the individual patient, with attention given to age, fertility plans, concomitant medications, and comorbidities.

Toxicity from DMARD therapy can cause significant morbidity and rarely mortality; thus, appropriate dosing and monitoring for toxicity are essential.

Combination therapy in RA can be more effective than mono-DMARD therapy in groups of patients with early and established RA.

The appropriate timing and combinations of DMARD therapy in individual patients is still not defined.

METHOTREXATE

KEY POINTS

An important mechanism of action for methotrexate (MTX) is the upregulation of adenosine, which is a potent inhibitor of inflammation.

MTX is polyglutamated in cells, and this is responsible for its long therapeutic effect.

The effects of MTX may be enhanced by splitting the dose (within a 12-hour window) when levels greater than 15 mg/wk are used, or by using a subcutaneous route of administration.

Concomitant use of folic acid abrogates some of the side effects of MTX without decreasing efficacy.

The dose of MTX must be adjusted for reduced renal function.

Although rare, MTX pneumonitis is a serious and potentially fatal complication of therapy.

It would be difficult to overstate the importance of methotrexate (MTX) in the contemporary management of rheumatic disease and, in particular, rheumatoid arthritis (RA). Because of its antiproliferative effects, MTX was introduced
functions after stressful injurious stimuli. Adenosine, a potent inhibitor of inflammation, induces vasodilation. Adenosine’s anti-inflammatory effects include regulation of endothelial cell inflammatory functions, including cell trafficking, counterregulation of neutrophils and dendritic cells, and cytokine modulation of monocytes and macrophages. Adenosine receptor ligation on monocytes and macrophages suppresses interleukin (IL)-12, a strong proinflammatory cytokine. Adenosine also suppresses the proinflammatory mediators tumor necrosis factor (TNF), IL-6, IL-8, macrophage inflammatory protein (MIP)-1α, leukotriene (LT)B₄, and nitric oxide and enhances production of the anti-inflammatory mediators IL-10 and IL-1 receptor antagonist. Furthermore, adenosine receptor-mediated processes result in inhibition of the synthesis of collagenase, including tissue inhibitors of metalloproteinases. In sum, adenosine appears to promote a self-limiting, healthy immune response, hastening the transition from neutrophil-mediated inflammation to a more efficient and highly specific dendritic cell–mediated response. Ultimately adenosine leads to the resolution of inflammation by downregulation of macrophage activation and promotes a shift from a T helper (Th)1 cell to a T helper (Th)2 cell response.

Evidence that the anti-inflammatory effects of MTX are mediated through adenosine has accumulated in in vitro and in animal studies. However, owing to adenosine’s short blood half-life of 2 seconds and MTX’s long latent period for active metabolites that modulate adenosine, it has been difficult to demonstrate changes in blood adenosine levels directly related to MTX. Recent evidence using forearm blood flow as a surrogate marker for adenosine release in RA patients treated with MTX demonstrated that MTX inhibits deamination of adenosine and potentiates adenosine-induced vasodilation. Demonstration of altered adenosine kinetics in patients treated with MTX coupled with adenosine’s known anti-inflammatory effects lends further credence to the hypothesis that MTX increases extracellular adenosine, which likely mediates some of the anti-inflammatory effects of MTX.

In addition to vasodilation, adenosine’s cardiovascular effects include negative inotropic and chronotropic cardiac synthesis, and (3) inhibition of dihydrofolate reductase (DHFR) results in inhibition of transmethylation reactions essential for cellular functioning (Figure 61-2).

Inhibition of ATIC by MTX-PG leads to accumulation of AICAR and ultimately to increased levels of adenosine. Three possible mechanisms are postulated and likely work in combination: (1) AICAR inhibition of adenosine monophosphate (AMP) deaminase leads to excess production of adenosine from AMP; (2) AICAR inhibition of adenosine deaminase (ADA) leads to decreased breakdown of adenosine to inosine; and (3) AICAR stimulation of the ecto-5′-nucleotidase converts extracellular AMP to adenosine (Figure 61-3).

Adenosine, a purine nucleoside, has been termed a “retaliatory metabolite” because of its tissue protective

![Figure 61-1](image1.png) **Figure 61-1** Chemical structure of folic acid and methotrexate.

![Figure 61-2](image2.png) **Figure 61-2** Methotrexate (MTX) enters cells primarily via the reduced folate carrier (RFC) but can use the folate receptor (FR). Once inside the cell, it becomes polyglutamated and can interfere with several cellular enzymes, including 5-aminimidazole-4-carboxamide ribonucleotide (AICAR) transformylase (ATIC), thymidylate synthetase (TYMS), and dihydrofolate reductase (DHFR). ABC, ATP-binding cassette; FPGH, folylpolyglutamate hydrolase; FPGS, folyl-polyglutamyl synthetase; MTX-PG, polyglutamation of MTX.
accumulation of polyamines in B cells is associated with agents, including ammonia and hydrogen peroxide, of polyamines by mononuclear cells gives rise to toxic synovial fluid and tissue in urine, Spermine and spermidine have been shown to accumulate upon each varies in vivo and function, although specific cellular dependence upon DHFR, to generate THF are required for cellular survival and function, and interference with intracellular transmethylation reactions. 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) acts as a proximal methyl donor for several reactions by donating the methyl group for the conversion of homocysteine to methionine. Methionine is then converted to S-adenosylmethionine (SAM), which acts as a methyl donor for the following: methylation of RNA, DNA, amino acids, proteins, and phospholipids, and synthesis of the polyamines spermidine and spermine. Upon demethylation of SAM to S-adenosylhomocysteine (SAH), SAH is converted to adenosine and homocysteine. Methylation products that are dependent upon SAM, and thus indirectly upon DHFR, to generate THF are required for cellular survival and function, although specific cellular dependence upon each varies (see Figure 61-3). The role of polyamines deserves further discussion. Spermine and spermidine have been shown to accumulate in urine, in peripheral blood mononuclear cells, and in synovial fluid and tissue in patients with RA. Metabolism of polyamines by mononuclear cells gives rise to toxic agents, including ammonia and hydrogen peroxide, which may impair lymphocyte function. Additionally, accumulation of polyamines in B cells is associated with enhanced production of rheumatoid factor (RF) in vitro, and incubation of these cells with methotrexate diminishes their ability to secrete both immunoglobulin and RF. These effects are seen with high in vitro concentrations of MTX and may not translate into the in vivo therapeutic effects of MTX in RA.

In addition, MTX inhibits methylation of 5′-deoxyuridylate (dUMP) into 5′-deoxythymidylate (dTMP) by TYMS, resulting in a further mechanism for disruption of DNA synthesis and proliferation of anti-inflammatory cells. This effect has been shown in vitro in human peripheral blood mononuclear cells incubated with low concentrations of MTX. Cell cycle disruption may lead to apoptosis of mononuclear cells via CD95 (APO-1/Fas) ligand-dependent and -independent mechanisms.

Therefore, inhibition of transmethylation reactions may lead to MTX efficacy via antiproliferative and anti-inflammatory mechanisms. Disruption of DNA, RNA, amino acid, and phospholipid synthesis results in its antiproliferative effect, which may be mediated via cellular apoptosis. Decreased levels of polyamines may downregulate the production of toxic agents, as well as RF secretion, leading to its anti-inflammatory effect.

In theory, the anti-inflammatory and antiproliferative properties of MTX already described should make it a potent inhibitor of the immune response that characterizes many rheumatic diseases. Indeed, MTX has become the cornerstone of therapy for RA and is efficacious in multiple other rheumatic diseases. Direct evidence for the immunomodulatory effects of MTX exists, whether studied in in vitro or in vivo systems.

Treatment with MTX has been shown to modulate monocyctic and lymphocytic cytokines and their inhibitors.

Figure 61-3  Simplified schema of the effects of polyglutamation of methotrexate (MTX-PG) on intracellular and extracellular adenosine production and interference with intracellular transmethylation reactions. 5′-NT, 5′-nucleotidase; ADA, adenosine deaminase; ADP, adenosine deaminase; AICAR, α-N-formylglycinamide ribonucleotide; IMP, inosine monophosphate; RNA, ribonucleic acid; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine; THF, tetrahydrofolate.
MTX has been shown to inhibit proinflammatory cytokine IL-1 secretion and to induce the IL-1 receptor antagonist, effectively inhibiting cellular responses to IL-1.\textsuperscript{35,36} Soluble TNF receptor (sTNFR p75) synthesis upregulation has also been shown as a result of MTX treatment from cultured monoblastic leukemia cells, which results in a diminished TNF inflammatory effect.\textsuperscript{37} MTX also inhibits production and secretion of the proinflammatory cytokine, IL-6, by cultured human monocytes.\textsuperscript{38,39} Reverse transcriptase polymerase chain reaction has been used to study the effects of MTX on gene expression for lymphocytic cytokines.\textsuperscript{40,41} MTX increases anti-inflammatory Th2 cytokine (IL-4 and IL-10) gene expression and decreases proinflammatory Th1 cytokine (IL-2 and interferon [IFN]-γ) gene expression in peripheral blood mononuclear cells (PBMCs) of patients with RA.\textsuperscript{42}

Prostaglandins (PGs) and leukotrienes (LTs) are important mediators of joint destruction in RA. MTX has been shown to modulate the inflammatory enzymes cyclooxygenase (COX) and lipooxygenase (LOX), and their products PG and LT. Thromboxane B\textsubscript{2} and prostaglandin E\textsubscript{2} activities were reduced in the whole blood of RA patients treated with MTX when compared with healthy controls.\textsuperscript{43} MTX also reduces LTB\textsubscript{4} synthesis by neutrophils, resulting in a decrease in total plasma LTB\textsubscript{4} levels in patients with RA treated weekly with MTX.\textsuperscript{44} In addition to possible direct effects on COX and LOX, MTX has been shown to exert an inhibitory effect on neutrophil chemotaxis, which may result in a further reduction of these enzymes in sites of inflammation.\textsuperscript{45}

Tissue destruction at sites of inflammation is thought to be related to increased synthesis and activity of proteolytic enzymes released by inflammatory cells, particularly in RA. MTX treatment has been shown to reduce gene expression of collagenase, metalloproteinase-1, and stromelysin, and to upregulate expression of tissue inhibitor of metalloproteinase-1 (TIMP-1).\textsuperscript{46} MTX may exert direct effects on messenger RNA (mRNA) for certain enzymes, such as collagenase. MTX also likely exerts indirect effects on gene expression via upstream cytokine modulation (IL-1 and IL-6), in the case of matrix metalloproteinase (MMP)-1 and TIMP-1.\textsuperscript{46}

**Pharmacology**

**Absorption and Bioavailability**

At low doses, MTX can be administered either orally or parenterally (subcutaneous or intramuscular), and absorption is rapid, peaking at 1 to 2 or 0.1 to 1 hour, respectively. The absorption of low-dose oral and parenteral MTX (<15 mg/wk) is roughly equivalent, but once the oral dose exceeds 15 mg/wk, absorption diminishes by as much as 30%.\textsuperscript{47} Absorption is not reduced by concomitant food intake, except for milk, which may be inhibitory,\textsuperscript{48} but may be reduced in the setting of intestinal pathology, such as inflammatory bowel disease or malabsorptive conditions.

Orally administered MTX is absorbed via the GI tract and passes through the liver via the portal vein; parenterally administered MTX passes through the liver via the hepatic artery. Although not prospectively studied in RA patients receiving long-term MTX treatment, the parenteral route should have diminished potential for hepatotoxicity. This effect has been seen in a retrospective study wherein more elevations in transaminases were noted when oral MTX was administered to the same individuals versus when given parenterally.\textsuperscript{49} A recent study at equivalent MTX doses showed improved efficacy in RA clinical end points for parenteral over orally administered MTX.\textsuperscript{50}

Forty-one RA patients who received 10 mg/m\textsuperscript{2} of oral MTX had a mean bioavailability of 70% with a range of 40% to 100%.\textsuperscript{51} The mean absorption time was 1.2 hours. Four hours after MTX administration, synovial fluid concentrations equal serum levels.\textsuperscript{52} A recent study of high-dose oral MTX (median dose, 30 mg/wk) has shown that mean bioavailability is improved by splitting the dose by 8 hours compared with one single dose (0.90 and 0.76, respectively).\textsuperscript{53} The pharmacokinetics of subcutaneous MTX are equivalent to those of intramuscular MTX; maximum serum concentration is attained within 2 hours of injection by either route.\textsuperscript{54} Also, bioavailability is equivalent between tablets and orally administered parenteral solution.\textsuperscript{54}

**Distribution and Half-Life**

MTX is 50% to 60% bound to plasma proteins and has a half-life of approximately 6 hours.\textsuperscript{51} An increase in free MTX caused by displacement from albumin by more highly protein-bound drugs such as aspirin, nonsteroidal anti-inflammatory drugs (NSAIDs), and sulfonamides can occur. This is generally of limited clinical significance with low MTX doses because the increase in free MTX is usually only modest.

MTX accumulates in third space fluids, which can serve as a reservoir for redistribution into the circulation long after the last dose is administered.\textsuperscript{55} Caution should be used when administering MTX to patients with pleural effusions or ascites. Furthermore, unexpectedly high levels of MTX have been seen in patients with bladder cancer who have undergone ileal conduit surgery owing to enhanced intestinal absorption through the newly fashioned conduit.\textsuperscript{56}

The biologically active form of MTX occurs after intracellular polyglutamation (MTX-PG). MTX undergoes up to five polyglutamations, and recent studies have looked at this aspect of MTX pharmacology. Once on a stable dose of MTX, the median time until 90% of the maximum steady-state concentration of MTX-PG was reached was found to be 27.5 weeks (range, 6.6 to 62.0 weeks).\textsuperscript{57}

**Elimination**

Most MTX is excreted in the urine within the first 12 hours after administration, except for MTX-PG. MTX undergoes some hepatic metabolism by the enzyme aldehyde oxidase to the 7-hydroxymethotrexate metabolite; this metabolite has unknown significance in RA. MTX and metabolites are excreted by the kidney by glomerular filtration and proximal tubular secretion but also undergo distal tubular reabsorption. The estimated median half-life of elimination of MTX-PG is 3.1 weeks (range, 0.94 to 4.1 weeks), and MTX-PG is undetectable at 15 weeks.\textsuperscript{57} MTX-PG is the most common subtype seen (30% of total MTX-PG) and has a median half-life elimination of 4.1 weeks.
Indications

Rheumatoid Arthritis. The efficacy of MTX in RA has been clearly established. Four well-designed, blinded, placebo-controlled trials published in 1984 and 1985 had a tremendous impact on the treatment of RA. These trials varied in design and duration: Two of these trials used oral MTX and two used intramuscular (IM) MTX, two trials had a crossover and two were parallel, and the duration of treatment varied from 6 to 28 weeks. Although the design and duration of therapy in these trials varied, the conclusions did not, as all showed MTX to be superior to placebo in the short-term treatment of RA. A meta-analysis of these trials by Tugwell and coworkers showed that MTX-treated patients had a 37% greater improvement in swollen joint and tender joint scores, a 39% greater improvement in joint pain, and a 46% greater improvement in morning stiffness. MTX was generally well tolerated in these trials; withdrawal rates ranged from 0 to 32% and were mostly related to minor toxicities (i.e., stomatitis, nausea). Taken together, the results of these trials firmly established MTX as an effective therapy for the treatment of RA, at least in the short term.

Numerous trials have compared MTX with other DMARDs. A meta-analysis done by Felson and co-workers showed that MTX was superior to placebo, auranofin, and probably hydroxychloroquine (HCQ), and was comparable with penicillamine, sulfasalazine, and IM gold. No trial has ever suggested that any other synthetic DMARD is superior to MTX.

Accumulating evidence suggests that the short-term benefit of most DMARDs is not sustained, and few patients continue to take these drugs after 3 years. MTX appears to have the best durability. Pincus and colleagues have shown that 60% of patients continued MTX at 5 years, compared with less than 25% for penicillamine, gold, HCQ, and azathioprine. Of all the DMARDs, MTX appears to have the best efficacy-to-toxicity ratio. However, despite all the favorable efficacy reports, MTX alone rarely induces remissions of RA, and it has become the cornerstone of combinations of DMARD therapies as discussed later.

Rheumatoid Arthritis–Related Conditions. MTX has been used successfully in treating Felty’s syndrome and the large granular lymphocyte syndrome when it is found in patients with RA. Improvement in neutrophil count occurs within 4 to 8 weeks of MTX initiation in both cases. MTX has been used successfully in adult-onset Still’s disease and for the cutaneous vasculitis of RA.

Juvenile Idiopathic Arthritis. MTX is efficacious in juvenile idiopathic arthritis. A definitive, randomized, placebo-controlled trial demonstrated that MTX at a dose of 10.0 mg/m² was superior to 5.0 mg/m² or placebo. Sixty-three percent of children receiving the higher dose (10.0 mg/m²) of MTX improved, compared with 32% in the lower-dose group (5.0 mg/m²) and 36% in the placebo group.

Psoriatic Arthritis (PsA). Numerous prospective and retrospective trials have showed a benefit for MTX in PsA. The largest double-blind randomized trial compared weekly oral MTX with placebo, and showed statistically significant results only for physician global assessment of arthritis activity and the amount of affected skin surface area; however, this study was small and may have been underpowered to detect differences in joint count, pain, and swelling. Despite the paucity of randomized controlled trial data, MTX remains a commonly used systemic agent in the treatment of PsA.

Systemic Lupus Erythematosus (SLE). MTX has been shown to be efficacious in controlling cutaneous and/or articular manifestations of SLE, particularly in disease resistant to antimalarial or requiring high doses of systemic steroids. Concomitant folic acid should be administered to abrogate the elevated levels of homocysteine that may be a side effect of MTX therapy, and is considered a risk factor for cardiovascular disease in SLE. The role of MTX in treating more severe SLE involvement, including renal, hematologic, or central nervous system disease, has not yet been established. Extreme caution should be employed in patients with renal disease.

Vasculitis. MTX in conjunction with corticosteroids has shown efficacy in treating early and non–life-threatening granulomatosis with polyangiitis (formerly Wegener’s granulomatosis), including upper airway disease and mild renal disease. In addition to induction of remission, MTX has been shown to maintain remission in granulomatosis with polyangiitis, although vigilance for relapse is warranted. Despite a lack of well-designed studies, MTX has been shown to be efficacious in corticosteroid-resistant Takayasu’s arteritis and in relapsing polychondritis.

The use of MTX for polymyalgia rheumatica (PMR) and giant cell arteritis (GCA) has been controversial. Multiple open-label studies have shown differing results. Recently, randomized, placebo-controlled trials (RCTs) of MTX in addition to corticosteroids in PMR and GCA have shown conflicting results. Thus, the routine use of MTX in either PMR or GCA has not been adopted, but some advocate for its use in an effort to more rapidly taper corticosteroid in those patients with intolerable side effects.

Inflammatory Myopathies. A review of the published reports of MTX and the inflammatory myopathies polymyositis (PMS) and dermatomyositis (DMS) shows overall positive results. However, despite the frequent use of MTX in the inflammatory myopathies, a recent Cochrane database review reveals a paucity of well-designed trials.

Other Rheumatic Diseases. MTX has been used in systemic sclerosis. One RCT looking at MTX use in early systemic sclerosis showed a trend in benefit for skin scores and pulmonary diffusion capacity and a significant benefit for physician global assessment; a second RCT in established systemic sclerosis showed significant benefit for skin scores and total creatinine clearance. In addition, prospective trials have shown that MTX is efficacious in the treatment of corticosteroid-resistant multisystem sarcoidosis, and a recent RCT showed that if initiated early in sarcoidosis, MTX is an effective steroid-sparing agent. MTX has also been shown to be effective as primary treatment and as a corticosteroid-sparing agent in inflammatory ocular disease. Finally, MTX in combination with corticosteroids is effective in the treatment of multicentric reticulohistiocytosis.

Dose and Drug Administration

MTX is available as 2.5, 5, 7.5, 10, and 15 mg tablets, and as a solution of 25 mg/mL for subcutaneous or intramuscular
injection. The starting dose is usually 5 to 10 mg given as a single weekly dose. More frequent administration is associated with a significantly increased risk of liver toxicity.50 If the oral dose of MTX exceeds 15 mg, consideration should be given to splitting the dose, with each half given 6 to 12 hours apart, for improved bioavailability. The dosage of MTX can be escalated, usually every 4 to 8 weeks to 25 mg/wk to achieve the desired clinical response. MTX may be administered orally via tablet or parenteral solution; the latter is less costly. Because its bioavailability is variable and appears to decrease at higher doses, parenteral MTX is generally recommended if patients have active disease despite oral doses of approximately 20 mg/wk. Recent data have shown that parenteral MTX is clinically superior to orally administered MTX, and that this may be a preferred route of administration, especially if oral MTX is not optimally effective.50

Concomitant administration of folic acid (1 to 3 mg/day) decreases the frequency of toxicities, including mucositis, nausea, hematologic abnormalities, and liver enzyme elevations, without seeming to interfere with clinical efficacy.99,100 Folic acid administration also decreases hyperhomocysteinemia in patients on MTX, and this may be important to help decrease the already high cardiovascular risk of patients with RA. Low-dose folic acid has also been used and can markedly reduce MTX toxicity in rheumatic disease therapy without interfering with efficacy if given in doses of 2.5 to 5 mg/wk and if not administered until 24 hours after the MTX dose. Because folic acid is widely available and less expensive, it is preferred by most.

Measurement of MTX-PG levels is commercially available (MTXGlut). It would be valuable to have a marker to predict response and adverse events associated with MTX therapy, but mixed results have been obtained in the search for a trend in the dose-response relationship for MTXGlut levels and RA disease activity. A recent study showed no relationship between MTXGlut concentrations and reduced disease activity in RA.101 Furthermore, no relationship was identified between MTXGlut levels and adverse events. Disease activity was influenced by red blood cell (RBC) folate level, and further study is warranted to determine whether this may serve as a marker for MTX efficacy.

### Geriatric Patients

Patients older than 65 years represent a special subset of patients receiving pharmacotherapy. Pharmacokinetic profiles, including drug distribution, are changed in the elderly as the result of decreases in end-organ blood flow and lean body mass, decreased hepatic drug metabolism, and decreased renal drug excretion. Furthermore, these patients are more likely to have multiple comorbidities, polypharmacy, noncompliance, increased risk for dosage errors, and limited access to medication for financial reasons.102

In practice, recommended doses should be reduced when therapy is initiated and should be adjusted for renal function based on creatinine clearance (CrCl).103 The serum creatinine may be a misleading measure of renal function in older patients owing to an overall reduction in lean muscle mass. Dosing recommendations are as follows: Initial doses should be around 5 to 7.5 mg/wk and should not exceed 20 mg/wk. Dosage adjustments for CrCl are as follows: For a CrCl of 61 to 80 mL/min, reduce the dose by 25%; for a CrCl of 51 to 60 mL/min, reduce the dose by 30%; for a CrCl of 10 to 50 mL/min, reduce the dose by 50% to 80%; and for a CrCl less than 10 mL/min, avoid use104 (Table 61-1).

### Pediatric Patients (See Table 61-1)

MTX is commonly used in the pediatric population. It can be given orally or subcutaneously, with a starting dose of 0.3 to 1 mg/kg/dose once weekly. It is common to start at 0.3 mg/kg/dose and escalate to a dose of 25 mg/wk. It is suggested that at doses greater than 15 mg/wk, parenteral application should be considered because of better bioavailability and tolerability. Toxicity monitoring is similar to adult recommendations.

### Toxicity

Despite initial concerns, when given once a week in doses used for rheumatic diseases and monitored correctly, MTX is very well tolerated. Some of the toxicities of MTX (stomatitis, nausea, bone marrow depression) are dose

### Table 61-1 Special Considerations

<table>
<thead>
<tr>
<th>Fertility</th>
<th>Pregnancy</th>
<th>Lactation</th>
<th>Elderly</th>
<th>Pediatrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methotrexate</td>
<td>W: no effect; M: reversible sterility; stop 3 mo before conception</td>
<td>Contrainicated; FDA category X; abortifacient; teratogenic</td>
<td>Contrainicated; present in breast milk</td>
<td>Lower initial dose (5-7.5 mg/wk); dose based on CrCl</td>
</tr>
<tr>
<td>Leflunomide</td>
<td>No effect; test levels before conception; may require washout</td>
<td>Contrainicated; FDA category X; embryolethal; teratogenic</td>
<td>Contrainicated; unknown concentrations in breast milk</td>
<td>No dosage adjustment required</td>
</tr>
<tr>
<td>Sulfasalazine</td>
<td>W: no effect; M: reversible sterility</td>
<td>Relatively safe; FDA category B, C</td>
<td>Relatively safe; present in breast milk</td>
<td>No dosage adjustment required</td>
</tr>
<tr>
<td>Hydroxychloroquine</td>
<td>No effect</td>
<td>Relatively safe; FDA category C</td>
<td>Relatively safe; present in breast milk</td>
<td>No dosage adjustment required</td>
</tr>
</tbody>
</table>

CrCl, Creatinine clearance; FDA, U.S. Food and Drug Administration; M, men; W, women.
dependent, appear to be related to folate deficiency, and respond to folate replacement. Other toxicities appear to be idiosyncratic or allergic and in most cases require discontinuation of MTX (e.g., pneumonitis). Still other toxicities, such as liver fibrosis and cirrhosis, appear to be multifactorial and may depend on the presence of concomitant risk factors, total dose, and frequency of administration.

### Gastrointestinal and Hepatic Side Effects

Gastrointestinal symptoms, including dyspepsia, nausea, and anorexia, are common, occurring in up to 20% to 70% of patients within the first year of therapy. These symptoms may be attenuated by adding folic acid or by changing to a parenteral dosing regimen.

The risk of significant liver toxicity appears to be lower when MTX is given once weekly to patients who abstain from alcohol consumption and are monitored carefully, and is on the order of 1 case per 1000 after 5 years of use. However, a recent review of a large North American database of patients with RA and PsA found that elevations in aspartate aminotransferase (AST)/alanine aminotransferase (ALT) greater than 1 times the upper limit of normal (1 × ULN) occurred in 22% and 31% of patients taking MTX or MTX and leflunomide. Elevations greater than 2 × ULN occurred in 1% to 2% on monotherapy and in greater than 5% on combination therapy. Elevated liver function tests (LFTs) were more likely in patients with PsA. Alcohol consumption, α1-antitrypsin deficiency, morbid obesity, diabetes, concomitant hepatotoxic drugs, and chronic hepatitis B or C have all been implicated as possible risk factors for MTX toxicity.

### Hematologic Side Effects

Bone marrow toxicity, in most cases, is dose dependent and responds to folic acid administration. Pancytopenia, leukopenia, anemia, and thrombocytopenia can occur, but are rare. In a review by Gutierrez-Urena and associates, clinically significant pancytopenia was found to develop in up to 1% to 2% of RA patients on MTX therapy. Severe, life-threatening bone marrow toxicity can be treated with folinic acid (leucovorin) and, if necessary, granulocyte-stimulating factor (GSF). Because the elimination of MTX is dependent on the kidney, decreases in renal function may precipitate bone marrow toxicity in patients who have been previously stable. Additional risk factors include hypoalbuminemia, dosing errors, and concomitant use of probenecid or trimethoprim/sulfamethoxazole (TMP/SMX).

### Pulmonary Side Effects

Five clinical pulmonary syndromes have been associated with MTX treatment: acute interstitial pneumonitis (hyper-sensitivity pneumonitis), interstitial fibrosis, noncardiogenic pulmonary edema (seen in high-dose treatment for malignancy with rare reports in RA), pleuritis and pleural effusions, and pulmonary nodules. Time lapse from initiation of therapy to cumulative dose before the onset of pulmonary toxicity is extremely variable, at 1 to 480 weeks and 7.5 to 3600 mg of MTX, respectively. MTX-induced pulmonary disease is rare and is difficult to quantify, but estimates suggest an incidence of 3.9 cases per 100 patient-years of MTX exposure and a prevalence of 2.1% to 5.5%.

Patients generally present with shortness of breath, tachypnea, dry cough, and fever. Chest radiographs most typically show a bilateral interstitial infiltrate (although this varies). Infectious causes, including opportunistic organisms, must always be ruled out. If routine evaluations for infection, including sputum studies, and for other medical conditions to explain the pulmonary symptoms are negative, bronchoscopy with bronchoalveolar lavage and transbronchial biopsy is recommended. If MTX pulmonary toxicity is suspected, MTX should be discontinued and supportive treatment initiated with the use of corticosteroids in more severe cases. Some patients with pulmonary toxicity have been successfully restarted on MTX, but clinicians have reported mortality in up to 50% of retreated patients.

Factors that appear to predispose to MTX lung toxicity include age, blue collar occupation, smoking (in women), diabetes, pleuropulmonary rheumatoid disease, and skin rashes from MTX.

### Malignancies

The induction of malignancies by MTX is a concern, and several studies have examined this question with conflicting conclusions. Recently, reports of lymphoma in MTX-treated RA patients have appeared. Because the incidence of lymphoma is increased in RA patients anyway, these reports are difficult to interpret. The case for a causative role of MTX has been strengthened, however, because a number of these cases have been B cell lymphomas of the type commonly seen in association with immunosuppression (associated with Epstein-Barr virus) and that may regress after discontinuation of MTX. Subsequently, lack of a causal relationship between MTX treatment and the development of lymphoma has been seen in two large series of RA patients—one prospective study and one retrospective study. The potential benefits of MTX for most RA patients thus far outweigh these statistically small risks.

### Miscellaneous

#### Methotrexate Flu

Patients taking MTX may describe flu-like symptoms shortly after taking their weekly dose. Nausea, low-grade fevers, myalgias, and chills are the most common signs of the so-called methotrexate flu. These side effects usually respond to supplementation with folic acid, decreasing the dose, switching from oral to parenteral administration, or changing the time of the dose (so the patient takes MTX right before going to bed).
**Nodulosis.** The development of, or increase in, the number or size of rheumatoid nodules has been reported to occur in RA patients treated with MTX with a prevalence of up to 8%. This may occur in rheumatoid factor-negative patients and in those in whom the synovitis is under excellent control. The mechanism of this nodule formation has been suggested to be due to an increase in adenosine, which appears to promote nodule formation. Conversely, nodules have been reported to decrease during MTX therapy.

**Vasculitis.** Despite efficacy in the treatment of the cutaneous vasculitis associated with RA, leukocytoclastic vasculitis has also been attributed to MTX therapy.

**Fertility, Pregnancy, and Lactation**

MTX does not seem to adversely affect female fertility but can cause reversible sterility in men. Women and men should discontinue MTX for at least 3 months before attempting to conceive, because of its large distribution and long half-life in the liver. Folic acid supplementation is essential before conception. MTX is included in U.S. Food and Drug Administration (FDA) Pregnancy Category X and is contraindicated in pregnancy. Women of childbearing age who are considered for MTX therapy should receive extensive counseling regarding teratogenic risk and should be placed and maintained on adequate contraception, before therapy is begun. Toxicities include fetal abnormalities such as “aminopterin syndrome” (multiple craniofacial, limb, and central nervous system abnormalities) and embryonic or fetal loss, and MTX at high doses (1 mg/kg) is an effective abortifacient. MTX is also contraindicated during pregnancy because small amounts are excreted in breast milk (see Table 61-1).

**Toxicity Monitoring**

The American College of Rheumatology (ACR) has recently revised recommendations for the use of DMARDs, and these serve as an excellent resource. Toxities that require monitoring include myelosuppression, hepatotoxicity, and pulmonary toxicity. Baseline evaluation should include a complete blood count (CBC) with platelets, hepatitis B and C serology in high-risk patients, liver transaminases, and creatinine. Although these guidelines make no recommendations on the need for a baseline chest x-ray (CXR), this is a reasonable approach. Liver biopsies are not routinely recommended before MTX is initiated. The rare patients whom one wants to treat with MTX despite abnormalities in screening laboratory or other significant risk factors may require liver biopsy before MTX is initiated. In addition, biopsies are recommended only in those patients who continue to have enzyme abnormalities, and for whom continuation of MTX therapy is contemplated.

Monitoring for toxicity should be done every 2 to 12 weeks and is based on the duration of therapy, with more frequent monitoring provided earlier in the course of treatment. Systems review and physical examination should include monitoring for symptoms or signs of myelosuppression (fever, infection, bruising, and bleeding), pulmonary toxicity (shortness of breath, cough, rales), gastrointestinal intolerance (nausea, vomiting, diarrhea), and lymphadenopathy. Laboratory parameters that should be followed include a CBC with platelets, liver transaminases, and creatinine (Table 61-2).

It is important to consider vaccination status in any patient who is going to use MTX. RA patients have an increased incidence of death from pneumonia, and MTX may reduce the immune response to pneumococcal infections.

### Table 61-2 Safety Monitoring

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>&lt;3 Months of Drug Therapy*</th>
<th>Monitoring Interval 3-6 Months of Drug Therapy*</th>
<th>&gt;6 Months of Drug Therapy*</th>
<th>Contraindications</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTX</td>
<td>CBC, LFT, Cr, HBV, HCV; vaccinate: influenza; <em>Pneumococcus, HBV</em></td>
<td>Every 2-4 wk</td>
<td>Every 8-12 wk</td>
<td>Every 8-12 weeks</td>
<td>Active infection, symptomatic pulmonary disease, WBC &lt;3000/mm³, Pt &lt;50,000/mL, CrCl &lt;30 mL/min, history of myelodysplasia or recent lymphoproliferative disorder, LFT &gt;2 ULN, acute or chronic HBV or HCV, pregnancy, lactation</td>
</tr>
<tr>
<td>Leflunomide</td>
<td>CBC, LFT, Cr, HBV, HCV; vaccinate: influenza; <em>Pneumococcus, HBV</em></td>
<td>Every 2-4 wk</td>
<td>Every 8-12 wk</td>
<td>Every 8-12 wk</td>
<td>Active infection WBC &lt;3000/mm³, Pt &lt;50,000/mL, history of myelodysplasia or recent lymphoproliferative disorder, LFT &gt;2 ULN, acute or chronic HBV or HCV, pregnancy, lactation</td>
</tr>
<tr>
<td>Sulfasalazine</td>
<td>CBC, LFT, Cr; vaccinate: influenza; <em>Pneumococcus</em></td>
<td>Every 2-4 wk</td>
<td>Every 8-12 wk</td>
<td>Every 8-12 wk</td>
<td>Sulfur allergy, Pt &lt;50,000/mL, LFT &gt;2 ULN, acute HBV/HCV, some classes of chronic HBV/HCV</td>
</tr>
<tr>
<td>Hydroxychloroquine</td>
<td>CBC, LFT, Cr; complete ophthalmologic examination within 1 year</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>History of vision changes attributed to 4-aminoquinolone derivatives, some classes of untreated HBV/HCV</td>
</tr>
</tbody>
</table>

*Monitoring at <3, 3-6, and >6 mo need only include CBC, LFT, and Cr.

CBC, complete blood count; Cr, creatinine; CrCl, creatinine clearance; HBV, hepatitis B; HCV, hepatitis C; LFT, liver function test; Pt, platelets; ULN, upper limit of normal; WBC, white blood cell count.

antigen. Thus, any patient in whom MTX is going to be used should first receive the pneumococcal vaccination, with booster as appropriate. Vaccinations for hepatitis B virus for at-risk patients and yearly influenza vaccines are recommended as well. Caution should be exercised when administering live virus vaccinations to patients on MTX.

**Drug Interactions and Contraindications**

**Drug Interactions**

Drugs that are known hepatotoxins, such as sulfasalazine, leflunomide, and azathioprine, may potentiate liver toxicity when used in combination. Organic acids such as sulfonamides, salicylates, NSAIDs, penicillin G, piperacillin, and probenecid competitively inhibit tubular secretion, and this delays MTX clearance. MTX also undergoes distal tubular reabsorption, which may be enhanced by the addition of HCQ and blocked by the addition of folic acid. Drugs that affect renal function should be used with caution because of the renal clearance of MTX and, therefore, the increased risk of MTX toxicity that could occur because of decreased clearance.

Several of the aforementioned drugs deserve special mention. Trimethoprim-sulfamethoxazole should be avoided or used with extreme caution because of possible hematologic toxicity with MTX. Mechanisms for this toxicity include an additive antifolate effect from trimethoprim, decreased MTX clearance due to inhibition of tubular secretion by sulfamethoxazole, and altered MTX plasma protein binding. NSAIDs are commonly used in RA patients as adjunctive therapy. NSAIDs may increase MTX levels by displacing MTX from plasma proteins and limiting tubular secretion. Despite lack of a significant pharmacokinetic or clinical interaction between low-dose MTX and a variety of NSAIDs, vigilance for MTX toxicity should increase whenever NSAID dosages are changed in patients on stable weekly doses of MTX. Low doses of aspirin used for cardiovascular prophylaxis are not likely to be of concern. Furthermore, probenecid should be avoided because it inhibits tubular secretion of MTX.

**Contraindications**

MTX should not be used in severe renal, pulmonary, or hepatic impairment, pre-existing bone marrow suppression, alcoholic liver disease, and pregnancy or breastfeeding. Ongoing or active infection is also a contraindication. In most cases, patients who desire to continue drinking alcohol should not be treated with MTX. Mild to moderate renal insufficiency is a relative contraindication, and use of MTX in these patients may require more vigilant toxicity monitoring (Table 61-3).

**LEFLUNOMIDE**

<table>
<thead>
<tr>
<th>Key Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leflunomide reversibly inhibits dihydroorotate dehydrogenase (DHODH).</td>
</tr>
<tr>
<td>Loading doses often are not used in clinical practice because of gastrointestinal toxicity.</td>
</tr>
<tr>
<td>Because of enterohepatic recirculation, leflunomide has a very long half-life.</td>
</tr>
<tr>
<td>Leflunomide is absolutely contraindicated in pregnancy, and levels must be checked with a washout protocol if needed before conception.</td>
</tr>
<tr>
<td>Vigilance must be used for hepatotoxicity.</td>
</tr>
</tbody>
</table>

Leflunomide, an isoxazole derivative, is a synthetic DMARD approved for the treatment of RA. It emerged from a specific anti-inflammatory drug development program and has potent immunomodulatory effects.
malononitriloamide A77 1726. A77 1726 is also known as teriflunomide (Figure 61-4).

**Actions of Leflunomide**

As with MTX, the precise mechanism of action responsible for the effects of leflunomide in rheumatic disease is not completely understood. Leflunomide is immunomodulatory, with the net effect being a reduction in activated T lymphocytes. Its two in vitro mechanisms of action vary depending on concentration: (1) At the concentration of the active metabolite (A77 1726) achieved in patients, its major effect appears to be reversible inhibition of the enzyme dihydroorotate dehydrogenase (DHODH), which results in inhibition of pyrimidine synthesis; (2) at higher concentrations, A77 1726 also inhibits tyrosine kinases, interfering with cell signal transduction.

Activation of T cells results in progression from the resting phase (G₀) to the G₁ phase, where ribonucleotides are synthesized, and then to the S phase, where cellular DNA is replicated in preparation for mitosis. T cell activation requires significant increases in de novo pyrimidine and purine biosynthesis. Sensors such as proto-oncogenes (p53) and checkpoints (cyclins C and D) in this pathway monitor the level of nucleotide pools and prevent damaged cells from replicating.

Uridine monophosphate (rUMP) is a precursor for the formation of pyrimidine nucleotides and thus is essential for both RNA and DNA synthesis. The steps in de novo rUMP synthesis are seen in Figure 61-5. A critical step in this pathway is the generation of dihydroorotate in the

![Figure 61-5](image-url) The active metabolite of leflunomide, A77 1726, blocks the conversion of dihydroorotate to orotate within the mitochondria. ATP, adenosine triphosphate; DNA, deoxyribonucleic acid; RNA, ribonucleic acid; rUMP, uridine monophosphate.

![Figure 61-4](image-url) Leflunomide is rapidly and completely metabolized to its active metabolite, A77 1726.

**Chemical Structure**

Leflunomide is a low-molecular-weight isoxazole compound and is chemically unrelated to any previous immunosuppressant. Leflunomide is a pro-drug and is rapidly and completely converted to its active metabolite, the
cytoplasm with subsequent diffusion into the mitochondria, where the enzyme dihydroorotate dehydrogenase (DHODH) is located. DHODH converts dihydroorotate to orotate, and the latter diffuses back into the cytoplasm and is subsequently converted to rUMP and ultimately to RNA and DNA.

The first postulated mechanism of action of leflunomide consists of inhibition of DHODH by A77 1726, which lowers orotate levels and leads to a decrease in rUMP and subsequent nucleotide synthesis, resulting in T cell cycle arrest. This mechanism of action has been substantiated by experimental evidence. In vitro mitogen-stimulated activation of T cells is blocked by levels of A77 1726 that inhibit DHODH, and this inhibition can be reversed by the addition of uridine, suggesting that A77 1726 works by disruption of pyrimidine biosynthesis. Further, the only enzyme inhibited by A77 1726 in this pathway, at concentration obtained in vivo, is DHODH.

Evidence also exists to support that inhibition of DHODH produces an arrest of lymphocytes in the G1 phase of the cell cycle. If the level of ribonucleotides, including rUMP, falls below a critical point, cytoplasmic p53 activation occurs, and p53 will translocate to the nucleus and initiate cellular arrest by ultimately preventing transcription of cyclins D and E. In cultures of human T cells, A77 1726 depletes rUMP pools and results in an accumulation of nuclear p53 with resultant cell cycle arrest. In comparison, treatment of cell lines lacking p53 with A77 1726 does not cause a G1 phase arrest.

Resting lymphocytes maintain ribonucleotide requirements largely through salvage pathways and are essentially unaffected by leflunomide. Active, or autoimmune, lymphocytes rely on the de novo pathway and are affected by leflunomide. Over the course of treatment with this slow-acting agent, autoimmune lymphocytes should be removed progressively.

At higher concentrations, A77 1726 inhibits phosphorylation of tyrosine kinases that are critical for cell growth and differentiation of activated cells. This inhibition has been proposed to partially or completely explain the anti-proliferative effects of leflunomide; however, it is unclear whether concentrations sufficient to achieve this effect are obtained in vivo.

Several other additional anti-inflammatory properties of leflunomide have been noted. Leflunomide has the ability to block the activation of nuclear factor κB (NFκB), which regulates the expression of genes important in inflammatory processes, including those seen in inflammatory arthritis. Ex vivo and in vitro studies in humans have shown that both leflunomide and MTX inhibit neutrophil chemotaxis, which may decrease the recruitment of inflammatory cells into the joints.

Leflunomide has also been shown to decrease the ratio of MMP-1 to TIMP-1. Finally, leflunomide alters the synthesis of cytokines by augmenting the immunosuppressive cytokine-transforming growth factor-β1 and suppressing the immunostimulatory cytokine IL-2.

**Pharmacology**

**Absorption and Bioavailability**

The gastrointestinal tract and the liver rapidly and completely convert ingested leflunomide into A77 1726. Food does not interfere with absorption. Circulating A77 1726 is highly bound (more than 99%) to plasma proteins, predominantly albumin. Its plasma concentration is linearly correlated with a single oral dose over a range of 5 to 25 mg; steady state is reached in 7 weeks after daily dosing.

**Distribution and Half-Life**

A77 1726 has a half-life of approximately 2 weeks (mean, 15.5 days), with a low apparent volume of distribution. A77 1726 undergoes enterohepatic recirculation. In healthy subjects, 90% of leflunomide is excreted by 28 days, but some may be present for much longer periods.

**Elimination**

In healthy subjects, the proportions excreted by the kidney and the gut are nearly equal. Because detectable A77 1726 may be present in the body months or years later, the ability to rapidly and effectively eliminate A77 1726 with cholestyramine is important. Oral administration of cholestyramine 8 g three times daily can lower the apparent half-life of A77 1726 to 1 to 2 days. Furthermore, activated charcoal 50 g every 6 hours can reduce plasma levels by 50% within 24 hours.

**Indications**

**Rheumatoid Arthritis.** Leflunomide was first shown to be safe and effective in treating RA in a placebo-controlled, dose-ranging, 6-month trial. Two pivotal trials, one in Europe and one in the United States, have compared leflunomide with sulfasalazine (SSZ) and MTX. The European trial had three arms: leflunomide (20 mg/day after a loading dose), sulfasalazine (escalated to 2 g/day), and placebo. In this trial, both leflunomide and sulfasalazine were superior to placebo in terms of swollen and tender joint counts, as well as physicians’ and patients’ overall assessments. It is important to note that both the leflunomide and sulfasalazine groups reported significant effects on slowing radiographic progression of disease compared with placebo. The U.S. trial compared patients treated with leflunomide (20 mg/day after a loading dose), MTX (7.5 to 15 mg/wk), or placebo. Again, both active drugs were found to be superior to placebo but not different from each other. Both leflunomide and MTX also slowed radiographic progression of disease compared with the placebo group. Another trial compared leflunomide (20 mg/day with loading) with MTX (10 to 15 mg/wk) in a 1-year trial with a 1-year extension. In this trial, MTX was shown to be statistically superior to leflunomide for the clinical outcomes measured, as well as the rate of radiographic progression after 2 years.

**Other Rheumatic Diseases.** Leflunomide has been reported to be efficacious in systemic lupus erythematosus. In a randomized, controlled trial, leflunomide was more effective than placebo in improving markers of lupus disease activity, and was safe and well tolerated. A subsequent small, prospective, open-label trial of patients with lupus nephritis unresponsive to conventional therapy showed that leflunomide was efficacious and well tolerated.
Leflunomide has been shown to be effective in treating psoriatic arthritis and psoriasis when compared with placebo. In an open-label trial of ankylosing spondylitis, leflunomide was shown to be effective in treating peripheral arthritis, but axial symptoms did not improve.

Both an open-label trial and a randomized, controlled clinical trial have shown the efficacy of leflunomide in maintaining remission in granulomatosis with polyangiitis after successful induction with cyclophosphamide. In the latter trial, leflunomide was superior to methotrexate in preventing relapse.

Leflunomide has also been shown to be safe and effective in patients with juvenile idiopathic arthritis (JIA) who did not respond to or could not tolerate MTX.

### Dose and Drug Administration

Leflunomide is available in oral tablets at doses of 10, 20, and 100 mg. Oral leflunomide is rapidly metabolized to A77 1726, which has a very long half-life; therefore, the standard recommendation is to start therapy with a loading dose of 100 mg daily for 3 days, then switch to the standard maintenance dose of 20 mg daily. Despite this recommendation, many clinicians no longer prescribe a loading dose because it is believed to increase the drug’s gastrointestinal toxicity. Also, it is common practice to decrease the dose to 10 mg daily if toxicity occurs, or if complete control of the disease can be maintained. Because of its long half-life, some clinicians give leflunomide less often (three to five times per week).

### Geriatric Patients (See Table 61-1)

No pharmacokinetic studies specific to geriatric patients have been done for leflunomide; hence, dosing recommendations are the same as for the general population. No clinical experience in patients with renal insufficiency has been reported, so those patients should be monitored carefully.

### Pediatric Patients (See Table 61-1)

Although not approved in the United States to treat patients with JIA, leflunomide is used off-label for this condition at a dose of 10 to 20 mg/day. This dosing is often based on a patient’s body weight. A recently published example of dosing by body weight is as follows: A patient weighing less than 20 kg receives 10 mg every other day, a patient weighing more than 20 kg but less than 40 kg receives 10 mg/day, and a patient weighing more than 40 kg receives 20 mg/day.

### Toxicity

For major controlled trials that have used leflunomide at a dose of 20 mg/day, the incidence of adverse events that resulted in trial withdrawal are shown in Table 61-4. Leflunomide-associated withdrawals (19%) were more frequent than those associated with MTX (14%), were similar in frequency to those associated with sulfasalazine (19%), but were more frequent than those associated with placebo treatment (8%).

### Gastrointestinal and Hepatic Side Effects

The most common side effect that limits the use of leflunomide is diarrhea, which responds to dose reduction and may be less common if the loading dose is not used. Abdominal pain, dyspepsia, and nausea from leflunomide appear to be slightly increased over placebo rates.

Liver toxicity can occur in association with leflunomide administration. Data from a large U.S. cohort of RA and PsA patients show that elevations in ALT/AST levels greater than 1 times the upper limit of normal (1 × ULN) occurred in 17% and elevations greater than 2 × ULN occurred in 1% to 2% of patients taking leflunomide. Leflunomide given in combination with MTX resulted in ALT/AST elevations greater than 1 × ULN in 31% and greater than 2 × ULN in 5% of patients. Furthermore, this change in transaminases was more commonly seen in PsA patients.

The European Agency for Evaluation of Medicinal Products (EMEA) reported 296 patients with hepatic abnormalities and 15 patients with liver failure and death while taking leflunomide. The U.S. Food and Drug Administration (FDA) reviewed adverse event reports between August 2002, and May 2009, and found 49 cases of severe liver injury, 14 of which resulted in death. Most patients with hepatotoxicity have risk factors, including concomitant administration of another hepatotoxic agent or underlying liver disease.

### Cardiovascular Side Effects

Hypertension has consistently been reported to occur more frequently in leflunomide-treated compared with placebo-treated patients. Additionally, elevation of cholesterol levels has been reported in association with leflunomide use. Both of these effects should be monitored because of the excess cardiovascular mortality reported in patients with RA.

### Miscellaneous

**Dermatologic.** Skin rashes have been reported, most commonly occurring between the second and fifth months and necessitating discontinuation of the drug. Severe skin reactions such as Stevens-Johnson syndrome or toxic epidermal necrolysis require leflunomide washout with cholestyramine.

An increased incidence of alopecia has been reported in clinical trials in association with leflunomide.

**Pulmonary.** Rare postmarketing reports have described interstitial lung disease, including interstitial pneumonitis and pulmonary fibrosis.

**Hematologic.** Rare cases of pancytopenia have been reported in postmarketing surveillance, primarily in patients...
with known risk factors for blood dyscrasias. An increased risk of lymphoproliferative disorders has not been associated with leflunomide.

**Weight Loss.** Significant weight loss has also been reported to occur in patients taking leflunomide.\(^{162}\)

**Fertility, Pregnancy, and Lactation**

Leflunomide is rated Pregnancy Category X. Animal studies have demonstrated substantial teratogenic and embryolethal effects with small doses of leflunomide. Therefore, women of childbearing potential should be strongly counseled about this, and leflunomide should not be prescribed for women who are not practicing reliable birth control methods. A pregnancy test should be considered before therapy is initiated. Leflunomide excretion in milk is unknown; therefore, nursing mothers should not receive leflunomide.

It is critically important to note that the active metabolite of leflunomide (A77 1726), largely because of its enterohepatic circulation, may remain in the body for years. Therefore, if a woman who has previously received leflunomide wishes to become pregnant, A77 1726 levels should be measured. Active elimination of leflunomide from the body should be considered for levels above 0.02 mg/L. This can be achieved by the oral administration of cholestyramine for 11 days (8g three times daily).\(^{163}\) Before pregnancy is attempted, verification of levels below 0.02 mg/L should be confirmed on two separate occasions, at least 14 days apart, and women should then wait an additional three full menstrual cycles.\(^{164}\) Patients may require more than one course of cholestyramine to achieve this level. Although no data exist, men wishing to father children should undergo the same washout procedure as women and should wait an additional 3 months after the second drug plasma level is verified below 0.02 mg/L (see Table 61-1).

**Toxicity Monitoring**

Similar to MTX, the ACR has published guidelines on the initiation and monitoring of leflunomide.\(^{124}\) Patients taking leflunomide should have a baseline CBC and liver enzyme monitoring, including AST, ALT, and albumin. Serum creatinine is important because leflunomide is partially eliminated by the kidney. The frequency of monitoring depends on the duration of therapy (see Table 61-2). More frequent monitoring may be warranted if concomitant immunosuppressive agents, such as MTX, are given. If patients experience significant toxicity, a washout procedure is indicated to more rapidly eliminate the drug. Caution should be exercised when live virus vaccinations are administered to patients on leflunomide.

**Drug Interactions and Contraindications**

**Drug Interactions**

Cholestyramine interferes with enterohepatic recycling of leflunomide, resulting in lower serum concentrations. Concomitant use with hepatotoxic agents, including MTX, increases the risk of liver toxicity, and leflunomide must be used with caution and monitored judiciously. Rifampin may increase the serum concentration of A77 1726. Leflunomide may potentiate warfarin therapy.

**Contraindications**

Leflunomide should not be used in patients with impaired liver function, severe renal impairment, bone marrow dysplasia, severe immunodeficiency, severe hypoproteinemia, or known hypersensitivity to the drug. The liver is involved in enterohepatic recirculation and biliary excretion, thus leflunomide use in liver disease is contraindicated. In renal insufficiency, the levels of circulating A77 1726 do not appear to be increased, but the component of free A77 1726 is increased. Leflunomide is contraindicated in the setting of serious infection and should be discontinued in patients with new or worsening pulmonary symptoms or rash. Leflunomide is absolutely contraindicated in pregnancy and breastfeeding (see Table 61-1).

**SULFASALAZINE**

**KEY POINTS**

**Sulfasalazine (SSZ) has antimicrobial and anti-inflammatory properties, but the exact mechanism of action is unknown.**

**SSZ is commonly used as part of combination therapy for RA.**

**Gastrointestinal intolerance is a common side effect.**

**Although rare, surveillance for leukopenia is important early in the course of the treatment.**

Sulfasalazine (SSZ) was the first agent to be synthesized specifically for rheumatoid arthritis in 1938, by Professor Nanna Svartz of Stockholm, in collaboration with the Swedish pharmaceutical company Pharmacia. The prevailing notion at that time was that RA was caused by infection, and SSZ was designed with both anti-inflammatory and antibacterial properties.

**Chemical Structure**

Salicylazosulfapyridine (SASP), now known as sulfasalazine, is a conjugate of the anti-inflammatory 5-aminosalicylic acid (5-ASA or mesalamine) and the antibacterial sulfapyridine joined by an azo bond (Figure 61-6). The abbreviation SASP is still in use as an alternative to SSZ.

**Actions of Sulfasalazine**

Despite more than seven decades of use, the mechanism of action of SSZ in rheumatic disease still is not fully elucidated. When SSZ was designed, its antimicrobial properties were thought to be fundamentally important in successful treatment of RA, which was postulated to be an entero-pathic arthropathy. Through alteration of gut flora, SSZ may downregulate the immune response leading to inflammatory arthritis. Although this hypothesis has never been disproved, this mechanism of action has fallen out of favor for several reasons: To date, no conclusive evidence has been found for an infectious cause of RA, other
The synovitis of RA.

proliferation and angiogenesis, which likely contributes to the location to the nucleus.

scription of key cytokines, adhesion molecules, and chemokines IL-1, TNF, and IL-6.

5-ASA are less active than SSZ. Cytokine profiles are also altered by SSZ, resulting in inhibition of the T cell cyto

production. With resultant declines in immunoglobulin synthesis and RF production.

erythrocyte sedimentation rate (ESR) and in some clinical parameters, including grip strength and joint circumference. Taken together, it would appear that both SSZ and sulfapyridine have a role in the therapeutic efficacy for RA.

One potential site of action for SSZ that may explain its systemic action despite low serum levels is the mucosa-associated lymphoid tissue (MALT) in the small bowel. In the gut lumen, the therapeutic concentration of SSZ is at least two times greater than in the serum, and drug concentration in the surrounding mucosal tissue is likely also high. The gut immune system is extensive, and active communication with the rest of the body occurs via migration and recirculation of activated lymphocytes. Furthermore, a link between MALT and the joints has been suggested. Evidence that some of the efficacy of SSZ may be mediated via MALT is described as follows: Treatment with SSZ has been shown to decrease circulating immunoglobulin (Ig) A-producing cells and serum levels of IgA, correlating with disease improvement. SSZ has been shown to reduce gut mucosa lymphocytes in treated patients, and SSZ has been shown to modulate an immune response elicited by an oral antigen in mice and healthy volunteers.

Pharmacology

Absorption and Bioavailability

Less than 30% of SSZ is absorbed by the small bowel; most undergoes enterohepatic circulation and is secreted unchanged in the bile, with a resulting bioavailability of 10%. The steady-state serum concentration of SSZ is 5 μg/mL after an oral dose of 2 g/day. Most SSZ reaches the colon, where intestinal bacteria reduce the azo bond and release the two active components: sulfapyridine and 5-ASA. However, the SSZ levels needed in vitro to see the anti-inflammatory and immunomodulatory effects are far greater than those obtained in vivo. There has not been a demonstrable relationship shown between plasma SSZ, sulfapyridine, and 5-ASA levels that correlate with clinical efficacy, so serum levels may be irrelevant. In an open, nonrandomized study comparing two groups of patients who received either sulfapyridine or 5-ASA in doses that would represent the molar equivalent of 2 g of SSZ, the group receiving sulfapyridine showed significant improvement in erythrocyte sedimentation rate (ESR) and in some clinical parameters, including grip strength and joint circumference. Taken together, it would appear that both SSZ and sulfapyridine have a role in the therapeutic efficacy for RA.

One potential site of action for SSZ that may explain its systemic action despite low serum levels is the mucosa-associated lymphoid tissue (MALT) in the small bowel. SSZ also has been demonstrated in vitro to possess multiple anti-inflammatory properties. First, SSZ weakly inhibits the proinflammatory effects of the arachidonic acid cascade, with a slight inhibitory effect on prostaglandin E2 synthetase activity, as well as lipoxygenase products. SSZ also downregulates neutrophil chemotaxis, migration, and proteolytic enzyme production and degradation and has been shown to inhibit neutrophil activation by decreasing the flux of second messengers involved in intracellular signal transduction. Neutrophil migration to sites of inflammation can be downregulated by adenosine. It has been shown that SSZ inhibits folate-dependent enzymes, including ATIC and DHFR, resulting in increased adenosine release into the extracellular milieu. Gadangi and colleagues confirmed that SSZ, like MTX, modulates inflammation via increased adenosine release. This effect appears to be due solely to SSZ as sulfapyridine and 5-ASA are inactive. SSZ is a more potent inhibitor of ATIC than is MTX.

SSZ also has been shown to possess multiple immunomodulatory properties. In vitro, SSZ inhibits T cell proliferation, natural killer cell activity, and B cell activation, with resultant declines in immunoglobulin synthesis and RF production. In all of these systems, sulfapyridine and 5-ASA are less active than SSZ. Cytokine profiles are also altered by SSZ, resulting in inhibition of the T cell cytokines IL-2 and γ interferon, and the monocytic/macrophage cytokines IL-1, TNF, and IL-6. NFκB is a key transcription factor that when active mediates the transcription of key cytokines, adhesion molecules, and chemokines essential to mounting an immune response. In vitro, SSZ, but not sulfapyridine or 5-ASA, inhibits NFκB translocation to the nucleus.

SSZ more so than sulfapyridine inhibits endothelial cell proliferation and angiogenesis, which likely contributes to the synovitis of RA. The synovial fibroblast also plays a key role in the pathogenesis of RA, and SSZ inhibits fibroblast proliferation and metalloproteinase synthesis. Finally, SSZ has been shown to inhibit the formation of osteoclasts and may be antiresorptive in RA.

The active moiety of SSZ is controversial. The parent compound, SSZ, appears to have the most biologic activity when compared with sulfapyridine and 5-ASA. However, the SSZ levels needed in vitro to see the anti-inflammatory and immunomodulatory effects are far greater than those obtained in vivo. There has not been a demonstrable relationship shown between plasma SSZ, sulfapyridine, and 5-ASA levels that correlate with clinical efficacy, so serum levels may be irrelevant. In an open, nonrandomized study comparing two groups of patients who received either sulfapyridine or 5-ASA in doses that would represent the molar equivalent of 2 g of SSZ, the group receiving sulfapyridine showed significant improvement in erythrocyte sedimentation rate (ESR) and in some clinical parameters, including grip strength and joint circumference. Taken together, it would appear that both SSZ and sulfapyridine have a role in the therapeutic efficacy for RA.

One potential site of action for SSZ that may explain its systemic action despite low serum levels is the mucosa-associated lymphoid tissue (MALT) in the small bowel. SSZ also has been demonstrated in vitro to possess multiple anti-inflammatory properties. First, SSZ weakly inhibits the proinflammatory effects of the arachidonic acid cascade, with a slight inhibitory effect on prostaglandin E2 synthetase activity, as well as lipoxygenase products. SSZ also downregulates neutrophil chemotaxis, migration, and proteolytic enzyme production and degradation and has been shown to inhibit neutrophil activation by decreasing the flux of second messengers involved in intracellular signal transduction. Neutrophil migration to sites of inflammation can be downregulated by adenosine. It has been shown that SSZ inhibits folate-dependent enzymes, including ATIC and DHFR, resulting in increased adenosine release into the extracellular milieu. Gadangi and colleagues confirmed that SSZ, like MTX, modulates inflammation via increased adenosine release. This effect appears to be due solely to SSZ as sulfapyridine and 5-ASA are inactive. SSZ is a more potent inhibitor of ATIC than is MTX.

SSZ also has been shown to possess multiple immunomodulatory properties. In vitro, SSZ inhibits T cell proliferation, natural killer cell activity, and B cell activation, with resultant declines in immunoglobulin synthesis and RF production. In all of these systems, sulfapyridine and 5-ASA are less active than SSZ. Cytokine profiles are also altered by SSZ, resulting in inhibition of the T cell cytokines IL-2 and γ interferon, and the monocytic/macrophage cytokines IL-1, TNF, and IL-6. NFκB is a key transcription factor that when active mediates the transcription of key cytokines, adhesion molecules, and chemokines essential to mounting an immune response. In vitro, SSZ, but not sulfapyridine or 5-ASA, inhibits NFκB translocation to the nucleus.

SSZ more so than sulfapyridine inhibits endothelial cell proliferation and angiogenesis, which likely contributes to the synovitis of RA. The synovial fibroblast also plays a key role in the pathogenesis of RA, and SSZ inhibits fibroblast proliferation and metalloproteinase synthesis. Finally, SSZ has been shown to inhibit the formation of osteoclasts and may be antiresorptive in RA.

The active moiety of SSZ is controversial. The parent compound, SSZ, appears to have the most biologic activity when compared with sulfapyridine and 5-ASA. However, the SSZ levels needed in vitro to see the anti-inflammatory and immunomodulatory effects are far greater than those obtained in vivo. There has not been a demonstrable relationship shown between plasma SSZ, sulfapyridine, and 5-ASA levels that correlate with clinical efficacy, so serum levels may be irrelevant. In an open, nonrandomized study comparing two groups of patients who received either sulfapyridine or 5-ASA in doses that would represent the molar equivalent of 2 g of SSZ, the group receiving sulfapyridine showed significant improvement in erythrocyte sedimentation rate (ESR) and in some clinical parameters, including grip strength and joint circumference. Taken together, it would appear that both SSZ and sulfapyridine have a role in the therapeutic efficacy for RA.

One potential site of action for SSZ that may explain its systemic action despite low serum levels is the mucosa-associated lymphoid tissue (MALT) in the small bowel. In the gut lumen, the therapeutic concentration of SSZ is at least two times greater than in the serum, and drug concentration in the surrounding mucosal tissue is likely also high. The gut immune system is extensive, and active communication with the rest of the body occurs via migration and recirculation of activated lymphocytes. Furthermore, a link between MALT and the joints has been suggested. Evidence that some of the efficacy of SSZ may be mediated via MALT is described as follows: Treatment with SSZ has been shown to decrease circulating immunoglobulin (Ig) A-producing cells and serum levels of IgA, correlating with disease improvement; SSZ has been shown to reduce gut mucosa lymphocytes in treated patients; and SSZ has been shown to modulate an immune response elicited by an oral antigen in mice and healthy volunteers.

Pharmacology

Absorption and Bioavailability

Less than 30% of SSZ is absorbed by the small bowel; most undergoes enterohepatic circulation and is secreted unchanged in the bile, with a resulting bioavailability of 10%. The steady-state serum concentration of SSZ is 5 μg/mL after an oral dose of 2 g/day. Most SSZ reaches the colon, where intestinal bacteria reduce the azo bond and release the two active components: sulfapyridine and 5-ASA. Most of the sulfapyridine is absorbed from the colon (>90%) and appears in plasma 4 to 6 hours after an oral dose (steady-state serum concentration of 30 μg/mL after 2 g/day dose). Most of the 5-ASA (80% to 90%) remains in the bowel. Sulfapyridine and SSZ are likely the active components in rheumatic disease, and 5-ASA is the active component in ulcerative colitis.

The bioavailability of standard SSZ is similar to that of enteric-coated SSZ. Administration of SSZ with food decreases the blood concentration of both SSZ and sulfapyridine.
Distribution and Half-Life

Both SSZ and sulfapyridine are widely distributed in the body with more than 99% and 50% to 70% plasma protein binding, respectively. Serum and synovial concentrations are comparable. The half-life of SSZ is 6 to 17 hours with upper range values reflecting older patients, and the half-life of sulfapyridine is 8 to 21 hours with upper range values reflecting slow acetylators.

Elimination

Sulfapyridine is extensively metabolized in the liver by N-acetylation and ring hydroxylation with subsequent glucuronidations and because of genetic variations in acetylator phenotype, wide variability is seen among individuals. Slow acetylators have a reduced clearance rate and higher serum sulfapyridine concentrations. Sulfapyridine is excreted in the urine, and 5-ASA is eliminated primarily in the feces. The small portion of 5-ASA that is absorbed is excreted in the urine as N-acetylmesalamine.

Indications

Rheumatoid Arthritis. Multiple published trials since the early 1980s have showed significant benefit in both clinical and laboratory parameters for SSZ (2 to 3 g daily) versus placebo in the treatment of RA. A meta-analysis of randomized controlled clinical trials of SSZ for the treatment of RA was published in 1999. In this analysis, SSZ was compared with placebo and other single DMARD therapies, including HCQ, d-penicillamine (d-Pen), and gold sodium thiomalate or aurothioglucone. In trials looking at SSZ versus placebo, SSZ was shown to be superior to placebo in terms of improvement in multiple clinical parameters. The withdrawal rate due to lack of efficacy was significantly greater in the placebo group than in the SSZ group; however, more patients in the SSZ group than in the placebo group withdrew because of adverse effects (P < .0001). No single DMARD emerged as clinically superior in this meta-analysis. Subsequent studies have demonstrated equivalent clinical efficacy with SSZ and MTX, as well as with SSZ and leflunomide, although a 2-year extension trial did show that beneficial effects are sustained to a greater extent with leflunomide than with SSZ.

Spondyloarthropathies

Psoriatic Arthritis (PsA). A recently published systematic review of therapies for psoriatic arthritis looked at six randomized controlled trials comparing SSZ with placebo in PsA. Results show that SSZ is efficacious in treating the peripheral arthritis of PsA, but does not appear to influence the axial manifestations. Few data on the prevention of radiographic progression are available.

Ankylosing Spondylitis (AS). A recent meta-analysis reviewed 11 trials that included 895 patients with AS treated with SSZ or placebo. In all patients with AS, SSZ showed some benefit in reducing ESR and easing spinal stiffness. The authors concluded that SSZ may benefit patients at an early disease stage with a higher ESR and peripheral arthritis. This finding is in agreement with those of Clegg and co-workers, who found a significant benefit for SSZ in peripheral arthritis, but not in axial arthritis, in AS.

Reactive Arthritis (ReA). Most cases of ReA resolve spontaneously; others become chronic with peripheral or axial arthritis. In a randomized, controlled trial of 134 male veterans with ReA (predominantly peripheral arthritis) unresponsive to NSAIDs, SSZ was more effective than placebo.

Inflammatory Bowel–Associated Arthritis. SSZ has been used effectively to treat ulcerative colitis and distal Crohn’s disease. No randomized controlled trials have investigated SSZ in terms of peripheral or axial arthritis manifestations of these diseases, although SSZ is used in clinical practice for peripheral arthritis.

Juvenile Inflammatory Arthritis. SSZ has been shown to be effective in polyarticular and pauciarticular juvenile inflammatory arthritis. A literature review by Brooks in 2001 found reports of 550 patients with juvenile inflammatory arthritis (half with polyarticular and one-third with pauciarticular disease) treated with SSZ. Results showed at least some drug-associated benefit in all subtypes, with the best response noted in late-onset pauciarticular disease and the least benefit observed in systemic-onset disease. Toxicity and intolerance were similar to those noted with adult use of SSZ, except for a substantial incidence of serum sickness in patients with systemic-onset disease.

Dosing

SSZ is available in regular and enteric-coated tablets of 500 mg and in a suspension of 50 mg/mL. To minimize side effects, most clinicians will prescribe 500 mg daily and will escalate the dose by 500 mg/day every week to the standard dose of 1500 to 3000 mg divided daily. Dose reduction may ameliorate side effects.

Geriatric Patients (See Table 61-1)

Dosing recommendations for SSZ in geriatric patients are the same as for the general adult population. Studies of pharmacokinetics in the elderly have shown that although the elimination half-life is longer, it is primarily dependent on acetylator phenotype; however, the dose should be reduced for renal insufficiency.

Pediatric Patients (See Table 61-1)

In pediatric patients, SSZ is dosed initially at 10 to 12.5 mg/kg/day, with a weekly dose increase of 50 mg/kg/day in two divided doses, until a maintenance dose of 2 g/day is achieved. This can be escalated to 3 g/day if no response is seen with lower doses.

Toxicity

In general, a majority of adverse effects from SSZ occur within the first several months of treatment and decrease with continued use. The most common early adverse effects include gastrointestinal (GI) effects, headache, dizziness, and rash. They tend to decrease with continued use.
Gastrointestinal and Hepatic

Nausea and upper abdominal discomfort are the most common adverse effects with SSZ. Nausea frequently occurs with central nervous system (CNS) effects, including dizziness and headache. In one large cohort of 1382 RA patients, GI/CNS effects were reported as transient in 8% of patients; SSZ was continued and in another 18% led to discontinuation of therapy. Nausea is more common in patients who achieve higher sulfapyridine levels and in slow acetylators. Diarrhea can occur, usually in the first several months. GI effects may be decreased by administration of enteric-coated preparations. Elevations in liver transaminases may occur and are usually transient. However, they may be accompanied by fever, rash, hepatomegaly, and possibly eosinophilia.

Hematologic

Hematologic disturbances are rare, occurring in less than 3% of patients; they usually occur within the first 3 months. The most common abnormalities include leukopenia that usually reverses upon cessation of the drug, although some cases of fatal agranulocytosis have been reported, warranting continued surveillance. Macrocytosis and hemolysis have been reported with SSZ; its use should be avoided in patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency. Folic acid supplementation may be reasonable because of SSZ effects on folate metabolism. Thrombocytopenia is rare.

Dermatologic

Rashes occur in less than 5% of patients, usually in the first 3 months of therapy. Rashes are usually maculopapular, pruritic, and generalized, although some patients develop urticaria. Desensitization to SSZ has been reported. Anecdotally, erythema multiforme, toxic epidermal necrolysis, and Stevens-Johnson syndrome have been rarely reported, but were not seen in the major clinical trials. Photosensitivity has also been reported. Patients who develop a rash while on SSZ should be cautioned to avoid other sulfonamide-containing agents, such as thiazide diuretics, cefaloxin, and antibiotics.

Pulmonary

Pulmonary toxicity from SSZ is rare and manifests as reversible infiltrates with peripheral eosinophilia, cough, dyspnea, fever, and weight loss. Pathology reveals an eosinophilic pneumonia with interstitial infiltrates with or without fibrosis. Most cases resolve with discontinuation of SSZ with or without corticosteroids.

Miscellaneous

Minor reactions, such as irritability, anxiety, headache, and difficulty sleeping, may occur. Rare cases of drug-induced lupus, hypogammaglobulinemia, and aseptic meningitis have been reported. Advise patients that they may develop orange discoloration of their urine, sweat, and tears.

Fertility, Pregnancy, and Lactation

No reports have described diminished fertility in women taking SSZ; however, men can have oligospermia, impaired sperm motility, and abnormal sperm morphology that returns to normal 2 to 3 months after cessation of the drug. SSZ is considered FDA Category B, C for pregnancy. SSZ and sulfapyridine cross the placenta, and fetal concentrations are equivalent to maternal concentrations; however, SSZ does not seem to cause or increase fetal abnormalities or spontaneous abortions, and may be one of the DMARDs of first choice for treating rheumatic disease in women of childbearing age who are or wish to become pregnant. Sulfapyridine is excreted into breast milk, and one report described a child who developed bloody diarrhea, which led the American Academy of Pediatrics to classify SSZ as a drug that must be given with caution to nursing women (see Table 61-1).

Toxicity Monitoring

Most side effects from SSZ occur early in the course of treatment. The ACR guidelines recommend a baseline CBC with platelets, liver enzyme monitoring (including AST, ALT, and albumin), creatinine, and consideration for G6PD. The frequency of monitoring depends on the duration of therapy (see Table 61-2). Vaccination with Pneumococcus should be given if appropriate at the initiation of therapy, and yearly influenza vaccination is recommended (see Table 61-2).

Drug Interactions and Contraindications

Drug Interactions

Few SSZ-drug interactions are known. SSZ may impair absorption of digoxin and decrease its bioavailability. Rarely, SSZ can increase the effects of oral hypoglycemics and the anticoagulant effects of warfarin. Broad-spectrum antibiotics may alter gut flora and decrease the bioavailability of sulfapyridine and 5-ASA through reduced cleavage of the azo bond.

Contraindications

Patients with hypersensitivity to any component of SSZ, or with a sulfonamide or salicylate allergy, should not be prescribed SSZ. Caution should be used in patients with porphyria or gastrointestinal or genitourinary obstruction. SSZ should not be used in thrombocytopenia, severe liver disease, and active viral hepatitis (see Table 61-2).

ANTIMALARIALS

Table 61-1

| Hydroxychloroquine (HCQ) is a well-tolerated DMARD that is commonly used in combination therapy regimens for RA. |
| HCQ is more commonly used than chloroquine. |
| HCQ has a very long half-life, attributed to its affinity for melanin-containing cells in the skin. |
The aminoquinolones, including quinine, were first derived from the bark of the Peruvian cinchona tree, and were originally used to treat malaria. To reduce toxicity, the 4-aminoquinolines, chloroquine (CQ), and HCQ were developed. CQ and HCQ are the most common antimalarials prescribed, although quinacrine is used occasionally.

**Chemical Structure**

HCQ and CQ are very similar in their chemical structure, differing only by the substitution of a hydroxyethyl group for an ethyl group on the tertiary amino nitrogen of the side chain of CQ. Quinacrine includes the CQ structure, although it is not a 4-aminoquinolone derivative (Figure 61-7).

**Actions of Hydroxychloroquine**

Antimalarial agents have both immunomodulatory and anti-inflammatory properties, although their precise mechanism of action in rheumatic disease is unknown. Because HCQ and CQ are weak bases, they can pass through cytoplasmic membranes into cytoplasmic vesicles and accumulate, thereby increasing the vesicle pH from around 4.0 to 6.0 and interfering with acid-dependent subcellular functions. This increased pH has several postulated immunoregulatory effects, including stabilization of lysosomal membranes, attenuation of antigen processing and presentation, and inhibition of cell-mediated cytotoxicity. Macrophages and monocytes require precise pH concentrations for protein digestion and antigen processing, which is altered with an increased pH. Furthermore, receptor assembly is disrupted, including the major histocompatibility complex (MHC) class II molecules, because a higher pH in the endoplasmic reticulum stabilizes the MHC protein with invariant chains and prevents their displacement by low-affinity autoantigens. This, combined with decreased membrane receptor recycling, leads to downregulation of antigen presentation. Furthermore, antimalarial treatment has been shown to decrease circulating immune complexes.

Antimalarials also have inhibitory effects on proinflammatory cytokines. CQ inhibits IL-1 and IFN-γ production by monocytes and T cells. CQ also inhibits macrophage TNF mRNA transcription and endotoxin-induced secretion of TNF, IL-1, and IL-6. Studies of HCQ have yielded conflicting results in terms of the ability to inhibit TNF, but HCQ has been shown to block IL-1, IL-6, and IFN-γ production by monocytes.

Apoptosis or cell death plays an important role in regulation of the immune system, and defects in apoptosis may allow for longevity and persistence of autoreactive lymphocyte clones and perpetuation of autoimmunity. Both CQ and HCQ have been shown to upregulate apoptosis and may downregulate autoimmunity by elimination of autoreactive lymphocytes. Furthermore, antimalarials have been shown to inhibit the proliferative response of human lymphocytes and natural killer cell activity.

Anti-inflammatory properties of the antimalarials include effects on the arachidonic acid cascade caused by downregulation of phospholipases A2 and C, which contribute to the production of proinflammatory prostaglandins and lipid peroxidation. Lipid peroxidation is thought to play a role in apoptosis, particularly in response to ultraviolet (UV) A and UVB irradiation. Antimalarial agents also have antioxidant properties and may protect against tissue damage from free radicals.

Antimalarials have several other beneficial effects relevant to rheumatic disease that warrant further discussion. First, they are photoprotective, and this is likely the result of locally induced anti-inflammatory effects. Second, both HCQ and CQ inhibit platelet adhesion and aggregation, leading to an antithrombotic effect. Third, HCQ...
and CQ favorably alter the lipid profile, with reductions in total cholesterol, triglycerides, very low-density lipoproteins (VLDLs), and low-density lipoproteins (LDLs), particularly in patients on concomitant corticosteroid therapy.\(^{214,231}\) Finally, HCQ and CQ have been reported to decrease plasma glucose levels through inhibition of insulin degradation in the Golgi apparatus.\(^{214,232}\)

**Pharmacology**

**Absorption and Bioavailability**

HCQ and CQ are administered orally and are rapidly and completely absorbed, with the peak plasma concentration for both occurring within 8 hours.\(^{233}\) Considerable variability in blood concentrations has been reported among patients treated with the same dose, but higher plasma levels do not correlate with a better therapeutic response.\(^{214}\)

**Distribution and Half-Life**

Antimalarials accrue in different concentrations in various tissue compartments. Relatively small concentrations, similar to those seen in the plasma, are contained in fat, bone, tendon, and brain. Increased concentrations are seen in the kidney, bone marrow, spleen, lungs, adrenal glands, and liver. The highest concentration occurs in melanin-containing cells, such as those in the skin and the retina.\(^{214}\) In fact, the skin can serve as a long-term reservoir whereby the drug can exert its effect, or toxicity, even after it has been stopped.\(^{214}\) The half-life of HCQ and CQ is 40 to 50 days, and plasma levels will increase gradually and equilibrate after 3 to 4 months.\(^{233}\)

**Elimination**

Most of the absorbed drug is excreted in the urine unchanged, but some is metabolized to a desethyl derivative. The remainder is excreted in the feces.\(^{236}\)

**Indications**

**Rheumatoid Arthritis.** The efficacy of antimalarial agents in rheumatoid arthritis has been seen in their ability to control signs and symptoms of the disease;\(^{237-239}\) however, they have not been shown to retard bone erosions.\(^{240}\) In a meta-analysis of antimalarials, HCQ was found to be less toxic but also less effective than CQ.\(^{241}\) In comparison with other DMARDs, symptomatic efficacy is equal to or slightly less than that of other agents, and they have the slowest onset of action.\(^{241}\) Antimalarial agents are particularly suited for use in early, mild rheumatoid arthritis and in combination therapy.

**Systemic Lupus Erythematosus.** Although antimalarials are not appropriate as monotherapy for severe manifestations of SLE, they are used frequently to control constitutional symptoms, arthritis, fever, fatigue, and rash. The most convincing data supporting the efficacy of antimalarials in SLE come from studies in which the medication was discontinued in successfully treated patients. In one double-blind, placebo-controlled drug discontinuation study of 47 SLE patients in remission on HCQ, the risk of disease flare was increased by a factor of 2.5 in the placebo group.\(^{243}\)

Antimalarials are particularly suited to treat SLE because of their photoprotective effects, and they are especially useful in dermatologic manifestations of SLE.

**Discoid Lupus.** Antimalarials are effective in discoid lesions, with remission or major improvement reported in 60% to 90% of treated patients.\(^{223}\) When HCQ or CQ treatment alone is unsuccessful, the addition of quinacrine may be helpful, but long-term use may be limited by the development of yellowish skin pigmentation.\(^{244}\)

**Antiphospholipid Antibody Syndrome.** HCQ use has been associated with a decreased incidence of thrombosis in patients with antiphospholipid (aPL) antibodies.\(^{245}\) HCQ has also been shown to diminish thrombus size and time in mice injected with aPL antibodies, and to reverse aPL antibody-mediated platelet activation.\(^{246,247}\) Although clinical trials are needed to further establish efficacy, it is reasonable to consider the use of HCQ in patients with antiphospholipid antibody syndrome, especially if they are unable to tolerate high levels of anticoagulation, or if they develop thrombosis despite oral anticoagulation.\(^{248}\)

**Sjögren’s Syndrome.** HCQ has been shown to improve local eye and mouth symptoms, arthralgia, and myalgia in a prospective open-label study of patients with Sjögren’s syndrome.\(^{249}\) In addition to causing the immunomodulatory and anti-inflammatory effects discussed earlier, HCQ may inhibit glandular cholinesterase activity and enhance salivary gland secretion.\(^{250}\)

**Miscellaneous.** Antimalarials have been reported to be efficacious in small and uncontrolled trials for the treatment of palindromic rheumatism,\(^{251}\) childhood systemic lupus erythematosus,\(^{252}\) childhood dermatomyositis,\(^{253}\) eosinophilic fasciitis,\(^{254}\) and erosive osteoarthritis.\(^{255}\) In a controlled trial of 17 patients, the arthritis of calcium pyrophosphate crystal deposition disease improved more in the HCQ than in the placebo group.\(^{256}\)

**Dosing**

HCQ comes in 200-mg tablets, CQ is available in 250-mg and 500-mg tablets, and quinacrine is available from compounding pharmacies. To prevent ocular toxicity, HCQ should be maintained at a dose of 6.5 mg/kg or less for body weight or ideal body weight, whichever is less, and the dose of CQ should be 3 mg/kg or less.\(^{257}\) In practice, doses of HCQ rarely exceed 400 mg daily, in divided doses, and doses of CQ rarely exceed 250 to 500 mg, dosed daily.

**Geriatric Patients** (See Table 61-1)

No specific pharmacokinetic studies have explored the use of antimalarials in the elderly; thus dosing recommendations are the same as for the general population.\(^{204}\) Elderly patients should be screened at baseline for pre-existing ocular disease.

**Pediatric Patients** (See Table 61-1)

HCQ is used at a dose of 3 to 5 mg/kg/day, typically to a maximum of 400 mg/day. To achieve this dosing in younger children and use the 200-mg tablets, the drug can be dosed every other day. Alternatively, HCQ can be compounded in a 25 mg/mL suspension. It is generally recommended that
ophthalmologic screening occur every 6 months; however, some suggest that screening every 12 months is sufficient.

**Toxicity**

**Ophthalmologic**

Early eye symptoms can include defects in accommodation or conversion or blurred vision, which resolves. Retinal toxicity is the most feared side effect, although it is rarely seen if proper dosage and monitoring protocols are followed. Fewer than 20 cases have been reported in the literature in more than 1,000,000 patients prescribed the drug, and in a recent review of six cases of retinal toxicity, all arose as the result of dosing above published guidelines.\(^{257,258}\) CQ has a higher risk of retinal toxicity than HCQ.\(^{259}\) Risk factors for retinopathy include high dosage, longer duration of use (>5 years), renal or liver disease, and advancing age (>60 years old). HCQ or CQ retinopathy is described as bilateral bull’s eye maculopathy, with retinal pigment epithelial (RPE) cell depigmentation in the central macula and sparing of a small foveal island. Testing of the paracentral visual field can usually show toxicity before RPE changes are visible. When advanced, visual loss may be irreversible and may continue despite cessation of the drug because of its long half-life in the retina; so early detection is essential. CQ and HCQ can also cause corneal deposits, which can be associated with halos around lights and are benign and reversible.

**Dermatologic**

Rash is a common side effect leading to discontinuation of therapy. HCQ may also cause photosensitivity, alopecia, and depigmentation of hair.\(^{241}\)

**Neuromuscular**

Common neuromuscular symptoms include headache, insomnia, nightmares, and irritability, which are mild and reversible with lowering of the daily dose. Tinnitus and deafness can occur. Neuromyotoxicity has been reported and presents as proximal weakness of insidious onset with a normal creatine phosphokinase (CPK), which may be associated with peripheral neuropathy and cardiac myotoxicity. Muscle biopsy shows curvilinear bodies and muscle fiber atrophy with vacuolar changes.\(^{260}\)

**Cardiovascular**

Rarely, conduction disturbances and cardiomyopathy have been reported.\(^{261}\)

**Gastrointestinal**

Anorexia, nausea, vomiting, diarrhea, and abdominal cramping have been reported.\(^{241}\)

**Metabolic**

HCQ can reduce blood glucose levels and hemoglobin \(A_1C\), so diabetic patients need to be cautioned to watch their blood sugars closely with initiation of HCQ and may need adjustments in their diabetic medications.\(^{262}\)

**Fertility, Pregnancy, and Lactation**

No reports have described adverse effects on fertility. HCQ and CQ are considered FDA Pregnancy Category C.\(^{123}\) Because of the long half-life of HCQ, discontinuation of the drug at the time of pregnancy does not avoid fetal exposure. HCQ does cross the placenta, but there have been no reports of adverse outcomes or teratogenic effects in women who continue HCQ during pregnancy.\(^{123}\) CQ also crosses the placenta and binds more tightly in tissue than HCQ; fetal anomalies in women who took CQ during pregnancy have been reported.\(^{263}\) Quinacrine should not be used in pregnancy because it is mutagenic. Current recommendations are that HCQ may be continued throughout pregnancy, particularly in SLE, where discontinuation could precipitate a flare, which would be dangerous to both mother and baby.\(^{164}\) HCQ is also found in low concentration in breast milk, but the American Academy of Pediatrics classifies it as compatible with breastfeeding\(^{121}\) (see Table 61-1).

**Toxicity Monitoring**

The ACR recommends baseline screening with CBC, liver transaminases, and creatinine. Subsequently, no routine laboratory monitoring is recommended.\(^{124}\) In 2002, the American Academy of Ophthalmology (AAO) released an information statement on screening recommendations for CQ and HCQ retinopathy.\(^{257}\) Revised guidelines were published in 2011. With new data showing the risk of retinal toxicity increasing toward 1% after 5 to 7 years of use, or a cumulative dose of 1000 g of HCQ, a baseline examination is advised to rule out maculopathy, which may be a contraindication to usage. Subsequently, annual eye examinations should commence at 5 years or sooner depending on other risk factors.\(^{263}\) At a minimum, this should include a dilated eye examination, visual field testing with Humphrey 10-2 fields, and screening for color blindness in male patients. In addition, newer objective tests should be employed that are more sensitive than visual fields (multifocal electroretinogram [mfERG], spectral domain optical coherence tomography [SD-OCT], and fundus autofluorescence [FAf]). Despite these recommendations, many rheumatologists would favor annual screening or, at a minimum, screening every 2 years. Vaccination with a yearly influenza vaccine is recommended (see Table 61-2).

**Drug Interactions and Contraindications**

**Drug Interactions**

HCQ should be used with caution in diabetic patients on hypoglycemic agents. HCQ has been shown to increase digoxin levels. CQ may increase cyclosporine levels and reduce MTX levels. CQ can interfere with cytochrome P450 enzymes and should be used with caution when combined with other agents metabolized by this pathway.

**Contraindications**

Hypersensitivity to or previous retinal or visual field changes attributable to 4-aminoquinoline agents and severe liver disease due to viral hepatitis are contraindications to further use (see Table 61-2).
**COMBINATION DMARD THERAPY IN RHEUMATOID ARTHRITIS**

**KEY POINTS**

- Monotherapy with MTX infrequently induces remission in RA.
- Combinations of DMARDs have been shown to have superior efficacy without increased toxicity.
- MTX is the cornerstone of combination therapy.
- Current research is focused on the best strategies for introduction, timing, and patient selection for combination therapy.

In the early 1990s, the use of combinations of DMARDs to treat RA was rare; now this strategy is employed by essentially all rheumatologists to treat many of their patients. Currently, the timing and make-up of combinations selected to treat RA is one of the most important decisions that clinicians face.

Monotherapy with MTX is considered by most as the initial treatment of choice for early RA. Four major studies have demonstrated the superiority of combinations of DMARDs over monotherapy in head-to-head comparisons. Trials with multiple conventional DMARDs and biologics have shown them to be more effective than placebo when added to the baseline MTX in patients who have active disease despite MTX therapy. With very few exceptions, successful combination trials have included MTX, and it remains the cornerstone of combination therapy.

**History of Combination DMARD Therapy**

Early combination DMARD studies were initiated in the late 1970s. The combination of cyclophosphamide, azathioprine, and HCQ produced substantial responses in a small group of patients; however, an unacceptably high number of malignancies were reported. Thus, early enthusiasm for combination therapy was limited. During the 1980s, inadequate dosing, the use of DMARDs with marginal efficacy, and problematic trial design contributed to the less-than-exciting results reported with combination therapy. In 1994, the first trial to convincingly show superior efficacy without increased toxicity of combination DMARD therapy in a head-to-head comparison with monotherapy was reported, and multiple trials showing the success of combination therapy have ensued.

**Early Rheumatoid Arthritis**

In the late 1990s, three pivotal trials showed the success of combination DMARD therapy in early RA. Researchers in the Netherlands reported success with a step-down approach in the COBRA (Combinatietherapie Bij Reuma-toide Artritis) trial. In this trial, patients with early disease were randomized to two groups: the combination of prednisolone, MTX, and SSZ was compared with SSZ alone. Prednisolone was started at 60 mg/day, rapidly tapered, and discontinued by week 28. MTX was given until week 40. The dose of SSZ was the same in the two groups. At 28 weeks, the combination group was significantly better than the SSZ-alone group. As prednisolone and MTX were tapered, clinical responses became similar in the two groups; however, significant benefits in certain parameters were noted in the combination group. It is important to note that combination therapy was not more toxic. Subsequent data confirm that the radiographic benefits conferred by COBRA in the initial trial continue for at least 5 years.

The second important early RA study is the Finland Rheumatoid Arthritis (Fin-RA) trial. In this open trial, patients were randomized to receive combination DMARD therapy (MTX, SSZ, HCQ, and low-dose prednisolone) or monotherapy with SSZ with optional prednisolone. The major end point of this trial was remission at 2 years. Significantly, patients who received combination therapy achieved more frequent remissions. A follow-up of this trial at 5 years demonstrated that patients treated initially with the combination were less likely to have evidence of C1-C2 subluxation on cervical spine radiographs.

In a third trial from Turkey, patients with early RA were randomized to single DMARD therapy (MTX, SSZ, or HCQ), two-drug therapy (MTX and SSZ or MTX and HCQ), or three-drug therapy (MTX, SSZ, and HCQ), with remission at 2 years as the major outcome. For all end points measured, two drugs were shown to be statistically superior to monotherapy, and three drugs were statistically superior to the two-drug regimens.

The TEAR (Treatment of Early Aggressive RA) trial is the first randomized head-to-head trial in early RA comparing initial combination therapy with MTX, SSZ, and HCQ or MTX and etanercept with initial MTX monotherapy with step-up to combination therapy at 6 months for patients with a Disease Activity Score (DAS) 28 of 3.2 or greater. Results show that initial combination therapy is superior to MTX monotherapy at 6 months, but at 2 years, no differences occur between any of the combination groups, regardless of initial or step-up to combination therapy.

In sum, combination therapy, whether step-up or initial therapy, has been shown in multiple trials to be superior to MTX alone in early rheumatoid arthritis. Predictors of who may respond to MTX monotherapy alone and who needs early combination therapy are still lacking.

**Patients with Active Disease Despite Methotrexate**

Combination DMARD therapy was first studied in groups of patients with active disease despite MTX therapy, or suboptimal MTX responders. The first study to show the advantage of combination therapy with MTX and another DMARD compared with continued therapy with MTX alone in this group of patients was the cyclosporine-MTX trial. Therapy with the combination of MTX, SSZ, and HCQ, so-called triple therapy, has been shown to be well tolerated and more effective than MTX monotherapy or SSZ monotherapy, and, in an open trial of patients with early disease, more effective than the double combination of MTX and SSZ or MTX and HCQ.
In 1996, a 2-year, randomized, double-blind, parallel study of 102 established RA patients was done to compare triple-drug therapy (MTX, SSZ, and HCQ) versus double therapy (HCQ and SSZ) versus monotherapy with MTX. 66 Significantly more patients receiving triple-drug therapy achieved a modified Paulus 50% response, 213 compared with those given double therapy. Triple therapy was well tolerated, and numerically fewer withdrawals occurred in the combination group compared with the other two groups. This therapy has been shown to be durable, with 62% of patients remaining on triple therapy for 5 years continuing to maintain a 50% efficacy response. 276

In follow-up, a 2-year, double-blind trial on patients with moderately advanced disease was done in 2002, in which the triple combination was compared with two double combinations (MTX plus SSZ and MTX plus HCQ) in a head-to-head comparison. 274 Patients were stratified for previous MTX use, and previous users had to have active disease despite receiving 17.5 mg/wk. The triple-therapy group and both double-therapy groups tolerated their treatments well, with only 8% withdrawing for toxicities, which were mostly minor. Triple therapy was shown to be superior to either of the double combinations.

A double-blind, placebo-controlled trial (see Figure 61-8) compared the addition of leflunomide or placebo to baseline MTX in suboptimal MTX responders. 277 Leflunomide or placebo was added to MTX. The combination group was statistically superior to the MTX-placebo group in terms of the American College of Rheumatology 20 set (ACR20) criteria for clinical response. The combination was reasonably well tolerated, but side effects including diarrhea, nausea, and dizziness were increased in the combination arm. Elevated ALT levels (>1.2 times normal) occurred more frequently in patients on combination therapy than in those on MTX alone, with increases leading to withdrawal in 2.3% of patients who received the combination.

In one arm of the previously discussed TEAR trial, combination triple therapy or MTX and etanercept was added for MTX nonresponders. Both groups responded, and eventually no differences were seen in the DAS28 for the step-up group compared with initial combination therapy patients. 278

Corticosteroids in DMARD Combinations

Corticosteroids have not traditionally been considered DMARDs. However, they clearly fulfill all of the criteria for DMARDs, including retarding radiographic progression. 278 Few clinicians who care for patients with RA dispute their efficacy. Indeed, they have been used as baseline therapy for well over half of the patients included in the combination trials discussed previously.

Prednisolone undoubtedly was a critical component for the success of the COBRA protocol 265 and may have played a role in the success of the combination group in the Fin-RA trial. 275 Kirwan and colleagues’ report of the ability of prednisolone to significantly retard radiographic progression of RA compared with placebo is testament to the efficacy of steroids when used in combination with other DMARDs. 278 Corticosteroids clearly deserve further formal investigation as a component of combination therapy. The COBRA trial and the Kirwan data have raised another interesting question: Should/could short courses of high-dose steroids be used as a form of induction therapy? 280

Biologic Agents in DMARD Combinations

Biologic agents that block TNF (etanercept, infliximab, adalimumab, and golimumab) and IL-1 (anakinra) have
been studied in early and established RA in combination with MTX.\textsuperscript{35,38} (see Figure 61-8). These trials have shown superior improvements in clinical and radiographic end points in the combination groups.\textsuperscript{286,287} Other biologic agents—rituximab, an anti-CD20 monoclonal antibody; abatacept, a T cell co-stimulatory inhibitor; and tocilizumab, an IL-6 receptor antagonist—have been studied in combination with MTX.\textsuperscript{289,290} In the REFLEX (Randomized Evaluation of Long-Term Efficacy of Rituximab in RA) trial, rituximab versus placebo was added to baseline MTX in RA patients who were suboptimal responders to MTX with failure of one or more TNF inhibitors.\textsuperscript{289} Results showed that the combination group had significant improvements in ACR-N responses. Abatacept versus placebo in addition to background MTX was studied in RA patients with a suboptimal response to MTX, and results at 1 year showed significant improvements in clinical and radiographic end points.\textsuperscript{290} The LITHE (Tocilizumab Safety and The Prevention of Structural Joint Damage) study compared the addition of tocilizumab versus placebo in MTX nonresponders and showed an improvement in structural outcomes at 1 year.\textsuperscript{291} Concerns do exist regarding the risks of infection and infusion reaction in combination therapy that includes biologic agents.

**Selecting the Right Patients for Combination Therapy**

Factors that predict a poor prognosis for patients with RA are well accepted and include rheumatoid factor, elevated erythrocyte sedimentation rate and C-reactive protein (CRP), the number of joints involved, erosions, and the presence of certain genetic markers. However, unless these factors can be shown to predict response to certain therapies in a differential fashion, they are of limited therapeutic use. Patient characteristics recommending one therapeutic regimen over another remain to be fully elucidated. Genetic differences have been suggested to influence outcomes in a differential fashion. Until this observation can be corroborated and factors that predict response to other therapies elucidated, choices will remain largely empiric.

Treatment of patients with RA using MTX combinations should be the gold standard against which future therapies are compared. Available data demonstrate that a variety of combinations are more effective than MTX alone. Many questions remain to be answered regarding the appropriate timing of combination therapy and the optimal combinations for specific patients and for specific clinical situations (e.g., induction, maintenance therapy, suboptimal response to MTX). Future research is needed to clarify the role of corticosteroids and, particularly, biologic response modifiers (specifically anti-TNF therapies) as components of and alternatives to MTX combination regimens.

**Selected References**


Full references for this chapter can be found on www.expertconsult.com.
References


PART 8


Immunosuppressive drugs comprise different classes of drugs that dampen the immune system—notably T and B lymphocytes—functionally and/or numerically (Table 62-1) but do not permanently correct the fundamental imbalance of immune regulation in autoimmune disease. As such, they do not have curative potential yet they can be effective in remission induction and control of specific rheumatic disease manifestations and remain cornerstone drugs in the management of rheumatic conditions. Many immunosuppressive drugs have outlasted the test of time, as attested by their ongoing use in transplantation medicine, nephrology, gastroenterology, ophthalmology, dermatology, and rheumatology. Consequently, their therapeutic potential and toxicity profiles hold few surprises. Apart from drug-specific toxicities, the main risk of immunosuppressive treatment is infection. In the absence of validated biomarkers of infection, sound clinical judgment and experience remain indispensable in monitoring patients who use immunosuppressive drugs, often for long periods of time. The use of live vaccines is contraindicated, and although other vaccinations are generally less effective, annual influenza vaccination is recommended in patients taking immunosuppressive medication.

This chapter outlines the clinical pharmacology and therapeutic use of immunosuppressive drugs used in rheumatology. These include cytostatic agents that affect bone marrow progenitor cells (cyclophosphamide, chlorambucil, and azathioprine) and drugs such as cyclosporine, sirolimus, tacrolimus, and mycophenolate mofetil (MMF) that target lymphocytes by inhibiting specific intracellular signaling pathways and/or proliferation. Their effects on the immune system overlap with those of traditional disease-modifying antirheumatic drugs such as methotrexate, glucocorticoids, and the newer biologics. Thalidomide is also discussed in this chapter, although its therapeutic utility is limited. The most commonly used immunosuppressive drugs—cyclophosphamide, azathioprine, and MMF—are discussed in more detail. Glucocorticoids, methotrexate, leflunomide, and biologic agents are discussed elsewhere.

**ALKYLATING AGENTS**

Alkylating agents substitute alkyl radicals into deoxyribonucleic acid (DNA), which ultimately results in cell death. Cyclophosphamide was introduced as an antitumor agent in 1958 and is still one of the most widely administered anticancer agents and one of the most potent immunosuppressants. It is the drug of choice for remission induction therapies in severe systemic lupus erythematosus (SLE) and necrotizing vasculitis. Chlorambucil is rarely used in...
Table 62-1  Mechanisms of Action of Immunosuppressive Drugs

<table>
<thead>
<tr>
<th>Drugs</th>
<th>Class</th>
<th>Mechanism of Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyclophosphamide, chlorambucil</td>
<td>Alkylating cytotoxics</td>
<td>Active metabolites alkylate DNA</td>
</tr>
<tr>
<td>Azathioprine, mercaptopurine</td>
<td>Purine analogue cytotoxics</td>
<td>Inhibit purine synthesis</td>
</tr>
<tr>
<td>Cyclosporine, tacrolimus (FK506)</td>
<td>Calcineurin inhibitors</td>
<td>Inhibit calcium-dependent T cell activation and interleukin-2 production</td>
</tr>
<tr>
<td>Sirolimus (rapamycin)</td>
<td>Noncalcineurin-binding macrolide immunoregulator</td>
<td>Blocks interleukin-2-mediated and growth factor–mediated signal transduction</td>
</tr>
<tr>
<td>Mycophenolate mofetil</td>
<td>Purine synthesis inhibitor</td>
<td>Mycophenolic acid inhibits inosine monophosphate dehydrogenase</td>
</tr>
<tr>
<td>Thalidomide</td>
<td>Glutamic acid derivative</td>
<td>Inhibition of tumor necrosis factor production and angiogenesis</td>
</tr>
</tbody>
</table>

Cyclophosphamide

**Structure**

Cyclophosphamide is an oxazaphosphorine-substituted nitrogen mustard and inactive prodrug requiring enzymatic bioactivation (Figure 62-1). Cyclophosphamide is the alkylating agent of choice for most rheumatic disease requiring such therapy.

**Mechanisms of Action**

Its DNA-alkylating effects are mediated predominantly through phosphoramid mustard and, to a lesser extent, other active metabolites. These positively charged, reactive intermediates alkylate nucleophilic bases, resulting in the cross-linking of DNA and of DNA proteins, breaks in DNA, and consequently decreased DNA synthesis and apoptosis. The cytotoxicity of alkylating agents correlates with the amount of DNA cross-linking, but the relationship between cytotoxicity and immunosuppressive effects is unclear. The effects of cyclophosphamide are not exclusively limited to proliferating cells or particular cell types. Sensitivity varies among cell populations, however; for example, hematopoietic progenitor cells are relatively resistant to even high doses of cyclophosphamide. The immunosuppressive effects of cyclophosphamide include decreased numbers of T lymphocytes and B lymphocytes, decreased lymphocyte proliferation, decreased antibody production, and suppression of delayed hypersensitivity to new antigens with relative preservation of established delayed hypersensitivity.

**Pharmacology**

**Absorption and Distribution.** Oral and intravenous (IV) administration of cyclophosphamide results in similar plasma concentrations. Peak plasma concentrations of cyclophosphamide occur 1 hour after oral administration. Protein binding of cyclophosphamide is low (20%), and it is widely distributed.

**Metabolism and Elimination.** Cyclophosphamide is rapidly metabolized, largely by the liver, to active and inactive metabolites. The formation of the active 4-hydroxy-cyclophosphamide is mediated by various cytochrome P-450 (CYP) enzymes, and genetic variations in the enzymes in lupus nephritis patients have been shown to affect responses to cyclophosphamide. 4-Hydroxy-cyclophosphamide, which is not cytotoxic at physiologic pH, readily diffuses into cells and spontaneously decomposes into the active phosphoramid mustard. The elimination half-life of cyclophosphamide is 5 to 9 hours, and alkylating activity is undetectable in the plasma of most patients 24 hours after a dose of 12 mg/kg. Plasma concentrations of cyclophosphamide are not clinically useful predictors of either efficacy or toxicity. Between 30% and 60% of the total cyclophosphamide is eliminated in the urine, mostly as inactive metabolites, although some cyclophosphamide and active metabolites such as phosphoramid mustard and acrolein can also be detected in urine.

**Pharmacokinetic Considerations in Special Circumstances**

Liver Disease. Although the half-life of cyclophosphamide is increased to 12 hours in patients with liver failure compared with 8 hours in controls, toxicity is not increased, suggesting that exposure to cytotoxic metabolites is not increased and dose modification in liver disease is generally not required.

Figure 62-1  The metabolism of cyclophosphamide. Cyclophosphamide is converted to 4-hydroxy-cyclophosphamide, in equilibrium with its tautomer aldophosphamide, by cytochrome P-450 enzymes. Subsequent nonenzymatic processes lead to the formation of phosphoramid mustard and acrolein. Oxidation of 4-hydroxy-cyclophosphamide and aldophosphamide through enzymes including aldehyde dehydrogenase results in inactive metabolites. Cytotoxic metabolites are shown in bold type.
Renal Impairment. Some studies have shown little alteration in drug disposition with no increased toxicity in patients with impaired renal function.\(^1\) In patients with autoimmune disease and a creatinine clearance of 25 to 50 mL/min and 10 to 25 mL/min, exposure to cyclophosphamide increased approximately 40% and 70%, respectively.\(^2\) In clinical practice, initial cyclophosphamide doses are therefore decreased by approximately 30% in patients with moderate-to-severe renal impairment and subsequent doses are titrated according to clinical response and effects on the leukocyte (white blood cell) count. Cyclophosphamide is removed by dialysis and is administered after dialysis, or, alternatively, dialysis can be initiated the day after cyclophosphamide administration.\(^5\)

Clinical Indications

Cyclophosphamide remains the drug of choice for most patients with systemic necrotizing vasculitis or Goodpasture's syndrome, for many patients with organ-threatening SLE, and for some patients with autoimmune disease-associated interstitial lung disease and inflammatory eye disease. In rheumatoid arthritis (RA) unless complicated by vasculitis, less toxic and more effective drugs have replaced cyclophosphamide. In SLE a remission induction course with IV cyclophosphamide followed by maintenance with azathioprine or mycophenolate to minimize cyclophosphamide toxicity is the most commonly used treatment for severe organ involvement including lupus nephritis, although remission induction regimens with MMF have been propagated as an effective and safe alternative for cyclophosphamide (discussed later). The original National Institutes of Health (NIH) protocol involved 6 monthly IV infusions with cyclophosphamide 1 g/m\(^2\) then once every 3 months for at least 24 additional months,\(^6\) whereas the Euro-Lupus protocol used in Europe involved administration of six IV infusions of 500 mg of cyclophosphamide every 2 weeks followed by azathioprine maintenance (Table 62-2). A comparison with 6 monthly IV infusions with cyclophosphamide 500 mg/m\(^2\), followed by two further infusions of slightly higher doses 3 and 6 months later, and azathioprine maintenance therapy resulted in similar rates of the end points of end-stage renal disease or doubling of creatinine concentration with up to 10 years of follow-up.\(^7\) Cyclophosphamide as either IV pulse therapy or orally can also be effective in patients with other serious complications of SLE including central nervous system involvement and thrombocytopenia and interstitial lung disease associated with systemic sclerosis and other autoimmune diseases.\(^8-10\) Several trials have investigated whether IV cyclophosphamide is as effective as oral cyclophosphamide as remission induction therapy for granulomatous with polyangiitis (GPA), the newly proposed name for Wegener's granulomatosis.\(^11\) Although early trial results suggested superiority of oral dosing, more recent clinical trial data pointed to equal efficacy, but slightly less hematologic toxicity with IV therapy.\(^12,13\) As in lupus nephritis, shorter induction courses of cyclophosphamide have been reported to be effective in GPA and microscopic polyangiitis.\(^13\) Cyclophosphamide has a steep dose-response curve, making it an ideal compound for dose escalation. High doses of cyclophosphamide, with or without stem cell rescue and lymphoablative antibodies or total body irradiation, have been used for severe juvenile idiopathic arthritis (JIA), RA, systemic sclerosis, and SLE.\(^16\)

With the introduction of effective biologics and new treatment paradigms for RA and JIA, the clinical need for immunosuppressant therapy has waned. Although large series have shown promising results of immunosuppressive therapy and stem cell rescue in patients with severe SLE, a recent randomized trial showed that standard-dose IV cyclophosphamide was not inferior to high-dose cyclophosphamide without stem cell rescue or lymphoablative antibodies.\(^17\) Prospective, randomized trials are in progress in systemic sclerosis to compare safety and efficacy of IV pulse cyclophosphamide and immunosuppressive therapy with stem cell rescue.

Dosage and Route of Administration

Typical dosage regimens are presented in Table 62-2. Dosages for IV pulse therapy with cyclophosphamide range from 0.5 to 1 g/m\(^2\) and for oral therapy 2 mg/kg. The bioavailability of oral cyclophosphamide is excellent.

Toxicity

Hematologic. Reversible myelosuppression manifesting as leukopenia and neutropenia is common and dose dependent. Generally, platelet counts are not affected with IV pulse doses of less than 50 mg/kg, but with long-term oral use, a mild decrease in platelet count is common. After a single IV dose of cyclophosphamide, the approximate times to nadir and recovery of leukocyte counts are 8 to 14 days and 21 days, respectively.\(^18\) The white blood cell nadir is about 3000 cells/mm\(^3\) after a dose of 1 g/m\(^2\) (=25 mg/kg) and 1500 cells/mm\(^3\) after a dose of 1.5 g/m\(^2\). With long-term use,
there is increased sensitivity to the myelosuppressive effects of cyclophosphamide and doses usually need to be decreased over time.

**Infection.** Infection with a range of common and opportunistic pathogens is a frequent complication. In 100 patients with SLE, infection occurred in 45 patients during treatment with a cyclophosphamide-based regimen and was the primary cause of death in 7 patients. In this study, infection was equally common in patients receiving oral or IV cyclophosphamide and was associated with a white blood cell nadir at some point in treatment of less than 3000/mm³ (55% infection rate vs. 36%). At the time of infection, the average white blood cell count was normal, however. A higher maximal corticosteroid dose was also associated with increased risk of infection. Half of the infections occurred at prednisone doses of less than 40 mg/day, and a quarter of the infections occurred at doses less than 25 mg/day. Lower rates of infection (25% to 30%) have been reported in SLE patients receiving cyclophosphamide in National Institutes of Health (NIH) protocols. Oral cyclophosphamide regimens generally pose a greater risk of infection than IV pulse regimens. Serious infections occurred in 41% and 70% of patients with GPA treated with pulse IV and daily oral cyclophosphamide, respectively. These rates of infection are higher than rates reported in long-term NIH protocols, in which 48% of 158 patients experienced 140 infections requiring hospitalization. The reported frequency of cyclophosphamide-associated infection varies, probably as a function of the stage and severity of the underlying disease, the degree of cyclophosphamide-induced immunosuppression, and variations in concomitant glucocorticoid regimens. *Pneumocystis jiroveci* pneumonia has been recognized as a preventable, serious opportunistic infection that complicates treatment of systemic vasculitis with regimens using cyclophosphamide and methotrexate. The risk is highest during the remission induction phase and is greater with oral than IV cyclophosphamide regimens. Surprisingly, in two placebo-controlled, randomized clinical trials in scleroderma lung disease, active treatment for 1 year with either oral cyclophosphamide or sequential treatment with prednisolone plus IV cyclophosphamide followed by azathioprine was not associated with more toxicity, however, suggesting disease-specific differences in toxicity.

**Urologic.** The bladder toxicities of cyclophosphamide, hemorrhagic cystitis, and bladder cancer are related to route of administration, duration of therapy, and cumulative cyclophosphamide dose. Bladder toxicity, a particular problem with long-term oral cyclophosphamide, is largely due to acrolein, a metabolite of cyclophosphamide. It is commonly accepted that bladder toxicity can be minimized in patients receiving pulse doses of IV cyclophosphamide by administering mesna, a sulfhydryl compound that binds acrolein in the urine and inactivates it. Direct evidence for the effectiveness of mesna in preventing cystitis, however, comes from its use with ifosfamide in patients with cancer and data from animal models. The data from rheumatology series are consistent with a protective effect but are inadequate to come to firm conclusions, which explains differences between national guidelines. The short half-life of mesna renders it suboptimal for the prevention of bladder toxicity in patients receiving daily oral cyclophosphamide—but oral mesna administered three times a day with daily oral cyclophosphamide decreased the incidence of bladder toxicity to 12%.

Nonglomerular hematuria, which may range from minor, microscopic blood loss to severe, macroscopic bleeding, is the most common manifestation of cyclophosphamide-induced cystitis. Nonglomerular hematuria occurred at some time in 50% of 145 patients treated with oral cyclophosphamide and was related to the duration of therapy and cumulative cyclophosphamide dose. The risk of bladder cancer was increased 31-fold (95% confidence interval [CI], 13-fold to 65-fold), and 7 patients (5%) had developed bladder cancer anytime between 7 months and 15 years after initiating therapy. The cancer was preceded by nonglomerular hematuria in all patients. Six of the seven patients had a cumulative dose of more than 100 g of cyclophosphamide and a duration of therapy of more than 2.7 years. Smokers were at increased risk of hemorrhagic cystitis and bladder cancer.

**Malignancy.** Cyclophosphamide increases the risk of malignancies (other than bladder cancer) twofold to fourfold. In the largest study, 119 patients with RA who had been treated with oral cyclophosphamide were followed for 20 years. There were 50 cancers in 37 patients in the cyclophosphamide group compared with 26 cancers in 25 of 119 control RA patients. Bladder, skin, myeloproliferative, and oropharyngeal malignancies occurred more commonly in the cyclophosphamide group. The risk of malignancies increased with the cumulative dose of cyclophosphamide, and 53% of patients who received more than 80 g of cyclophosphamide developed malignancy. Few malignancies have been reported in patients treated with pulse IV cyclophosphamide regimens. Current data do not allow quantification of the long-term risk of malignancy associated with pulse IV cyclophosphamide treatment, but it is likely to be substantially smaller than that associated with oral regimens.

**Reproductive.** Cyclophosphamide, as used in autoimmune disease, results in significant gonadal toxicity. The risk of sustained amenorrhea after cyclophosphamide therapy has ranged from 11% to 59%. The risk of ovarian failure depends more on age of the patient and cumulative dose of cyclophosphamide than on route of administration. Patients younger than 25 years old receiving 6 pulses of IV cyclophosphamide had a low frequency of ovarian failure (none of four patients), whereas patients older than 31 years receiving 15 to 24 pulses all had ovarian failure (four of four patients). The use of alkylating agents in male patients leads to azoospermia, and, if the clinical situation allows, referral to a fertility clinic for banking of sperm (or ova in female patients) should be considered before cyclophosphamide treatment. There was no increase in genetic disease in the offspring of adults who underwent cancer chemotherapy in childhood.

**Pulmonary.** Cyclophosphamide-induced pulmonary toxicity occurs in less than 1% of patients. Early-onset pneumonitis 1 to 6 months after exposure to cyclophosphamide may respond to withdrawal of the drug and treatment with corticosteroids. A more insidious, irreversible, late-onset pneumonitis and fibrosis with radiographic findings of
diffuse reticular or reticulonodular infiltrates may occur after treatment with oral cyclophosphamide for 1 to 13 years.28

Miscellaneous. A varying degree of reversible alopecia can occur with daily oral and monthly pulse cyclophosphamide. Cardiotoxicity, a dose-limiting adverse effect in oncology, and water intoxication, owing to inappropriate antidiuretic hormone secretion, are rare at standard doses.30 Unusual hypersensitivity reactions include urticaria and anaphylaxis, although the bladder protectant mesna is a more likely cause of allergic responses in patients receiving both drugs.31,32

Strategies to Minimize Toxicity. Strategies to minimize toxicity include adjusting the dose of cyclophosphamide to avoid a significant degree of leukopenia (white blood cell count <3000/mm³ for daily oral therapy or a nadir of <2000/mm³ for pulse IV therapy) and granulocytopenia.33 The blood count is monitored initially at 1- to 2-week intervals and monthly thereafter in patients on stable oral doses. To decrease the risk of infection added by concomitant high-dose corticosteroids, the dose of corticosteroids should be reduced after a clinical response has been obtained. Alternate-day glucocorticoids can be considered in the maintenance phase. Oral cyclophosphamide is best administered as a single dose in the morning with the patient drinking plenty of fluids and emptying the bladder frequently to dilute the urinary concentration of acrolein and to minimize the time the bladder is exposed to it. Prophylaxis against P. jiroveci pneumonia is often prescribed, particularly during the induction phase when doses of cyclophosphamide and corticosteroids are higher. The use of mesna to prevent bladder toxicity is described earlier. Urinalysis should be performed monthly, and nonglomerular hematuria should be evaluated by a urologist. All patients who receive cyclophosphamide, particularly patients who develop hemorrhagic cystitis, are at increased risk of developing bladder cancer, and lifelong surveillance is required with urinalysis, urine cytology, and, if indicated, cystoscopy.34 Lastly, drugs that are less toxic than cyclophosphamide such as methotrexate and MMF are an option for inducing remission or remission maintenance in patients with GPA or lupus nephritis.35,36

Pregnancy and Lactation
Cyclophosphamide is a U.S. Food and Drug Administration (FDA) Pregnancy Category D drug. Cyclophosphamide is teratogenic, particularly in the first trimester, and should be avoided in pregnancy and during lactation.37,38 If a patient becomes pregnant while taking (receiving) this drug, the patient should be apprised of the potential hazard to the fetus. Women of childbearing potential should be advised to avoid becoming pregnant.

Drug Interactions
Cimetidine inhibits the activity of several hepatic enzymes. It has resulted in increased exposure to cyclophosphamide metabolites in a rabbit model.39 Ranitidine and presumably other H₂-receptor antagonists that have little effect on hepatic drug metabolism are not associated with increased cyclophosphamide toxicity.40 Allopurinol increases the half-life of cyclophosphamide and the frequency of leukopenia.41 Cyclophosphamide decreases plasma pseudocholinesterase activity and can potentiate the effect of succinylcholine.42

Chlorambucil

Structure
Chlorambucil (4(4-bis(2-chloroethyl)aminophenyl)-butyric acid) has occasionally been used as an alternative to cyclophosphamide in some rheumatologic diseases, but its clinical utility is limited now that effective immunosuppressive drugs and biologics with more favorable safety profiles are available.

Mechanism of Action
The mechanism of action of chlorambucil is similar to that of cyclophosphamide, but slower.

Pharmacology
Chlorambucil is well absorbed (>70%) after oral administration with peak concentrations occurring within 2 hours.43 Chlorambucil is extensively metabolized by β oxidation to a metabolite, phenylacetic acid mustard, which is also cytotoxic.43,44 Less than 1% of the oral dose of chlorambucil appears in the urine as unchanged drug.41 The plasma half-life of chlorambucil and the phenylacetic acid mustard metabolite is 30 to 180 minutes.43,44

Clinical Indications
Chlorambucil can be effective for patients with inflammatory eye disease including Behçet’s syndrome and occasionally for refractory dermatomyositis.47 Chlorambucil is used as an alternative alkylating agent for patients unable to tolerate cyclophosphamide because of bladder toxicity or gastrointestinal intolerance. Chlorambucil does not seem to be as effective as cyclophosphamide, however, in the treatment of vasculitis or glomerulonephritis, or systemic sclerosis. The use of chlorambucil in RA is obsolete given the availability of a multitude of alternative drugs with a more favorable risk-benefit profile.

Dosage
Chlorambucil is often started at an oral dose of 0.1 mg/kg/day; the dose is increased or decreased according to clinical response and toxicity. Doses of 0.2 mg/kg/day or greater are associated with more frequent myelosuppression. Alternatively, a “start-low, go-slow” approach, with chlorambucil started at a dose of 4 mg/day and increased in 1-mg increments at 1- to 2-month intervals, if required, may be better tolerated. Even with this approach, 75% of patients discontinued therapy because of chlorambucil-related toxicity.38 Regular monitoring, particularly of the white blood cell count, at approximately 2-week intervals initially, and then monthly when stable, is required.
Toxicity

**Hematologic.** Myelosuppression is common and may be abrupt in onset. The degree of leukopenia and neutropenia is dose related, but there are considerable interindividual differences in sensitivity. Myelosuppression is usually reversible; however, it may take several months for the white blood cell count to return to the normal range, and some patients remain relatively leukopenic. Irreversible, fatal bone marrow suppression has been reported in patients receiving chlorambucil for rheumatic disease.48

**Infection.** The average frequency of herpetic zoster infection is 13%. As is the case with cyclophosphamide, infections resulting from a wide range of bacterial and nonbacterial pathogens occur.48

**Malignancy.** Treatment with chlorambucil increases the risk of leukemia, particularly myeloid leukemia, and has been associated with a range of lymphomas.48,49 Various solid organ tumors have occurred in association with chlorambucil, but no causal relationship has been established.48

**Miscellaneous.** Other adverse effects of chlorambucil include azospermia and amenorrhea, which are usually reversible. Pulmonary fibrosis, oral ulceration, hepatotoxicity, nausea, fever, and rashes are other adverse effects.

**Pregnancy and Lactation**

Chlorambucil is an FDA Pregnancy Category D drug. It should not be used in pregnancy or lactation. Chlorambucil can cause fetal harm when administered to a pregnant woman. If a patient becomes pregnant while taking this drug, the patient should be apprised of the potential hazard to the fetus. Women of childbearing potential should be advised to avoid becoming pregnant.

**PURINE ANALOGUES**

**Azathioprine**

*Structure*

Azathioprine is a prodrug that is converted to 6-mercaptopurine involving the removal of an imidazole group.60 6-Mercaptopurine is a purine analogue that acts as a cycle-specific antimetabolite chemotherapeutic agent interfering with the synthesis of nucleotides, thereby inhibiting proliferation of lymphocytes. Azathioprine has a better therapeutic index than 6-mercaptopurine and has replaced it in the treatment of rheumatic autoimmune disease.

**Mechanisms of Action**

The exact mechanism of action of the active thiopurine metabolites of azathioprine in autoimmune disease is unknown. Thiopurine metabolites such as thioguanine nucleotides decrease the de novo synthesis of purine nucleotides by inhibiting amidotransferase enzymes and purine ribonucleotide interconversion and are incorporated into DNA and ribonucleic acid (RNA).60 The incorporation of thioguanine nucleotides into the nucleic acids of cells is thought to mediate the cytotoxicity of azathioprine, whereas inhibition of purine synthesis may be more important in decreasing cellular proliferation. Leukopenia is unnecessary for immunosuppression. Azathioprine decreases the circulating lymphocyte count, suppresses lymphocyte proliferation, inhibits antibody production, inhibits monocyte production, suppresses natural killer cell activity, and inhibits cell-mediated and humoral immunity.

**Pharmacology**

**Absorption and Distribution.** Oral azathioprine is well absorbed and rapidly converted to 6-mercaptopurine, which is further metabolized to several compounds including 6-thiourate (Figure 62-2) that are excreted in urine. The plasma half-life of azathioprine is less than 15 minutes but it is 1 to 3 hours for the active derivative, 6-MP.61 The bioavailability of azathioprine, measured as the concentrations of mercaptopurine achieved after oral administration, varies. In healthy volunteers, bioavailability ranged from 27% to 83% with an average of 47%.61 Mercaptopurine is widely distributed with a volume of distribution of 4 to 8 L/kg.67

**Metabolism and Elimination.** The metabolism of azathioprine is complex60,62 and has been simplified in Figure 62-2. Two enzymes, xanthine oxidase and thiopurine methyltransferase (TPMT), shunt mercaptopurine metabolites to relatively inactive compounds, whereas other enzymes such as hypoxanthine-guanine-phosphoribosyl-transferase lead to the formation of cytotoxic thiopurine nucleotides. Low TPMT activity or inhibition of xanthine oxidase by drugs such as allopurinol leads to decreased detoxification and increased formation of cytotoxic metabolites after the administration of azathioprine or mercaptopurine. Maximal concentrations of mercaptopurine occur 1 to 3 hours after administration of azathioprine, and the half-life of mercaptopurine is 1 to 2 hours.61 The half-life of the intracellular, active 6-thioguanine nucleotides is estimated to be 1 to 2

[Figure 62-2: Azathioprine is converted to mercaptopurine (6-MP) enzymatically by glutathione-S-transferase and nonenzymatic mechanisms. Xanthine oxidase and thiopurine methyltransferase metabolize 6-MP to the inactive metabolites 6-methylthiouric acid and 6-methylmercaptopurine. Hypoxanthine-guanine-phosphoribosyl-transferase (HGPR7) metabolizes 6-MP to active, cytotoxic thiopurine nucleotides.]

---

**Absorption and Distribution.** Oral azathioprine is well absorbed and rapidly converted to 6-mercaptopurine, which is further metabolized to several compounds including 6-thiourate (Figure 62-2) that are excreted in urine. The plasma half-life of azathioprine is less than 15 minutes but it is 1 to 3 hours for the active derivative, 6-MP.61 The bioavailability of azathioprine, measured as the concentrations of mercaptopurine achieved after oral administration, varies. In healthy volunteers, bioavailability ranged from 27% to 83% with an average of 47%.61 Mercaptopurine is widely distributed with a volume of distribution of 4 to 8 L/kg.67

**Mechanisms of Action**

The exact mechanism of action of the active thiopurine metabolites of azathioprine in autoimmune disease is unknown. Thiopurine metabolites such as thioguanine nucleotides decrease the de novo synthesis of purine nucleotides by inhibiting amidotransferase enzymes and purine ribonucleotide interconversion and are incorporated into DNA and ribonucleic acid (RNA).60 The incorporation of thioguanine nucleotides into the nucleic acids of cells is thought to mediate the cytotoxicity of azathioprine, whereas inhibition of purine synthesis may be more important in decreasing cellular proliferation. Leukopenia is unnecessary for immunosuppression. Azathioprine decreases the circulating lymphocyte count, suppresses lymphocyte proliferation, inhibits antibody production, inhibits monocyte production, suppresses natural killer cell activity, and inhibits cell-mediated and humoral immunity.

**Pharmacology**

**Absorption and Distribution.** Oral azathioprine is well absorbed and rapidly converted to 6-mercaptopurine, which is further metabolized to several compounds including 6-thiourate (Figure 62-2) that are excreted in urine. The plasma half-life of azathioprine is less than 15 minutes but it is 1 to 3 hours for the active derivative, 6-MP.61 The bioavailability of azathioprine, measured as the concentrations of mercaptopurine achieved after oral administration, varies. In healthy volunteers, bioavailability ranged from 27% to 83% with an average of 47%.61 Mercaptopurine is widely distributed with a volume of distribution of 4 to 8 L/kg.67

**Metabolism and Elimination.** The metabolism of azathioprine is complex60,62 and has been simplified in Figure 62-2. Two enzymes, xanthine oxidase and thiopurine methyltransferase (TPMT), shunt mercaptopurine metabolites to relatively inactive compounds, whereas other enzymes such as hypoxanthine-guanine-phosphoribosyl-transferase lead to the formation of cytotoxic thiopurine nucleotides. Low TPMT activity or inhibition of xanthine oxidase by drugs such as allopurinol leads to decreased detoxification and increased formation of cytotoxic metabolites after the administration of azathioprine or mercaptopurine. Maximal concentrations of mercaptopurine occur 1 to 3 hours after administration of azathioprine, and the half-life of mercaptopurine is 1 to 2 hours.61 The half-life of the intracellular, active 6-thioguanine nucleotides is estimated to be 1 to 2
weeks, however, and concentrations do not change over the 24-hour dose period in patients receiving daily azathioprine.\textsuperscript{51} At conventional rheumatologic doses, approximately 1% of mercaptopurine is excreted unchanged in the urine.\textsuperscript{51} Increased toxicity can occur with renal impairment (creatinine clearance <25 mL/min), and a modest dose reduction is usually necessary. The substantial interindividual variability in azathioprine disposition and TPMT activity are more important determinants of sensitivity to azathioprine than renal function.\textsuperscript{59} Azathioprine is only slightly dialyzable (10%) through conventional hemodialysis membranes.

**Dosage**

Azathioprine is often started at a dose of 1 mg/kg daily, and if this is tolerated, the dose is increased to 2 to 2.5 mg/kg after 2 to 4 weeks. A gradual increase in dose is often better tolerated. The onset of immunosuppressive effects is relatively slow, over several weeks, presumably because the active thioguanine metabolites slowly accumulate intracellularly.

**Clinical Indications**

In current practice, azathioprine is mainly used in the treatment of connective tissue disease rather than inflammatory joint disease. In RA azathioprine is less effective than methotrexate and has a slow mode of onset when compared with other disease-modifying antirheumatic drugs and biologics. Nevertheless, it remains a treatment option for RA patients with refractory disease or as a corticosteroid-sparing agent in those with organ involvement. Azathioprine is used to treat some patients with lupus nephritis, and although it is more effective than corticosteroids alone, it is not as effective as IV pulse cyclophosphamide to induce remission.\textsuperscript{55} It is effective in maintenance therapy, however,\textsuperscript{56} even after low-dose cyclophosphamide induction. For other manifestations of SLE including cutaneous disease, azathioprine is widely used as a glucocorticoid-sparing agent.\textsuperscript{57} Azathioprine in combination with corticosteroids is useful in the treatment of a range of other autoimmune diseases including inflammatory muscle disease, inflammatory eye disease including Behçet’s syndrome,\textsuperscript{58} psoriatic arthritis,\textsuperscript{59} reactive arthritis, and various forms of vasculitis. In systemic vasculitis, azathioprine is less effective as a remission-induction agent than cyclophosphamide\textsuperscript{15} but safer as a remission-maintenance and glucocorticoid-sparing drug.\textsuperscript{37} A small study has shown efficacy, however, of high-dose (1200- to 1800-mg infusions) monthly azathioprine as initial treatment of GPA and lupus nephritis.\textsuperscript{60} Azathioprine is frequently used in systemic sclerosis and overlap syndromes, especially in those with interstitial lung disease or joint involvement.\textsuperscript{10,61}

**Toxicity**

**Hematologic.** Reversible myelosuppression is dose related but varies among individuals. Low-dose azathioprine (1 to 2 mg/kg/day) rarely results in leukopenia or thrombocytopenia. Pure red cell aplasia is also rare. Severe myelosuppression is uncommon and caused by low or absent TPMT activity. Decreased TPMT activity leads to a decreased ability to detoxify mercaptopurine and results in increased formation of cytotoxic thioguanine metabolites and clinical toxicity.\textsuperscript{62} TPMT activity is polymorphic with a trimodal distribution. Approximately 90% of subjects show high activity, 10% show intermediate activity, and 0.3% (the subjects homozygous for the poorly functional polymorphisms) show very low activity.\textsuperscript{62,63} The median TPMT activity in African-Americans is approximately 17% lower than in white Americans.\textsuperscript{64} The 1 in 300 subjects with low or absent TPMT activity is at great risk of severe azathioprine-induced myelosuppression, which has a delayed but sudden onset, most commonly 4 to 10 weeks after azathioprine has been started.\textsuperscript{64} More than half of all cases of leukopenia in patients receiving azathioprine have a normal TPMT genotype and phenotype, however.

**Gastrointestinal.** Liver test abnormalities occur in 34% but are seldom serious. Serious liver toxicity, severe cholestasis, hepatic veno-occlusive disease, and nodular regenerative hyperplasia and pancreatitis are rare.

**Malignancy.** Data regarding the risk of malignancy in patients treated with azathioprine for rheumatologic disease are conflicting. Some studies found an increased risk, particularly of lymphoproliferative malignancies, whereas others did not.\textsuperscript{65} A 24-year retrospective study of 358 SLE patients found no difference in malignancy rates between patients who had received azathioprine and patients who had not, and there were no lymphomas in the azathioprine group.\textsuperscript{66}

**Azathioprine Hypersensitivity.** Acute hypersensitivity syndromes, usually occurring within 2 weeks of starting therapy, with a range of manifestations including shock, fever, rash, pancreatitis, renal failure, and hepatitis are rare.\textsuperscript{67}

**Others.** Infection is less common with azathioprine than with alkylating agents; however, infections with a range of bacterial and nonbacterial pathogens including herpes zoster and cytomegalovirus may occur. The rate of infection when azathioprine is administered alone or with low doses of glucocorticoids is approximately 2.5 per 100 person-years of exposure. Maculopapular or urticarial rashes can occur. Eosinophilia and drug fever are rare.

**Strategies to Minimize Toxicity.** TPMT activity testing is the most commonly used method to identify patients at risk of serious toxicity. More than 23 variants in the TPMT gene, associated with decreased TPMT activity, have been identified, with the TPMT*2, TPMT*3A, and TPMT*3C alleles accounting for most of the intermediate- and low-activity cases. The concordance between TPMT genetics and phenotypes is slightly less than 100%. TPMT activity (phenotype) can be measured directly in red blood cell membranes. Alternatively (e.g., in patients who have undergone blood transfusions), genetic polymorphisms can be identified by polymerase chain reaction. Guidelines for TPMT activity testing vary among different specialties, but it is generally recommended to test TPMT status before starting azathioprine therapy. Whether testing is cost-effective and clinically useful when compared with traditional monitoring of white blood counts is still a matter of debate.\textsuperscript{68-71} In the absence of TPMT testing, a low initial dose and careful monitoring of the white blood cell count in patients starting azathioprine is required. Some authors
suggest weekly monitoring during the first 15 weeks of azathioprine treatment.64 When patients are on a stable dose of azathioprine, blood counts are monitored monthly and liver function tests are monitored every 3 to 4 months.

**Drug Interactions**

One of the most important, and potentially fatal, drug interactions in rheumatology is the ability of allopurinol, through inhibition of xanthine oxidase-mediated inactivation of mercaptopurine, to increase dramatically the cytotoxic effects of azathioprine and mercaptopurine.72 Various strategies have been employed to treat hyperuricemia and gout in patients receiving azathioprine, a common clinical problem after transplantation. Reduction of the dose of azathioprine by at least two thirds in patients who also are receiving allopurinol is advocated. Because myelosuppression can still occur after a 75% reduction in dose, however, careful monitoring is required.72 Alternatively, uricosurics such as benzbromarone have been effective and safe,72 and MMF has been substituted for azathioprine as an alternative immunosuppressant.73 Combination of azathioprine with several other drugs may also increase the risk of myelosuppression: sulfasalazine, ganciclovir, angiotensin-converting enzyme (ACE) inhibitors, carbamazepine, co-trimoxazole, and clozapine. Azathioprine has been associated with resistance to warfarin in case reports.74

**Pregnancy and Lactation**

Azathioprine is an FDA Pregnancy Category D drug. Azathioprine and mercaptopurine cross the placenta, but drug and metabolite concentrations are lower in the fetal circulation, suggesting placental metabolism.38,75 There are limited data in rheumatologic diseases, and although it is being used in pregnancy, azathioprine is better avoided in pregnancy and lactation if possible.76 In a prospective observational study in 189 pregnant women treated with azathioprine for various autoimmune conditions, the rate of major malformation was 3.5%, which was similar to the general population.77 Its use was associated with prematurity, however. Azathioprine may be considered in cases where the benefits of disease control in the mother give the best chances of term pregnancy and fetal survival.

**Cyclosporine, Tacrolimus (FK506), and Sirolimus (Rapamycin)**

**Cyclosporine**

**Structure**

Cyclosporine is a lipophilic endecapeptide derived from a fungus; it is effective in refractory RA, psoriatic arthritis, systemic lupus erythematosus, and autoimmune eye disease.

**Mechanism of Action**

Cyclosporine impairs production of interleukin-2 and other cytokines, reducing lymphocyte proliferation. Cyclosporine complexes with cyclophilin, one of a group of cytosolic-binding proteins known as immunophils. This complex binds to and inhibits calcineurin, a serine/threonine phosphatase. Inhibition of calcineurin phosphatase activity prevents the translocation of cytosolic nuclear factor of activated T cells to the nucleus, a translocation that is required for the transcription of genes for cytokines such as interleukin-2 and for T cell activation (Figure 62-3).78,79

**Pharmacology**

There are two formulations of cyclosporine for oral use: (1) the older, oil-based Sandimmune formulation and (2) the newer, microemulsion Neoral formulation. The active drug, cyclosporine, is identical in both formulations, but the microemulsion formulation results in better bioavailability and less intersubject and intrasubject variability. Neoral or equivalent generic formulations are replacing the Sandimmune formulation of cyclosporine, which is no longer universally available.

**Absorption and Distribution.** Cyclosporine is poorly and variably absorbed from the gut with a bioavailability of approximately 30%. A high-fat meal increases absorption. There is substantial interindividual (threefold) and intraindividual (twofold) variability in cyclosporine disposition.80 The time to peak concentration (1 to 8 hours) and the elimination half-life (3 to 20 hours) vary. Cyclosporine is lipophilic and widely distributed in body tissues, particularly in the lean body mass.80 Higher concentrations of cyclosporine are found in red blood cells than in plasma, and whole-blood concentrations are used for monitoring, although this is seldom required in autoimmune diseases. The time to maximal concentration is approximately 25% shorter, and the maximal cyclosporine concentration and area under the concentration curve are increased by approximately 50% with the microemulsion formulation.81,82 Also, interindividual and intraindividual variability in cyclosporine disposition is decreased by approximately 50% with the microemulsion formulation.81,82

**Metabolism and Elimination.** There are two major determinants of cyclosporine disposition. First, P-glycoprotein (Pgp), a drug efflux pump, pumps substrates such as cyclosporine out of cells. Pgp, the product of the multidrug resistance gene, is expressed on intestinal epithelial cells and in the liver. Second, cyclosporine is extensively metabolized by the CYP3A enzyme system, which is active not only in the liver but also in intestinal epithelium. Pgp, by limiting drug uptake, and CYP3A4, by assisting drug metabolism in the gut and liver, act to limit the bioavailability of cyclosporine and to determine its disposition.83 Cyclosporine is extensively metabolized to more than 20 metabolites. The rate-limiting step in the elimination of cyclosporine is the formation of metabolites, not their clearance. Cyclosporine elimination is not altered in renal failure; however, because of its nephrotoxicity, cyclosporine is avoided in patients with impaired renal function. Liver disease impairs the excretion of cyclosporine metabolites.

**Dosage**

Effective use of cyclosporine requires that appropriate patients be selected for treatment and monitored carefully (Table 62-3). The starting dosage of cyclosporine is 2.5 mg/
kg/day, usually administered in divided doses. In obese patients, dosage is based on the approximate ideal body weight. Clinical response is slow, occurring over 4 to 8 weeks, and may be only maximal after 12 weeks or more of treatment. To improve efficacy, the dosage can be increased by 0.5 mg/kg/day at 4- to 8-week intervals to a maximal dosage of 4 mg/kg/day of the microemulsion formulation. If there is no clinical response in 4 to 6 months, cyclosporine should be discontinued. In patients who are well controlled, the dosage of cyclosporine can be decreased by 0.5 mg/kg/day at 4- to 8-week intervals to determine the minimal effective dose for the individual patient. In patients receiving the older Sandimmune formulation of cyclosporine who convert to the microemulsion formulation, a 1:1 dose conversion is generally used. Because of the greater and more predictable bioavailability of the microemulsion formulation, however, a greater exposure to cyclosporine is likely. Blood pressure and creatinine should be monitored initially at 2-week intervals after the conversion, and the dose of cyclosporine should be decreased if required.

**Clinical Indications**

Cyclosporine is effective in the treatment of RA as a single agent and in combination with methotrexate or hydroxychloroquine, but it is used less commonly now because of the availability of more effective and safer treatment options in early RA. Nevertheless, it remains a useful drug in refractory RA. Cyclosporine has been shown to increase mean peak plasma methotrexate levels and area under the curve by about 20%, this may contribute to the efficacy of the combination. Data comparing the efficacy and safety of cyclosporine with other disease-modifying antirheumatic

---

**Table 62-3 Clinical Use of Cyclosporine in Rheumatic Disease**

<table>
<thead>
<tr>
<th>Select appropriate patients</th>
<th>Contraindications:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current or past malignancy other than basal cell carcinoma, renal impairment, uncontrolled hypertension, hepatic dysfunction</td>
<td>Elderly age, obesity, controlled hypertension, premalignant lesions, drugs that interact with cyclosporine, pregnancy</td>
</tr>
<tr>
<td>Cautions:</td>
<td>Obtain ≥2 creatinine concentrations before starting cyclosporine, and average these to provide baseline creatinine value</td>
</tr>
<tr>
<td></td>
<td>Start low — cyclosporine 2.5 mg/kg/day in divided doses</td>
</tr>
<tr>
<td></td>
<td>Stay low — maximum 4 mg/kg/day (microemulsion formulation)</td>
</tr>
<tr>
<td></td>
<td>Monitor blood pressure and creatinine initially every 2 wk for 3 mo, then monthly if stable</td>
</tr>
<tr>
<td></td>
<td>If serum creatinine increases &gt;30% above patient’s baseline, reduce dosage of cyclosporine by 1 mg/kg/day; recheck serum creatinine in 1-2 wk, and temporarily discontinue cyclosporine if creatinine remains &gt;30% above baseline</td>
</tr>
<tr>
<td></td>
<td>When creatinine level returns to within 15% of baseline, cyclosporine can be restarted at a lower dosage</td>
</tr>
</tbody>
</table>
drugs over long periods and in a large number of patients are limited. Cyclosporine is effective for the skin and joint manifestations of psoriasis. Less data are available regarding the use of cyclosporine in other rheumatic diseases. In uncontrolled, small studies in SLE, cyclosporine has been reported to improve disease activity; have a glucocorticoid-sparing effect; and improve proteinuria, thrombocytopenia, and leukopenia. The efficacy of cyclosporine as a steroid-sparing drug was confirmed in a randomized clinical trial in patients with severe SLE, but it was not found to be more effective or safer when compared with azathioprine. Cyclosporine has also been reported to be effective in small series of cases in many other autoimmune conditions including pyoderma gangrenosum, Behçet’s disease, maintenance therapy of antineutrophil cytoplasmic antibody-associated vasculitis, and macrophage-activation syndrome in juvenile RA.

Toxicity

Hypertension. Hypertension occurs in approximately 20% of patients with autoimmune disease receiving cyclosporine. The magnitude of increase in blood pressure is usually mild but clinically significant as it increases the risk of stroke, myocardial infarction, heart failure, and other adverse cardiovascular events associated with elevated BP. The hypertension should be controlled by reducing the dose of cyclosporine or by antihypertensive drug therapy.

Nephrotoxicity. Virtually all patients who take cyclosporine have a small but measurable decrease in renal function that is reversible after cyclosporine is discontinued. Serum creatinine concentrations have increased approximately 20% in 6- to 12-month clinical trials, but few patients have had to withdraw because of this. Long-term data regarding renal function in RA patients treated with cyclosporine are limited. In one 12-month study, an increase in serum creatinine of more than 30% occurred in 50% of patients; half of these patients responded to cyclosporine dose reduction, and half did not, requiring discontinuation of the drug. The small increase in serum creatinine observed in most studies occurs mainly during the first 2 to 3 months of treatment, and then creatinine remains relatively stable over 12 months. Other data suggest, however, that over periods of treatment longer than 1 year, many patients, who over the first year had a stable, acceptable increase in creatinine concentration, subsequently have an increase in creatinine to more than 30% of baseline that is not controlled by cyclosporine dose reduction; such patients have to discontinue treatment. Preventable risk factors for cyclosporine-induced nephrotoxicity are a high dosage of cyclosporine (>5 mg/kg/day) and an increase in serum creatinine concentration of more than 50% of the baseline value. The risk of cyclosporine nephropathy is low in patients treated according to the clinical guidelines (see Table 62-3). Renal biopsy specimens in 11 patients with RA who received cyclosporine (average dosage, 3.3 mg/kg/day) for 26 months and had an average increase in serum creatinine of 31% showed no significant cyclosporine-induced renal changes.

Gastrointestinal. Gastrointestinal upset is common but usually mild and transient. A few patients discontinue cyclosporine therapy for this reason, however.

Malignancy. In transplant recipients, cyclosporine use has been associated with an increased risk of skin cancer and lymphoma. In 208 patients with RA treated with cyclosporine for an average of 1.6 years, the incidence of malignancy and mortality was similar to that of RA controls, but a recent meta-analysis on the risk of immunomodulatory drugs in RA, psoriasis, and psoriatic arthritis did find an increased risk of nonmelanoma skin cancer in patients treated with cyclosporine. Epstein-Barr virus–induced B cell lymphoma, which may be reversible when cyclosporine is discontinued, has been reported in a few patients receiving cyclosporine for a variety of indications.

Others. Other adverse effects that are common but usually of minor significance include hypertrichosis, gingival hyperplasia, tremor, paresthesia, breast tenderness, hyperkalemia, hypomagnesemia, and increase in serum uric acid. Cyclosporine may result in a clinically insignificant increase in alkaline phosphatase concentrations but does not increase the frequency of abnormal transaminase concentrations in patients also receiving methotrexate.

Strategies to Minimize Toxicity. Because cyclosporine may increase liver enzymes, potassium, uric acid, and lipid concentrations and decrease magnesium concentrations, it is prudent to measure these before, and occasionally after, initiating therapy. At least two, and preferably more, recent normal blood pressure and serum creatinine determinations should be obtained before starting treatment. Many patients with RA have low serum creatinine concentrations, and it is important not to overlook significant cyclosporine-induced elevations in serum creatinine, which may remain within the normal laboratory reference range. If a patient has a baseline creatinine level of 0.6 mg/dL that after cyclosporine increases to 0.9 mg/dL (still in the normal range), this represents a 50% increase above baseline and requires dose reduction. Cyclosporine concentrations are not useful predictors of efficacy or toxicity in rheumatic diseases and are not routinely performed. Cyclosporine trough concentrations, measured approximately 12 hours after the last dose, can be useful if there are concerns about compliance or unusual drug disposition in individual patients.

Pregnancy and Lactation

Cyclosporine is an FDA Pregnancy Category C drug. Pregnancy outcomes in transplant recipients receiving cyclosporine-based and noncyclosporine-based regimens are similar. Cyclosporine use in pregnancy is not recommended, however, unless the potential benefit exceeds the potential risk to the fetus. Breastfeeding should be avoided.

Drug Interactions

Cyclosporine and tacrolimus, because of the influence of Pgp and CYP3A4 enzyme activity on their disposition, have many clinically important drug interactions (Table 62-4). Many drugs such as erythromycin, azole antifungal drugs, and some calcium channel antagonists that inhibit CYP3A4 (inhibiting the metabolism of cyclosporine) also inhibit Pgp. Drug interactions mediated by these dual mechanisms may result in a twofold to fivefold increase in
Tacrolimus (FK506)

Structure

Tacrolimus, previously known as FK506, is a macrolide derived from an actinomycete and is widely used in organ transplantation as an alternative to cyclosporine. Studies in autoimmune disease are less advanced.

Mechanisms of Action

Tacrolimus is about 100 times more potent than cyclosporine and, although structurally different, is also a calcineurin inhibitor. Tacrolimus binds to an intracellular binding protein (FK binding protein), and this immunophilin complex, in association with calcineurin, suppresses transcription of cytokines such as interleukin-2, inhibiting the early steps of T lymphocyte activation (see Figure 62-3).19

Pharmacology

Absorption of tacrolimus after oral administration is poor and highly variable (range, 4% to 93%; average, 25%).108 Tacrolimus is lipophilic, is widely distributed in tissues, and is almost completely metabolized with an elimination half-life of 5 to 16 hours.109 As with cyclosporine, Pgp and CYP3A4 in liver and gut are important determinants of the metabolism and disposition of tacrolimus. Drugs that inhibit CYP3A4 or Pgp can increase tacrolimus concentrations (see Table 62-4).108 Impaired hepatic function, but not impaired renal function, increases tacrolimus concentrations.109

Dosage

Tacrolimus is not routinely used for rheumatologic conditions. Different dosages have been used in clinical studies (see next paragraph).

Clinical Indications

Tacrolimus is effective in animal models of arthritis, but data in humans are limited.96,100,110 In a 6-month phase II, randomized, double-blind, placebo-controlled monotherapy study, patients with RA received 2 mg or 3 mg of tacrolimus or placebo daily for 24 weeks. An American College of Rheumatology 20% improvement criteria (ACR20) response was observed in 10.2% of patients receiving placebo and in 18.8% and 26.8% of patients receiving 2 mg and 3 mg of tacrolimus.111 A longer-term study of 3 mg of tacrolimus in 896 rheumatoid patients with a median duration of treatment of 359 days yielded ACR20, ACR50, and ACR70 responses of 38.4%, 18.6%, and 9%,112 Topical 1% tacrolimus has been used with moderate success in patients with resistant skin disease secondary to SLE, subacute cutaneous lupus erythematosus, and discoid lupus erythematosus.113 Tacrolimus 0.1 mg/kg/day was effective in seven out of nine patients with diffuse proliferative lupus nephritis refractory to IV cyclophosphamide,114 but large-scale placebo-controlled trials have not yet been performed in SLE.

Toxicity

The adverse effects of tacrolimus are dose related and include nephrotoxicity, hypertension, hyperkalemia, hyperuricemia, tremor, hyperglycemia, and gastrointestinal intolerance.108 Among RA patients taking 3 mg of tacrolimus daily for more than 1 year, 59% experienced a side effect, probably or possibly due to the drug including diarrhea.

Table 62-4 Clinically Important Drug Interactions with Cyclosporine*

<table>
<thead>
<tr>
<th>Increased Cyclosporine Concentrations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erythromycin, clarithromycin</td>
</tr>
<tr>
<td>Azole antifungals: ketoconazole, fluconazole, itraconazole</td>
</tr>
<tr>
<td>Calcium channel antagonists: diltiazem, verapamil, amiodipine†</td>
</tr>
<tr>
<td>Grapefruit juice</td>
</tr>
<tr>
<td>Others: amiodarone, danazol, allopurinol, colchicine</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Decreased Cyclosporine Concentrations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inducers of hepatic enzymes: rifampicin, phenytoin, phenobarbital, nafellin, St John’s wort</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Increased Cyclosporine Toxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased renal toxicity with aminoglycosides, quinolone antibiotics, amphotericin B, (?) nonsteroidal anti-inflammatory drugs, (?) angiotensin-converting enzyme inhibitors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cyclosporine Increasing Toxicity of Another Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased risk of myopathy and rhabdomyolysis with lovastatin and other statins</td>
</tr>
<tr>
<td>Increased risk of colchicine neuromyopathy and toxicity</td>
</tr>
<tr>
<td>Increased digoxin concentrations</td>
</tr>
<tr>
<td>Increased risk of hyperkalemia with K+ -sparking diuretics and K+ supplements</td>
</tr>
</tbody>
</table>

*Most interactions with cyclosporine also likely apply to tacrolimus.
†There are conflicting data that amiodipine does and does not increase cyclosporine concentrations.

Tacrolimus is lipophilic, is widely distributed in tissues, and highly variable (range, 4% to 93%; average, 25%). Tacrolimus is about 100 times more potent than cyclosporine and, although structurally different, is also a calcineurin inhibitor. Tacrolimus binds to an intracellular binding protein (FK binding protein), and this immunophilin complex, in association with calcineurin, suppresses transcription of cytokines such as interleukin-2, inhibiting the early steps of T lymphocyte activation (see Figure 62-3).19

Pharmacology

Absorption of tacrolimus after oral administration is poor and highly variable (range, 4% to 93%; average, 25%). Tacrolimus is lipophilic, is widely distributed in tissues, and is almost completely metabolized with an elimination half-life of 5 to 16 hours. As with cyclosporine, Pgp and CYP3A4 in liver and gut are important determinants of the metabolism and disposition of tacrolimus. Drugs that inhibit CYP3A4 or Pgp can increase tacrolimus concentrations (see Table 62-4). Impaired hepatic function, but not impaired renal function, increases tacrolimus concentrations.

Dosage

Tacrolimus is not routinely used for rheumatologic conditions. Different dosages have been used in clinical studies (see next paragraph).

Clinical Indications

Tacrolimus is effective in animal models of arthritis, but data in humans are limited. In a 6-month phase II, randomized, double-blind, placebo-controlled monotherapy study, patients with RA received 2 mg or 3 mg of tacrolimus or placebo daily for 24 weeks. An American College of Rheumatology 20% improvement criteria (ACR20) response was observed in 10.2% of patients receiving placebo and in 18.8% and 26.8% of patients receiving 2 mg and 3 mg of tacrolimus. A longer-term study of 3 mg of tacrolimus in 896 rheumatoid patients with a median duration of treatment of 359 days yielded ACR20, ACR50, and ACR70 responses of 38.4%, 18.6%, and 9%. Topical 1% tacrolimus has been used with moderate success in patients with resistant skin disease secondary to SLE, subacute cutaneous lupus erythematosus, and discoid lupus erythematosus. Tacrolimus 0.1 mg/kg/day was effective in seven out of nine patients with diffuse proliferative lupus nephritis refractory to IV cyclophosphamide, but large-scale placebo-controlled trials have not yet been performed in SLE.

Toxicity

The adverse effects of tacrolimus are dose related and include nephrotoxicity, hypertension, hyperkalemia, hyperuricemia, tremor, hyperglycemia, and gastrointestinal intolerance. Among RA patients taking 3 mg of tacrolimus daily for more than 1 year, 59% experienced a side effect, probably or possibly due to the drug including diarrhea.

Table 62-4 Clinically Important Drug Interactions with Cyclosporine*

<table>
<thead>
<tr>
<th>Increased Cyclosporine Concentrations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erythromycin, clarithromycin</td>
</tr>
<tr>
<td>Azole antifungals: ketoconazole, fluconazole, itraconazole</td>
</tr>
<tr>
<td>Calcium channel antagonists: diltiazem, verapamil, amiodipine†</td>
</tr>
<tr>
<td>Grapefruit juice</td>
</tr>
<tr>
<td>Others: amiodarone, danazol, allopurinol, colchicine</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Decreased Cyclosporine Concentrations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inducers of hepatic enzymes: rifampicin, phenytoin, phenobarbital, nafellin, St John’s wort</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Increased Cyclosporine Toxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased renal toxicity with aminoglycosides, quinolone antibiotics, amphotericin B, (?) nonsteroidal anti-inflammatory drugs, (?) angiotensin-converting enzyme inhibitors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cyclosporine Increasing Toxicity of Another Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased risk of myopathy and rhabdomyolysis with lovastatin and other statins</td>
</tr>
<tr>
<td>Increased risk of colchicine neuromyopathy and toxicity</td>
</tr>
<tr>
<td>Increased digoxin concentrations</td>
</tr>
<tr>
<td>Increased risk of hyperkalemia with K+ -sparking diuretics and K+ supplements</td>
</tr>
</tbody>
</table>

*Most interactions with cyclosporine also likely apply to tacrolimus.
†There are conflicting data that amiodipine does and does not increase cyclosporine concentrations.

Tacrolimus is lipophilic, is widely distributed in tissues, and highly variable (range, 4% to 93%; average, 25%). Tacrolimus is about 100 times more potent than cyclosporine and, although structurally different, is also a calcineurin inhibitor. Tacrolimus binds to an intracellular binding protein (FK binding protein), and this immunophilin complex, in association with calcineurin, suppresses transcription of cytokines such as interleukin-2, inhibiting the early steps of T lymphocyte activation (see Figure 62-3).19

Pharmacology

Absorption of tacrolimus after oral administration is poor and highly variable (range, 4% to 93%; average, 25%). Tacrolimus is lipophilic, is widely distributed in tissues, and is almost completely metabolized with an elimination half-life of 5 to 16 hours. As with cyclosporine, Pgp and CYP3A4 in liver and gut are important determinants of the metabolism and disposition of tacrolimus. Drugs that inhibit CYP3A4 or Pgp can increase tacrolimus concentrations (see Table 62-4). Impaired hepatic function, but not impaired renal function, increases tacrolimus concentrations.

Dosage

Tacrolimus is not routinely used for rheumatologic conditions. Different dosages have been used in clinical studies (see next paragraph).

Clinical Indications

Tacrolimus is effective in animal models of arthritis, but data in humans are limited. In a 6-month phase II, randomized, double-blind, placebo-controlled monotherapy study, patients with RA received 2 mg or 3 mg of tacrolimus or placebo daily for 24 weeks. An American College of Rheumatology 20% improvement criteria (ACR20) response was observed in 10.2% of patients receiving placebo and in 18.8% and 26.8% of patients receiving 2 mg and 3 mg of tacrolimus. A longer-term study of 3 mg of tacrolimus in 896 rheumatoid patients with a median duration of treatment of 359 days yielded ACR20, ACR50, and ACR70 responses of 38.4%, 18.6%, and 9%. Topical 1% tacrolimus has been used with moderate success in patients with resistant skin disease secondary to SLE, subacute cutaneous lupus erythematosus, and discoid lupus erythematosus. Tacrolimus 0.1 mg/kg/day was effective in seven out of nine patients with diffuse proliferative lupus nephritis refractory to IV cyclophosphamide, but large-scale placebo-controlled trials have not yet been performed in SLE.

Toxicity

The adverse effects of tacrolimus are dose related and include nephrotoxicity, hypertension, hyperkalemia, hyperuricemia, tremor, hyperglycemia, and gastrointestinal intolerance. Among RA patients taking 3 mg of tacrolimus daily for more than 1 year, 59% experienced a side effect, probably or possibly due to the drug including diarrhea.
(15%); nausea (10%); tremor (9%); headache (9%); abdominal pain (8%); increased creatinine (7%); hypertension (5%); and pneumonia, pancreatitis, hyperglycemia, and diabetes mellitus (all <1%). A creatinine increase of greater than 30% from baseline to end of study was found in 40%.114

Sirolimus (Rapamycin), Everolimus

Sirolimus, isolated from an actinomycete, and its derivative everolimus have been developed as immunosuppressants for organ transplantation, but the collective experience in rheumatologic conditions is limited. Sirolimus and everolimus, in contrast to cyclosporine and tacrolimus, do not act through calcineurin. They bind to FK binding protein and target proteins variously known as targets of rapamycin (TOR) proteins or FK-rapamycin associated proteins (FRAP), blocking progression of the cell cycle from G1 to S phase by inhibiting several downstream signal transduction pathways.78 Clinical trials have shown efficacy in transplant recipients who also had inflammatory arthritis. A randomized, placebo-controlled, proof-of-concept study with everolimus in 121 RA patients receiving methotrexate (MTX) showed a rapid onset of action with ACR20 response rates that were significantly higher in the everolimus group (36.1%) than in the placebo group (16.7%; P = 0.022) after 12 weeks. Adverse events were more common in the everolimus group (mainly gastrointestinal, skin, neurologic).115 Rapamycin was as effective in reducing skin thickening as methotrexate in a small randomized trial in 18 patients with early diffuse cutaneous systemic sclerosis, but 2 patients treated with rapamycin were withdrawn because of severe hypertriglyceridemia.116

Mycophenolate Mofetil

Structure

MMF, a prodrug, is the inactive 2-morpholinoester of mycophenolic acid (MPA), an antibiotic with immunosuppressive effects.117

Mechanism of Action

There are two pathways for the synthesis of guanine nucleotides: the de novo pathway and the salvage pathway. MPA reversibly inhibits inosine monophosphate dehydrogenase, a crucial enzyme for the de novo synthesis of guanosine nucleotides.117,118 Lymphocytes, in contrast to many other cells, are critically dependent on the de novo purine synthesis pathway and are a relatively selective target for MPA, accounting for the ability of the drug to inhibit reversibly B cell and T cell proliferation without myelotoxicity.119 MPA results in decreased guanine synthesis and decreased DNA synthesis, decreased lymphocyte proliferation, and decreased antibody production.118-122 MPA also inhibits proliferation of fibroblasts, endothelial cells, and arterial smooth muscle cells and prevents deposition and contraction of collagen, extracellular matrix proteins, and smooth muscle actin.121

Toxicity

MMF is generally well tolerated. The most common side effects are gastrointestinal such as diarrhea, nausea, abdominal pain, and vomiting. Occasional infections, leukopenia, lymphocytopenia, and elevated liver enzymes can occur. Of 54 SLE patients treated with MMF over a 3-year period, 16% withdrew because of adverse events, with 73% continuing treatment at 12 months.123 In patients with lupus nephritis, diarrhea was more common and serious infections were less common with MMF than with cyclophosphamide.38 Enteric-coated mycophenolate sodium and MMF have similar rates of side effects.129 Opportunistic infections including one that was fatal occurred in 3 of 10 patients with idiopathic dermatomyositis who were treated with glucocorticoids and MMF.130

Figure 62-4 Mycophenolate mofetil is converted to mycophenolic acid (MPA) and is subsequently metabolized to its glucuronide (Glu) conjugates, MPAG and AcMPAG, by different isoforms of UDP glucuronosyltransferases (UGT) in the liver. Pathways of enterohepatic circulation of MPA via the glucuronide conjugate metabolites are shown. The majority of MPAG is excreted in urine.
**Dosage**

Effective daily dosages of MMF range from 0.5 to 1.5 g twice a day. In 71 patients with lupus nephritis, the initial dose was 1 g/day with a target of 3 g/day. The mean maximal dose was 2680 mg/day, and 63% of patients tolerated 3 g/day.\(^{38}\)

**Clinical Indications**

In recent years MMF has emerged as a potentially safer alternative to cytotoxic agents in the treatment of several rheumatic diseases, notably systemic lupus erythematosus,\(^{10,13}\) systemic sclerosis,\(^{12}\) vasculitis,\(^{13\text{a}}\) and inflammatory muscle disease.\(^{13\text{b},13\text{c}}\) In a 24-week study in lupus nephritis, mycophenolate was more effective than monthly pulse cyclophosphamide with a failure rate (without complete or partial remission at 24 weeks, plus those who stopped treatment for any reason) of 34 of 71 (47.9%) compared with 48 of 69 in the cyclophosphamide group (69.6%; \(P=0.01\)).\(^{38}\) In a systematic review of four trials involving 618 patients, MMF was not superior to cyclophosphamide for renal remission and there was no significant difference for adverse events (infections, leukopenia, gastrointestinal symptoms, herpes zoster, end-stage renal disease, and death) except for a lower incidence of alopecia and amenorrhea with the use of MMF compared with cyclophosphamide.\(^{13\text{d}}\) In seven patients with myositis, six had a good clinical and biochemical response to mycophenolate,\(^{13\text{e}}\) an observation that was confirmed in another study in six patients with refractory myositis,\(^{13\text{f}}\) and in three studies in patients with interstitial lung disease associated with dermatomyositis (\(n=4\)) or other connective tissue diseases including rheumatoid arthritis (\(n=10\)) and systemic sclerosis (\(n=13\)), MMF was effective in improving signs and symptoms of lung disease.\(^{13\text{g},13\text{h}}\) MMF may be useful as an alternative immunosuppressant to azathioprine, particularly in patients with gout who require therapy with allopurinol because, in contrast to azathioprine, it does not seem to interact significantly with allopurinol.\(^{76,119}\) Mycophenolate 1 g twice daily was not more effective than placebo in two clinical trials involving 443 patients with refractory RA as assessed with ACR20 responses, and a larger mycophenolate-cyclosporine trial was stopped prematurely. Treatment-related adverse events were experienced by 51.6%, 73.1%, and 36.1% of patients receiving MMF, cyclosporine, and placebo, respectively. Hypertension, increased serum creatinine, muscle cramps, hirsutism, and hypertrichosis were more than twice as common with cyclosporin as with MMF. In all three trials the incidence of serious adverse events with MMF was 12.1% (compared with 11.3% and 7.5% for cyclosporin and placebo, respectively). Although mycophenolate is used in psoriasis as an effective alternative to methotrexate, there are only anecdotal reports of its utility in psoriatic arthritis.

**Pregnancy and Lactation**

MMF is an FDA Pregnancy Category C drug. Mycophenolic acid is associated with miscarriage and congenital malformations when used during pregnancy and should therefore be avoided whenever possible by women trying to conceive. It is transferred into the mother’s milk, and extreme caution should be used in women with childbearing potential and lactating mothers.

**Drug Interactions**

Because MPA is glucuronidated and not metabolized by CYP oxidation, there are few clinically significant drug interactions. Antacids reduce bioavailability by approximately 15%, and cholestyramine reduces bioavailability by approximately 40%.\(^{141}\) Rifampin treatment reduced MPA concentrations twofold to threefold.\(^{142}\) Co-administration with azathioprine is not recommended.

**THALIDOMIDE**

**Structure**

Thalidomide is a racemic glutamic acid analogue that was introduced in the 1950s as a sedative and antiemetic. The recognition that thalidomide was a potent teratogen resulting in characteristic congenital malformations led to its withdrawal in 1961. The rediscovery of the immunomodulating effects of thalidomide has led to the cautious and closely regulated, but controversial, reintroduction of thalidomide for the treatment of erythema nodosum leprosum. Preliminary studies have explored other potential therapeutic roles.

**Mechanism of Action**

Multiple mechanisms have been proposed for the immunosuppressive effects of thalidomide, the most plausible being the inhibition of angiogenesis and the inhibition of tumor necrosis factor production.\(^{143,144}\)

**Pharmacology**

Peak concentrations of thalidomide occur 2 to 4 hours after oral administration. The elimination half-life is approximately 5 hours, with elimination being virtually entirely through nonenzymatic hydrolysis.\(^{145}\) CYP2C19, a polymorphic enzyme, contributes to the formation of an active metabolite, 5-hydroxythalidomide. The pharmacokinetics of thalidomide are poorly characterized, and there is little information regarding drug interactions or use in patients with impaired renal or hepatic function. The sedative effects of other central depressants such as barbiturates are enhanced by thalidomide.

**Dosage**

Thalidomide is approved by the FDA only for the treatment of erythema nodosum leprosum. Prescription of thalidomide for this indication and its off-label use is closely regulated. In a randomized, controlled trial, thalidomide (100 mg/day and 300 mg/day) for 24 weeks improved the mucocutaneous lesions of Behçet’s syndrome. Clinical response was lost rapidly, however, after discontinuation of the drug.\(^{146}\) Small, largely uncontrolled reports suggest possible benefit in the skin manifestations of lupus, sarcoidosis, RA, Sjögren’s syndrome, ankylosing spondylitis, systemic-onset juvenile RA, and pyoderma gangrenosum.\(^{147}\)
Toxicity

The most serious, best-known, and preventable adverse effect of thalidomide is its ability to cause birth defects. In clinical studies, peripheral neuropathy has been the most common serious adverse effect, usually manifesting as painful, symmetric paresthesias. Electrophysiologic changes precede clinical neuropathy, and most studies reporting higher rates of neuropathy have used this diagnostic technique. The neuropathy may become evident only after thalidomide has been discontinued and in most patients (75%) may not resolve completely. Other common adverse effects include sedation, skin rash, limb edema, and constipation. Neutropenia is less common. Ovarian failure and arterial and venous thromboses have been reported.

Strategies to Minimize Toxicity

Strategies to prevent fetal exposure to thalidomide are outlined in the System for Thalidomide Education and Prescribing Safety (STEPS) program developed by the drug’s manufacturer, Celgene, and in guidelines developed in the United Kingdom. Registration in the STEPS program is required before thalidomide is used. The program involves mandatory patient registration, education, surveys, and contraception for men and women. Electrophysiologic monitoring for thalidomide-induced neuropathy should be considered if long-term therapy is planned. Thalidomide should be discontinued if peripheral neuropathy occurs. Thalidomide’s side effects, its teratogenic effects, and the rapid relapse of autoimmune disease after discontinuation of thalidomide severely limit its therapeutic potential.

CONCLUSION

Immunosuppressive drugs are key therapeutic tools in the management of many rheumatic diseases. They include alkylating agents such as cyclophosphamide and purine analogue cytotoxic drugs such as azathioprine with a long history of clinical use in rheumatology and relatively newer noncytotoxic immunosuppressants such as MMF. In contrast to extracellular, exquisitely targeted therapeutics represented by biologics, our understanding of the in vivo mechanism of action of immunosuppressive drugs is limited. In contrast, their potential efficacy and safety profiles are generally well known and serious toxicities can usually be prevented by careful monitoring of laboratory tests for white blood counts, liver and renal function, and electrolytes. As a general rule, combination therapy of the different immunosuppressants discussed earlier should be avoided. The individual response to immunosuppressive therapy can be highly variable, and decisions to continue a chosen immunosuppressant should be revisited on a regular basis weighing the benefits and side effects.

Selected References


Full references for this chapter can be found on www.expertconsult.com.
References


In recent years, discoveries delineating the immunopathophysiologic basis of various rheumatic diseases, combined with biopharmaceutical development, have allowed the introduction of biologic therapeutics. These agents target specific components of the immune response that are dysregulated and are thought to be central to the cause and sustenance of the disease process. In the rheumatoid synovium, for example, substantial evidence has been found of upregulation of key proinflammatory cytokines such as tumor necrosis factor (TNF), interleukin-1 (IL-1), IL-6, and others. Agents targeting these key mediators, in particular TNF, have considerable efficacy in the treatment of patients with rheumatoid arthritis (RA) and other systemic inflammatory disorders. The ability of TNF inhibitors not only to improve the signs and symptoms of disease but also to preserve functional status and quality of life and to inhibit disease progression has altered both physicians’ and patients’ expectations regarding antirheumatic treatment. Moreover, their success has driven research into the targeting of other cytokines relevant to the pathogenesis of autoimmune disorders. In this chapter, we focus on therapeutic agents that target TNF, IL-1, and IL-6.

**TUMOR NECROSIS FACTOR INHIBITORS**

TNF plays a central role in the pathogenesis of RA and other inflammatory disorders. Although it can be produced by numerous cell types, in inflammatory conditions such as RA, TNF is produced largely by activated macrophages. Human TNF is synthesized and expressed as a 26-kD transmembrane protein on the plasma membrane and is cleaved by a specific metalloproteinase (TNF-converting enzyme). After proteolytic cleavage, TNF is converted to a 17-kD soluble protein, which oligomerizes to form the active homotrimer. The actions of TNF are mediated through two structurally distinct receptors: TNF-RI (55 kD; CD120a) and TNF-RII (75 kD; CD120b). The two receptors differ in their binding affinities signaling properties, and primary functions. The binding of TNF to its receptor can initiate several signaling pathways. Signaling cascades include the activation of transcription factors (e.g., nuclear factor κB [NFκB]), protein kinases (intracellular enzymes that mediate cellular responses to inflammatory stimuli, such as c-JunN-terminal kinase [JNK] and p38 mitogen-activated protein [MAP] kinase), and proteases (enzymes that cleave peptide bonds, such as caspases).

TNF may contribute to the pathogenesis of RA through myriad mechanisms, including induction of other proinflammatory cytokines (e.g., IL-1, IL-6) and chemokines (e.g., IL-8); enhancement of leukocyte migration by increasing endothelial layer permeability and adhesion molecule expression and function; activation of numerous cell types; and induction of the synthesis of acute phase reactants and other proteins, including tissue-degrading enzymes (matrix metalloproteinase enzymes) produced by synoviocytes or chondrocytes. The pivotal role of TNF in mediating such diverse inflammatory activities provided the rationale for targeting this cytokine in systemic inflammatory diseases.

Initially, animal studies proved that inhibition of TNF with monoclonal antibodies or soluble TNF-R constructs ameliorated the signs of inflammation and prevented joint destruction. Subsequently, studies in humans confirmed the substantial efficacy of these compounds.
Currently, five anti-TNF agents are available for clinical use: infliximab, a chimeric anti-TNF monoclonal antibody; etanercept, a soluble dimeric p75-TNF-R/Fc fusion construct; adalimumab, a human anti-TNF monoclonal antibody; golimumab, a human anti-TNF monoclonal antibody; and certolizumab pegol, a Fab fragment of a recombinant, humanized anti-TNF monoclonal antibody linked to a 40-kD polyethylene glycol (PEG) moiety. For each agent, initial assessment in open-label studies of patients with RA was followed by double-blind, placebo-controlled, randomized clinical trials. Typically, early studies included patients with very active disease that was relatively chronic and refractory. Driven by success in the most difficult populations, studies of all TNF inhibitors have also been performed in patients with early RA. Most studies included patients whose disease remained active despite concurrent use of methotrexate (MTX); some studies assessed the efficacy of a drug as monotherapy. Building on the efficacy achieved in RA, TNF inhibitors have been tested in other inflammatory arthritides, including psoriatic arthritis and ankylosing spondylitis. Moreover, these agents have been studied in other autoimmune conditions, such as Crohn’s disease, ulcerative colitis, psoriasis, uveitis, and others.

Although all five available agents are macromolecule TNF inhibitors, differences among them have been noted. The monoclonal antibodies infliximab, adalimumab, golimumab, and certolizumab pegol are specific for TNF, whereas etanercept binds both TNF and lymphotoxin-α (LT-α; previously referred to as lymphotoxin). Given intravenously, infliximab has a high peak concentration followed by steady-state elimination, whereas etanercept, adalimumab, golimumab, and certolizumab pegol, because they are given subcutaneously, have “flatter” pharmacokinetic profiles. With the exception of certolizumab pegol, these agents are capable of affecting Fc-mediated functions, such as complement-dependent cytosis and antibody-dependent cell-mediated cytotoxicity, and all bind to both soluble and membrane forms of TNF, although some relative differences in affinity may be noted. Other differences, such as effects on cytokine secretion, have been observed in some in vitro studies. Regarding apoptosis, the data have been somewhat discrepant. In patients with RA, both the anti-TNF monoclonal antibody infliximab and the soluble receptor construct etanercept are capable of inducing apoptosis in synovial macrophages. However, in patients with Crohn’s disease, etanercept was not clinically effective at the doses studied and did not induce apoptosis. In contrast, the anti-TNF monoclonal antibodies infliximab and adalimumab were clinically effective and induced apoptosis in highly activated lymphocytes. However, certolizumab pegol is effective in Crohn’s disease and is not able to induce apoptosis. The extent to which these potential differences among TNF inhibitors correlate with any specific aspects of efficacy or toxicity remains to be established.

**Infliximab**

**Structure**

Infliximab is a chimeric mouse-human monoclonal antibody composed of constant regions of human immunoglobulin (Ig) G1κ coupled to the variable regions of a high-affinity neutralizing murine anti-human TNF antibody. The resulting construct is approximately 70% human (Figure 63-1).

**Pharmacokinetics**

Clinical pharmacology studies demonstrate that infliximab has a dose-dependent pharmacokinetic profile following infusions of 1 to 20 mg/kg. In combination therapy with MTX (7.5 mg once a week), serum infliximab concentrations tend to be slightly higher than when administered alone. Infliximab behaves in a consistent manner across different demographic groups (including pediatric vs. adult patients) and among patients with different diseases of varied severity. The half-life of infliximab is around 8 to 9.5 days at the 3 mg/kg dose, although longer values have been reported for higher doses. The volume distribution of infliximab at steady state (3 to 5 L) is independent of dose, suggesting a predominantly intravascular distribution. Concomitant use of MTX results in an increase in the area under the curve of infliximab of approximately 25% to 30%.

**Drug Dose**

The typical initial dose of infliximab in RA is 3 mg/kg given as an intravenous (IV) infusion in combination with MTX, followed by doses 2 and 6 weeks after the first infusion, then every 8 weeks thereafter. Some RA patients have received infliximab in combination with disease-modifying antirheumatic drugs (DMARDs) other than MTX or as monotherapy. For patients who have an incomplete response, dosing may be increased up to 10 mg/kg, or the drug may be administered as often as every 4 weeks. In the clinic, increasing the dose of infliximab or decreasing the interval of administration is not an uncommon practice; however, it is not clear to what extent such changes achieve clinical improvements. For patients with psoriatic arthritis and ankylosing spondylitis, the recommended dose is 5 mg/kg, with or without MTX, at 0, 2, and 6 weeks, then every 8 weeks.

**Efficacy**

**Rheumatoid Arthritis.** In the earliest controlled trials, the efficacy of single doses of 1, 5, 10, and 20 mg/kg of infliximab was demonstrated; however, disease activity recurred when therapy was discontinued. This, along with the growing safety record, provided the rationale for studies with longer durations of therapy. In a subsequent study, concurrent therapy with MTX, even at a relatively low dose of 7.5 mg/wk, seemed to enhance the clinical response to infliximab and to decrease its immunogenicity. Almost all subsequent studies in RA have used such combination therapy.

Multicenter double-blind, placebo-controlled, randomized clinical trials have evaluated the effects of multiple doses of infliximab over longer periods. In the Anti-TNF Trial in Rheumatoid Arthritis with Concomitant Therapy (ATTRACT) trial, the addition of infliximab in patients with long-standing, refractory, active disease was significantly superior to treatment with MTX alone. The results were promising: Substantial improvement in signs and symptoms of disease was noted soon after treatment and was sustained though 54 weeks of follow-up. In addition to
achieving substantial efficacy, the use of infliximab was associated with significant improvement in functional status and quality of life.\textsuperscript{16} Perhaps most remarkably, patients receiving infliximab had a dramatic reduction in the progression of joint damage as assessed by radiographic change scores. The median change in the Sharp score at 1 year for infliximab-treated patients was 0.0 unit (mean change, +0.55; baseline score, 50.5), indicating no significant progression. The median change in score for patients on MTX alone was +4.0 units (mean change, +7.0; baseline score, 55.5); this amount of progression is roughly what would have been predicted given the disease severity.\textsuperscript{14,16}

Following the success achieved in patients with longstanding RA, this therapy was tested in patients with early RA (<3 years’ duration). In the ASPIRE trial, significant improvement in disease activity was noted in patients treated with infliximab plus MTX in comparison with MTX alone, at 54 weeks. In addition, although a significant increase in radiographic destruction was observed in patients treated only with MTX, a reduction in disease progression was observed in the infliximab plus MTX groups.\textsuperscript{17} A detailed subanalysis of data from the ATTRACT study demonstrated that a significant radiographic benefit was achieved with infliximab plus MTX treatment even among patients who experienced no improvement in signs and symptoms of disease. This suggests that there may be an uncoupling among various outcomes, and that treatment with TNF inhibitors might lead to disease-modifying activity even in the absence of a clinical response.\textsuperscript{18}

In a safety-based study, a group of patients with active RA and disease more typical of clinic populations (i.e., patients with comorbidities were allowed to enroll) were treated with 3 or 10 mg/kg of infliximab for 46 weeks, resulting in comparable efficacy to that found in earlier trials.\textsuperscript{19} However, an increase in serious infectious events was reported in the 10 mg/kg group.

Psoriatic Arthritis (PsA). In the Infliximab Multinational Psoriatic Arthritis Controlled Trial 1 (IMPACT 1), patients with or without background DMARDs were randomized to receive infliximab or placebo for 16 weeks; thereafter, all patients received infliximab until week 50. The response to infliximab therapy defined by American College of Rheumatology 20% response criteria (ACR20) and PsA response criteria was evident as early as week 2, and improvement continued through week 50. Significant improvements were also seen in skin psoriasis and in dactylitis and enthesitis assessments.\textsuperscript{20,21} In the same group of patients, infliximab inhibited radiographic progression as early as 24 weeks. Thus, the change in the van der Heijde–Sharp score was $-0.70 \pm 2.53$ versus $0.82 \pm 2.62$ at week 24 for infliximab- and placebo-treated patients, respectively. At week 54, these scores continued to improve: $-0.94$ versus 0.53 for infliximab and placebo crossover patients, respectively.\textsuperscript{22}

In the same study, the effect of infliximab on health-related quality of life was assessed. The mean percentage improvement from baseline in the health assessment questionnaire (HAQ) was 48.6% in the infliximab group,
compared with a worsening of 18.4% in the placebo group at week 14. Overall, 58.6% of the infliximab group and 19.4% of the placebo group achieved a clinically meaningful improvement in HAQ at week 14.24

Ankylosing Spondylitis. The efficacy and safety of infliximab have been assessed in several multicenter clinical trials.24-27 In the initial double-blind, placebo-controlled, randomized part of one early study, disease activity measured by the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), functional ability measured by the Bath Ankylosing Spondylitis Functional Index (BASFI), and mobility of the spine measured by the Bath Ankylosing Spondylitis Metrology Index (BASMI) significantly improved with 5 mg/kg infliximab given at weeks 0, 2, 6, and 12 compared with placebo.25 Of patients who continued to receive infliximab every 6 weeks until week 102, a significantly higher percentage achieved 50% or greater improvement in the BASDAI.26,27 At week 102, 25% of the completers were in partial remission according to the Assessment in Ankylosing Spondylitis (ASAS) criteria. The clinical efficacy and the partial remission achieved in a substantial minority of patients were maintained through 3 years of treatment. In addition, the incidence of enthesitis and anterior uveitis significantly decreased during the third year of treatment in comparison with baseline.26 This group was also analyzed for the clinical response and the time to relapse after discontinuation of anti-TNF therapy. Of note, after infliximab was stopped, all patients experienced a return of disease activity, with a mean time to relapse of 17.5 weeks. Retreatment with infliximab was safe and resulted in clinical improvement in all patients to a state similar to that achieved previously.27

In another trial, most of the patients who received 5 mg/kg infliximab at 0, 2, 6, and 12 weeks reached ASAS 20% response criteria (ASAS20) at week 24, in comparison with 19% of the placebo group. Efficacy was noted as early as 2 weeks after the beginning of therapy and was maintained throughout the 24-week observation period. At week 24, 22.4% of patients in the infliximab group achieved ASAS partial remission, in comparison with 1.3% in the placebo group. In addition, patients receiving infliximab showed significant improvement in BASDAI, BASFI, and BASMI.20

To assess the effects of treatment on inflammation and structural damage, magnetic resonance images of the lumbar spine and sacroiliac joints were assessed. Spinal inflammation and clinical disease activity improved significantly from week 0 to week 30 in patients treated with infliximab plus MTX, in comparison with MTX alone.21 In one study, 2 years of continuous infliximab therapy resulted in persistent improvement in spinal inflammation determined by both T1-weighted gadolinium-enhanced and short tau inversion recovery magnetic resonance imaging (MRI) sequences in all ankylosing spondylitis (AS) patients (n = 20).22 Of note, disease activity parameters did not directly correlate with MRI results.

**Etanercept**

**Structure**

Etanercept is formed by the linkage of two soluble p75 TNF-R extracellular domains to the Fc portion of human IgG1 (see Figure 63-1). The resultant molecule binds both TNF and LT-α with high affinity and specificity.23 The TNF-R domains in etanercept bind to two of the three receptor binding sites on the TNF trimer, thus blocking the ability of TNF to interact with cell-bound TNF-R—a prerequisite for signal transduction24 (see Figure 63-1).

**Pharmacokinetics**

When administered subcutaneously, etanercept is absorbed slowly, reaching a mean peak concentration approximately 50 hours after a single 25-mg dose. The Ig structure affords a half-life of 3 to 4.8 days. The volume of distribution suggests predominantly intravascular distribution.25 The route of clearance from the circulation is unclear, although it is presumed to be mediated through Fc binding by the reticuloendothelial system. Different from what has been seen with anti-TNF monoclonal antibodies, concomitant use of MTX does not appear to alter the pK of etanercept.

**Drug Dose**

Etanercept is administered by subcutaneous injection in doses of 25 mg twice weekly or 50 mg once weekly in RA, PsA, and ankylosing spondylitis. Etanercept is approved for use as monotherapy or in combination with MTX. In skin psoriasis, a higher dose (50 mg twice weekly) is commonly used for the first 12 weeks of therapy. As is true of all available TNF inhibitors, in the clinic, etanercept has been used in combination with DMARDs other than MTX, for example, leflunomide, sulfasalazine, and others.

**Efficacy**

**Rheumatoid Arthritis.** Initial studies demonstrated the efficacy and tolerability of etanercept in both early and refractory disease and also established the optimal dose as 25 mg twice weekly.25,45 In addition to achieving substantial efficacy, as measured by ACR20 clinical response criteria, the use of etanercept has been associated with significant improvement in functional status and quality of life. In one double-blind, placebo-controlled, randomized clinical trial, patients with active and long-standing RA who were refractory to DMARD therapy were treated with etanercept (10 or 25 mg twice weekly) for 6 months. Etanercept rapidly reduced disease activity.27 In another trial, the addition of etanercept in patients with active disease, despite concurrent MTX, was significantly superior to treatment with MTX alone.28 At 6 months, disease activity was significantly reduced in the combination therapy group versus those who received only MTX. In the open-label extension part of this study, patients were able to maintain the improvement, and most were able to decrease their use of MTX, corticosteroid, or both.

Following the success achieved in patients with refractory disease, the role of TNF inhibitors in the treatment of early disease was demonstrated in a large clinical trial in which two doses of etanercept (10 or 25 mg twice weekly) were compared with an accelerated dosing of MTX in MTX-naïve RA patients with disease of less than 3 years’ duration.39,40 Clinical efficacy was achieved more quickly with the standard dose of etanercept than with MTX or
lower-dose etanercept. Moreover, radiographic assessments at 0, 6, 12, and 24 months showed that the rate of radiographic progression was significantly reduced with etanercept compared with MTX.

In long-term, open-label follow-up studies of patients from the clinical trials, responses to etanercept therapy have been sustained over a number of years, with some patients having received more than a decade of treatment. Etanercept has also proved efficacious in patients with juvenile arthritis. In a study known as TEMPO, etanercept plus MTX was shown to be superior to MTX or etanercept alone in patients with relatively early disease. Similarly, improvement in disability based on HAQ was greater with combination therapy. The etanercept plus MTX group showed significantly less radiographic progression than did either group receiving monotherapy, and radiographic progression was significantly less in the etanercept group compared with the MTX group. Because no pK interaction occurs between etanercept and MTX, this study conclusively shows an additive effect on various important outcome results with the combination of MTX and a TNF inhibitor.

Different dosing of etanercept has been assessed in clinical trials. In RA patients, 50 mg etanercept once weekly resulted in similar efficacy to 25 mg twice weekly. Of note, etanercept monotherapy at a dose of 50 mg twice weekly did not result in increased efficacy in RA patients compared with 25 mg twice weekly.

Psoriatic Arthritis. In the initial double-blind, placebo-controlled clinical trial, substantial improvements were observed in PsA response criteria, ACR20 response criteria, and PASI dermatologic scores in the etanercept-treated patients. The median PASI improvement was 46% in the etanercept group versus 9% in the placebo group.

Subsequently, 205 patients with PsA were randomized to treatment with placebo or 25 mg etanercept twice weekly. At week 12, 59% of etanercept-treated patients met the ACR20 criteria, compared with 15% of placebo patients. During the open-label extension through week 48, patients continuing with etanercept treatment maintained or improved their clinical responses, while those in the placebo group showed similar improvements once they began receiving etanercept. The primary radiographic end point was the annualized rate of change in the modified total Sharp score. Radiographic disease progression was inhibited in the etanercept group (−0.03 unit) compared with worsening (+1.00 unit) in the placebo group.

Ankylosing Spondylitis. Similarly positive results were observed with etanercept therapy in patients with ankylosing spondylitis. In a double-blind, placebo-controlled, randomized clinical trial, significant improvement in clinical response (defined by ASAS20) was demonstrated in etanercept-treated patients throughout the 24-week study. Sustained clinical response was maintained in the open-label extension of this study throughout 96 weeks of treatment. Health-related quality of life assessments also significantly improved with etanercept. The extent of clinical response was similar in subsequent clinical trials. Improvement measured by ASAS20 was noted as early as 2 weeks after etanercept was started, and the response was maintained throughout the trial. In a subsequent multicenter trial, significant improvements in BASDAI, BASFI, and BASMI scores were achieved initially in the etanercept-treated group, and later in all patients after the placebo group received etanercept as well. Relapses occurred at a mean of 6 weeks after cessation of etanercept. Re-administration of etanercept led to similar results in the same group of patients. Most patients were able to completely discontinue nonsteroidal anti-inflammatory drugs (NSAIDs).

In one clinical trial, magnetic resonance images of the lower thoracic and lumbar spine of 40 patients were evaluated. Spinal inflammation regressed by 54% in the etanercept group but worsened by 13% in the placebo group after 12 weeks of therapy. After switching to etanercept, placebo patients experienced a similar improvement in spinal inflammation. Finally, it was demonstrated that continuous treatment with etanercept for 24 weeks reduced active spinal changes by 69% as measured by different MRI sequences.

**Adalimumab**

**Structure**

Adalimumab is a human anti-TNF IgG1κ monoclonal antibody generated through repertoire cloning. Adalimumab neutralizes the biologic activity of TNF by binding with high affinity to the soluble and transmembrane forms of TNF and inhibiting the binding of TNF with its receptors (see Figure 63-1).

**Pharmacokinetics**

The peak serum adalimumab concentration and the area under the curve increase linearly with doses in the range of 0.5 to 10 mg/kg. Adalimumab appears to have a low clearance and distributes mainly in the vascular compartment. Its elimination half-life is comparable with that of native IgG1 (10 to 13.6 days). Concomitant use of MTX results in an approximate increase in the area under the curve of adalimumab of 25% to 30%.

**Drug Dose**

The recommended dosing for adalimumab in RA, PsA, and ankylosing spondylitis is 40 mg subcutaneously every other week. In patients who do not achieve an optimal response, it is possible to increase the dosing frequency to weekly. Adalimumab is approved for use as monotherapy or in combination with MTX. In the clinic, it has been used in combination with a number of DMARDs other than MTX.

**Efficacy**

**Rheumatoid Arthritis.** In a phase II trial, adalimumab (20, 40, or 80 mg) via weekly subcutaneous injection for 12 months demonstrated efficacy in comparison with placebo. In a subsequent trial, adalimumab provided significant, rapid, and sustained improvement in disease activity over 24 weeks compared with MTX alone in patients with active RA despite long-term MTX therapy. In a study known as ARMADA, combined treatment with adalimumab and MTX demonstrated sustained improvement in the signs and symptoms of RA as assessed by American College of
Rheumatology (ACR) criteria, as well as notable improvement in functional status as assessed by HAQ scores. In a phase III multicenter trial, 619 patients with active RA and inadequate response to MTX were randomized to receive adalimumab 40 mg every other week, adalimumab 20 mg weekly, or placebo. Both adalimumab regimens were significantly more effective at reducing signs and symptoms and improving physical function in comparison with placebo. In addition, joint damage was assessed radiographically using Sharp scores; patients treated with adalimumab showed significantly smaller changes, and significantly fewer adalimumab-treated patients had new erosions and overall progression of damage compared with those taking placebo. A study evaluating health-related quality of life in two clinical trials demonstrated that adalimumab plus MTX provides statistically significant improvements. Long-term follow-up of patients receiving open-label treatment after several controlled trials has demonstrated sustained improvement and good tolerability over several years.

Following the efficacy noted in patients with refractory RA, adalimumab was assessed in patients with early disease. In a study known as PREMIER, adalimumab plus MTX was superior to adalimumab or MTX monotherapy in MTX-naive patients with early RA (active disease <3 years’ duration) at 1 year. Significantly less radiographic progression was observed among patients in the combination group at both 1 year and 2 years in comparison with the monotherapy groups, although progression was less with adalimumab monotherapy than with MTX monotherapy. Because this study was the first to include all three possible treatment arms (TNF inhibitor, MTX, and TNF inhibitor plus MTX), it definitively established that combination therapy with MTX and TNF inhibitor achieves the best outcomes in patients with early RA. Although TNF inhibitor monotherapy was superior to MTX alone in terms of radiographic progression, clinical efficacy was comparable; however, each of these was substantially less effective than the combination.

Psoriatic Arthritis. Adalimumab given as 40-mg subcutaneous injections every other week has also been studied in patients with PsA. In a large placebo-controlled trial with 313 patients, 50% of patients were on background MTX. A significantly higher percentage of patients from the adalimumab-treated group reached ACR20, -50, and -70 response criteria compared with those receiving placebo. Moreover, the rate of radiographic damage progression was significantly attenuated with adalimumab. As was true in studies of the other TNF inhibitors in PsA, MTX was permitted but not required for enrollment in the study. With this design, there seemed to be no difference in efficacy or toxicity between the combination of MTX plus adalimumab and adalimumab monotherapy. In addition to improvements in various articular manifestations, dramatic improvements were noted in skin psoriasis with adalimumab treatment.

Ankylosing Spondylitis. In the adalimumab trial evaluating long-term efficacy and safety in ankylosing spondylitis (known as the ATLAS trial), a total of 315 patients received either 40 mg adalimumab or placebo every other week for 24 weeks. The number of subjects who met ASAS partial remission criteria was significantly higher in the adalimumab group at weeks 12 and 24 in comparison with the placebo group. Subjects meeting ASAS 5/6 criteria (20% improvement in five of six domains, without 20% worsening in the sixth domain) at weeks 12 and 24 were also significantly more in the adalimumab-treated group. Significant improvement in health-related quality of life was reported in this group of patients. The efficacy of these drugs was demonstrated not only by improved clinical indices but also by spinal inflammation as assessed by MRI.

Golimumab

**Structure**

Golimumab is a human IgG1κ monoclonal antibody specific for human TNF. Golimumab was created using genetically engineered mice immunized with human TNF, resulting in an antibody with human-derived antibody variable and constant regions (see Figure 63-1).

**Pharmacokinetics**

Following subcutaneous administration of golimumab to healthy subjects and patients with RA, the median time to reach time to peak concentration (Tmax) ranged from 2 to 6 days. Golimumab exhibited dose-proportional pharmacokinetics in patients with active RA over the dose range of 0.1 to 10.0 mg/kg following a single intravenous dose. The volume distribution for golimumab indicates that golimumab is distributed primarily in the circulatory system with limited extravascular distribution. Median terminal half-life values were estimated to be approximately 2 weeks in healthy subjects and patients with RA, PsA, or AS. When golimumab 50 mg was administered subcutaneously to patients with RA, PsA, or AS every 4 weeks, serum concentrations appeared to reach steady state by week 12. With concomitant use of MTX, treatment with golimumab every 4 weeks resulted in a mean steady-state trough serum concentration of approximately 0.4 to 0.8 µg/mL in patients with RA, PsA, or AS. Concomitant MTX usage resulted in higher mean steady-state trough concentrations of golimumab in patients with RA, PsA, or AS (52%, 36%, and 21% respectively), compared with those treated with golimumab alone.

**Drug Dose**

Golimumab is administered by subcutaneous injection, 50 mg once a month, in patients with RA, PsA, and AS.

**Efficacy**

Rheumatoid Arthritis. Results of a phase II study of golimumab in 172 patients with active RA despite MTX therapy demonstrated the efficacy of golimumab given every 4 weeks by subcutaneous injection in combination with MTX. The clinical effect was evident as early as 2 weeks after the first dose and was sustained to 1 year. The efficacy and safety of golimumab were evaluated in three multicenter, double-blind, placebo-controlled trials. Golimumab was administered subcutaneously at doses of 50 mg or 100 mg every 4 weeks. In the GO-FORWARD study, patients with active RA despite a stable dose of MTX who...
had not been previously treated with a biologic TNF inhibitor were randomized to receive golimumab 50 mg + MTX, golimumab 100 mg + MTX, or golimumab 100 mg monotherapy every 4 weeks. All golimumab regimens were significantly more effective at reducing signs and symptoms and in improving physical function. Patients who had less than 20% improvement in swollen and tender joint count entered early escape at week 16. The proportion of patients who achieved ACR20 and reached a low disease activity score at week 52 was significantly greater in the golimumab group. Patients who were treated with 100 mg of tocilizumab appeared to have increased risk of serious adverse events and serious infections.

Patients who were previously treated with one or more doses of TNF inhibitor without a serious adverse reaction were enrolled in the GO-AFTER study. Greater efficacy measured by ACR20 and Disease Activity Score (DAS) 28 was observed in patients who received golimumab 50 mg plus MTX at week 14 compared with patients who received MTX alone. Improvements in HAQ scores were also markedly higher for patients receiving golimumab plus MTX in comparison with MTX alone.

Patients with active RA who were MTX-naïve and had not been previously treated with TNF inhibitor were randomized to be treated with MTX alone, golimumab 50 mg plus MTX, golimumab 100 mg plus MTX, or golimumab 100 mg alone in the GO-BEFORE study. Golimumab plus MTX led to significant improvement in ACR responses (ACR50, 40%) at 24 weeks in MTX-naïve patients compared with patients who received MTX alone (ACR50, 29%).

An MRI substudy of the GO-BEFORE and GO-FORWARD trials demonstrated that patients who received golimumab plus MTX had improvements in inflammation (synovitis and osteitis) exceeding those observed with MTX only as early as week 12 and continuing through week 24. Significant improvements in erosions were also observed in the GO-BEFORE trial.

Psoriatic Arthritis. The GO-REVEAL study evaluated the efficacy and safety of golimumab in PsA patients. Four hundred five patients with active PsA despite NSAID or DMARD therapy who had not been previously treated with TNF inhibitor were randomized to receive golimumab 50 mg or 100 mg with or without MTX every 4 weeks. Fifty-one percent of patients who received golimumab reached ACR20 response at week 14 in comparison with 9% of patients who received only MTX. Significant improvements in enthesitis and HAQ scores were seen in patients treated with golimumab, in addition to significant improvements in the psoriasis skin lesions evaluated with PASI responses.

Long-term radiologic outcome was assessed using the van der Heijde–Sharp (vH-S) method modified for PsA. Golimumab 50 mg and 100 mg given to patients with active PsA showed no to minimal evidence of radiographic disease progression through week 104.

Ankylosing Spondylitis. The efficacy of golimumab in AS was evaluated in the GO-RAISE study. Three-hundred fifty-six patients with active AS according to modified New York criteria were randomized to receive placebo, golimumab 50 mg, or golimumab 100 mg every 4 weeks. Patients were allowed to be on stable doses of sulfasalazine, hydroxychloroquine, low-dose corticosteroids, or NSAIDs during the trial. The use of other DMARDs was prohibited. At weeks 14 and 24, golimumab 50 mg once a month demonstrated significant ASAS20 and ASAS40 responses (56% and 44%, respectively) versus placebo (23% and 15%, respectively).

Certolizumab Pegol

Structure

Certolizumab pegol is a Fab fragment of a recombinant, humanized anti-TNF monoclonal antibody that has been fused to a 40-kD polyethylene glycol (PEG) moiety. The Fab fragment is composed of a light chain with 214 amino acids and a heavy chain with 229 amino acids. Given this structure, certolizumab is not able to effect Fc-mediated activities that can be mediated by full monoclonal antibody and constructs containing Fc pieces such as complement-dependent or antibody-dependent cell-mediated toxicity or apoptosis.

Pharmacokinetics

Single intravenous and subcutaneous doses of certolizumab pegol have predictable dose-related plasma concentrations with a linear relationship between the dose administered and the peak concentration (Cmax). A mean Cmax of approximately 43 to 49 µg/mL occurred at week 5 in patients with RA at the recommended dose regimen. Following subcutaneous administration, peak plasma concentrations of certolizumab pegol were attained between 54 and 171 hours post injection. PEGylation delays the metabolism and elimination of these entities from the circulation by a variety of mechanisms, including renal clearance, proteolysis, and immunogenicity. The terminal elimination phase half-life was approximately 14 days for all doses tested.

Drug Dose

The recommended dose of certolizumab pegol is 400 mg initially and at weeks 2 and 4, followed by 200 mg every other week. For maintenance dosing, certolizumab pegol 400 mg every 4 weeks can be considered. A 400-mg dose is given as two subcutaneous injections of 200 mg at separate sites in the thigh or abdomen.

Efficacy

Intravenous ascending-dose monotherapy with certolizumab pegol was shown to effectively control the signs and symptoms of RA in a phase II trial. The efficacy and safety of certolizumab pegol were assessed in multiple randomized, placebo-controlled, double-blind studies. RAPID-I and RAPID-II studies included patients with active RA despite MTX. Both were 52 week studies that evaluated the lyophilized (RAPID-I) or the liquid (RAPID-II) preparation of certolizumab with endpoints of progression of structural damage (change from baseline in modified Sharp score) and signs and symptoms of RA measured by ACR20, -50, and -70. Patients were randomized to receive loading doses of 400 mg of certolizumab pegol or placebo at weeks 0, 2, and 4 followed by 200 mg of certolizumab pegol, 400 mg
certolizumab pegol, or placebo every other week in combination with MTX. Certolizumab pegol–treated groups showed rapid and sustained improvement in clinical disease activity starting from week 1 and sustained through week 52. Mean radiologic progression was significantly reduced in the certolizumab–treated groups versus MTX alone.\textsuperscript{78,79} Rapid and sustained improvements in health-related quality of life, fatigue, home and workplace productivity, and social activities were also reported in patients treated with certolizumab pegol in RAPID-I and -II trials.\textsuperscript{80,81} Sustainability of improvements in RA signs and symptoms and inhibition of joint damage progression, as well as tolerability of certolizumab pegol + MTX, were evaluated in the RAPID-II open-label extension study. Results at 3 years showed that 400 mg every other week dosing did not significantly increase the efficacy outcome; certolizumab pegol was generally well tolerated; and clinical and radiologic improvements were sustained over this period of time.\textsuperscript{82} In a broader group of RA patients with inadequate response to more than one DMARD, including prior TNF inhibitor exposure, addition of certolizumab pegol to the regimen resulted in rapid clinical response measured with ACR20 response rates and DAS28 scores, and improved function measured with HAQ scores.\textsuperscript{83}

**Mechanism of Action of TNF Inhibitors**

Several potential mechanisms of action may explain the efficacy of TNF inhibitors in RA and other conditions (Table 63-1). Although some support has been put forth for these varied mechanisms, the exact relationship between any particular mechanism and specific aspects of clinical efficacy remains to be delineated. Downregulation of local and systemic proinflammatory cytokine production and reduction of lymphocyte activation and migration into the joint may be the most relevant mechanisms; for example, serum levels of IL-6 and IL-1 are significantly reduced after administration of anti-TNF monoclonal antibody.\textsuperscript{84,85} The reduction in TNF and the consequent reduction in IL-1 would be expected to reduce the synthesis of matrix metalloproteinase (MMP) and the production of other degradative enzymes. Serial studies have shown that there is in fact a marked reduction in pro–MMP-3 and pro–MMP-1 after anti-TNF therapy.\textsuperscript{86-88} As noted, anti-TNF therapy is also associated with a reduction of lymphocyte migration into the joints of patients with RA. Using radiolabeled granulocytes, it has been demonstrated that anti-TNF monoclonal antibody significantly reduces cell movement into the affected joints.\textsuperscript{89} In addition, post-treatment synovial biopsies show reduced cellular infiltrates, with fewer T cells and macrophages present.\textsuperscript{90} These effects are thought to be secondary to a reduction in the expression of endothelial adhesion molecules in the synovial tissue. Treatment with anti-TNF monoclonal antibody results in a dose-dependent decrease in soluble forms of intercellular adhesion molecule-1 (ICAM-1) and E-selectin (CD62E).\textsuperscript{89} Changes in soluble E-selectin, soluble ICAM-1, and circulating lymphocytes with anti-TNF therapy correlate with clinical outcomes. Vascular endothelial growth factor (VEGF) is a potent endothelial cell–specific angiogenic factor. It is produced in the synovium and is an important regulator of neovascularization in the pannus. After anti-TNF therapy, VEGF serum levels are reduced in patients with RA. This decrease correlates significantly with the clinical benefit observed in these patients.\textsuperscript{91} Because angiogenesis is a prominent feature of rheumatoid synovium, the relationship between inflammation and angiogenesis has been investigated. Computerized image analysis of endothelium for multiple markers of endothelium (e.g., von Willebrand factor, CD31) and neovascularure (αβ3) has shown reduced vascularity after anti-TNF therapy. A number of other potential mechanisms of action have been suggested to be operative for TNF inhibitors (see Table 63-1), although debate on these aspects is ongoing.

**Table 63-1** Potential Mechanisms of Action of Tumor Necrosis Factor Inhibitors

<table>
<thead>
<tr>
<th>Decrease Production of Other Inflammatory Mediators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytokines (e.g., IL-1, IL-6, GM-CSF)</td>
</tr>
<tr>
<td>Chemokines (e.g., IL-8)</td>
</tr>
<tr>
<td>Degradative enzymes (e.g., MMPs)</td>
</tr>
<tr>
<td>Other mediators (e.g., C-reactive protein)</td>
</tr>
<tr>
<td>Alter Vascular Function; Leukocyte Traffic and Activation</td>
</tr>
<tr>
<td>Decreased adhesion molecule expression and function</td>
</tr>
<tr>
<td>Angiogenesis inhibition</td>
</tr>
<tr>
<td>Modulate the Function of Immunocompetent Cells</td>
</tr>
<tr>
<td><strong>T Cells</strong></td>
</tr>
<tr>
<td>Normalization activation threshold for CD3–T cell receptor signaling</td>
</tr>
<tr>
<td>Alter Th1/Th2 phenotype, cytokine secretion</td>
</tr>
<tr>
<td>Increase regulatory T cell number and function</td>
</tr>
<tr>
<td>Induce apoptosis (?)</td>
</tr>
<tr>
<td><strong>Monocytes and Macrophages</strong></td>
</tr>
<tr>
<td>Modulate HLA-DR expression</td>
</tr>
<tr>
<td>Possibly increase apoptosis (?</td>
</tr>
</tbody>
</table>

GM-CSF, granulocyte-macrophage colony-stimulating factor; HLA-DR, human leukocyte antigen DR; IL, interleukin; MMPs, matrix metalloproteinases; Th, T helper.

**Other Considerations**

**Treatment in Other Autoimmune Conditions**

The role of TNF inhibitors in the treatment of Crohn’s disease, juvenile idiopathic arthritis, and psoriasis has been clearly defined. Based on promising results in various immune conditions, these agents have been used in a variety of other disorders, including idiopathic and spondyloarthopathy-related anterior uveitis, sarcoidosis, Sjögren’s syndrome, Behçet’s syndrome, inflammatory myopathies, and various types of vasculitis. Although a number of case reports or small, uncontrolled clinical trials have reported on these conditions, there is a paucity of conclusive data from controlled trials. Perhaps the most notable clinical response has been observed in the treatment of anterior uveitis, especially with anti-TNF monoclonal antibody constructs.\textsuperscript{92} On occasion, promising results in uncontrolled trials have been disproved in controlled trials. This was observed in a placebo-controlled trial of etanercept given in addition to standard therapy for remission induction and maintenance in patients with granulomatosis with polyangiitis (formerly Wegener’s granulomatosis). Despite promising anecdotal evidence, the...
addition of the TNF inhibitor not only failed to achieve any clinical improvement but also resulted in greater risk of solid malignancies beyond that observed with cyclophosphamide alone. 93

Despite evidence that anti-TNF agents can result in the development of certain autoantibodies and even lupus-like syndromes, the safety and efficacy of anti-TNF agents have been assessed in a small group of patients with systemic lupus erythematosus (SLE). Patients with joint involvement experienced remission of arthritis, and a significant reduction in the level of proteinuria occurred with infliximab. In this small study, TNF inhibitor therapy did not lead to adverse events suggestive of an increase in SLE activity; however, as might have been expected, autoantibodies to double-stranded DNA and cardiopilin did increase. 94

Cardiovascular Risk and Lipid Profile

Cardiovascular morbidity and mortality appear to be increased in autoimmune conditions. This may be related both to increased prevalence of traditional risk factors for cardiovascular disease and to uncontrolled systemic inflammation, which appears to predispose independently to accelerated progression of atherosclerosis. Treatment with TNF inhibitors appears to be associated with overall improvement in cardiovascular disease risk, related to beneficial effects on lipid parameters and to control of systemic inflammation.95,96 Long-term investigations are needed to define the possible beneficial effects of TNF inhibitors on overall and cardiovascular survival in patients with autoimmune disease.95,96

Monitoring

All patients must be evaluated for active and inactive (latent) tuberculosis infection before initiation of therapy. Appropriate screening tests (e.g., tuberculin skin test, ex vivo testing for tuberculosis, chest x-ray) should be performed on all patients. Use of TNF inhibitor therapy may increase the risk of reactivation of hepatitis B virus (HBV) among those who are chronic carriers of HBV. Therefore evaluating patients for HBV before starting TNF inhibitor therapy is recommended. No other specific laboratory monitoring is currently required by regulatory agencies during therapy with TNF inhibitors. Nevertheless, because of the rare occurrence of myelosuppression and concern about the risk of infection, clinicians typically assess the complete blood count (CBC) intermittently during therapy. Assiduous monitoring of patients for any sign or symptom of infection, demyelinating disease, and malignancy is requisite during treatment with all TNF inhibitors. Repeat testing at regular intervals (e.g., annually) for exposure to tuberculosis has been recommended by some regulatory authorities.

Pregnancy and Breastfeeding

Developmental toxicity studies in rats, rabbits, and mice have not revealed any maternal toxicity, embryo toxicity, or teratogenicity associated with TNF inhibition. Minimal human pregnancy information has been published for these medications, and most data consist of isolated case reports, retrospective surveys, and uncontrolled studies. As the number of patients treated with TNF inhibitors increases, a growing number of pregnancies will be reported among them.97 Outcome data based on anecdotal observations of small numbers of pregnant women treated with infliximab, etanercept, and adalimumab reveal that the relative rates of live births, miscarriages, and therapeutic terminations were comparable with rates in a national cohort of age-matched healthy women. TNF inhibitors are classified as U.S. Food and Drug Administration (FDA) Pregnancy Category B (animal reproduction studies have failed to demonstrate a risk to the fetus, and no adequate and well-controlled studies have been performed in pregnant women). The use of anti-TNF agents in pregnancy is recommended only if such treatment is clearly needed. If TNF inhibitors are used during pregnancy, it must be noted that transfer to the fetus is possible, and monitoring might be considered on that basis. Because it is not currently known whether TNF blockers are excreted in human milk, or whether they are absorbed systemically after ingestion, it is recommended that TNF blockers not be used by nursing mothers.

Vaccinations

It is preferred to have all recommended vaccinations brought up to date before initiating treatment with TNF inhibitors. Because data on the response to live vaccinations or the secondary transmission of infection by live vaccinations in patients receiving TNF inhibitors are insufficient, concurrent administration of live vaccines with TNF inhibitors is not recommended.

Toxicity

In clinical trials, etanercept, infliximab, adalimumab, golimumab, and certolizumab pegol have generally been well tolerated.98 Longer-term follow-up of patients initially enrolled in clinical trials has provided additional safety data for these agents. However, TNF plays a key role not only in the pathogenesis of autoimmune disease but also in normal immune homeostasis. Therefore, a number of safety considerations, including the potential risk of infection and malignancy, are germane to the optimal clinical use of these agents.98 Additional information concerning adverse effects associated with these agents has been obtained through pharmacovigilance. Adverse events related to the use of TNF inhibitors can be grouped into those that are agent-related and those that are target-related (Table 63-2).99 Injection site and infusion reactions and immunogenicity and their sequelae vary, depending on the particular agent. A potentially increased predisposition to infection, development of malignancy, and induction of autoimmune disorders and an association with demyelinating disorders, myelosuppression, and worse outcomes with congestive heart failure might be considered target-related adverse events. Thus, any clinically effective TNF inhibitor might be expected to be

*References 12, 16, 35-37, 56, 71, 80, 81.
associated with such adverse events, although the relative risk among different agents may vary, depending on dose and other factors.

**Infusion and Injection Site Reactions.** Infliximab has been associated with infusion reactions, the most common of which are headache (20%) and nausea (15%). These are rarely severe, are usually transient, and typically can be controlled by slowing the rate of infusion or by treating with acetaminophen or antihistamines. With etanercept, adalimumab, golimumab, and certolizumab pegol, cutaneous reactions at injection sites represent the most frequent administration-related side effect; however, they rarely lead to discontinuation of therapy. Injection site reactions typically consist of erythematous or urticarial lesions. Although they can arise at sites of previous injections, these reactions seem to be limited to the skin and are not associated with other or systemic features of immediate hypersensitivity. Reactions typically occur close to treatment initiation and abate over time, even with continued dosing.

**Antigenicity.** As is true for any therapeutic agent (especially large protein molecules, some of which contain foreign sequences), antibodies to anti-TNF agents can develop. Although the clinical relevance of these antibodies is presently unclear, they can diminish the half-life of the therapeutic agent and consequently decrease its efficacy. Approximately 3% of etanercept-treated patients develop antibodies to the drug. In an early study, it was noted that antibodies to infliximab developed in 53%, 21%, and 7% of patients who were receiving 10, 3, and 1 mg/kg infliximab, respectively. RA trials of infliximab with or without concomitant MTX treatment revealed that immunogenicity was decreased by concomitant MTX, perhaps owing in part to the increase in the half-life of infliximab associated with MTX use. A multicenter trial of infliximab therapy in Crohn’s disease demonstrated that induction of these anti-infliximab antibodies might contribute to hypersensitivity reactions in some patients. Antibodies to adalimumab, golimumab, and certolizumab pegol developed in about 4% to 12% of patients; this rate was reduced to 1% with concurrent MTX treatment.

Although it is believed that there is a trend toward higher clearance of TNF inhibitors in the presence of antibodies to the construct, routine testing for antibodies to TNF inhibitors is not widely available, nor is it currently recommended.

**Infection.** Given that TNF is a key mediator of inflammation, a major concern surrounding the use of TNF inhibitors is their potential to increase the risk of infection. Although inhibition of TNF in animals does not appear to increase their risk for infection with most pathogens, it does interfere with the ability to mount an inflammatory response against intracellular organisms. In experimental models, TNF blockade impaired resistance to infection with mycobacteria, Pneumocystis carinii, fungi, Listeria monocytogenes, and Legionella. In patients with RA, infection with these types of opportunistic organisms has been observed. However, confounding the attribution of infection to any therapeutic agent is the fact that infections occur more frequently and are important contributors to the accelerated morbidity of RA patients compared with the normal population. It is difficult to determine how much of this susceptibility relates to the disease itself and how much is caused by the effects of immunomodulatory drugs (e.g., steroids, disease-modifying antirheumatic drugs [DMARDs]). The subset of RA patients with great susceptibility to infection (i.e., those with severe, active disease) is also the subset most commonly enrolled in trials of TNF inhibitors; this is the group of patients for whom these agents have the greatest clinical utility.

In RA trials with TNF inhibitors, a number of infections have occurred. In general, the most frequent infections have been those that occur most commonly among all persons, such as upper and lower respiratory tract infections and urinary tract infections. In most studies, a slightly greater propensity to develop infection was seen in patients receiving TNF inhibitors; however, this trend is common in most studies of effective therapies for RA. The incidence of serious infection, defined as infection requiring hospitalization or treatment with parenteral antibiotics, among RA patients treated with TNF inhibitors was similar to that of the control groups in individual studies; it also approximated the incidence noted among RA patients before the anti-TNF era. In certain subgroups, such as patients with early RA, the overall incidence of infection was less than in patients with more long-standing disease, and infections and serious infections were comparable among TNF inhibitor–treated patients and controls.

It is worth noting, however, that several characteristics of clinical trials may affect investigators’ ability to extrapolate their safety data to the clinic. In general, patients enrolled in clinical trials tend to be healthier and therefore less likely to develop adverse effects such as infection, compared with the general population of RA patients in the clinic. Therefore, postmarketing data provide an important complement to safety data obtained from clinical trials. Also, clinical trials are powered to assess efficacy and therefore may not include sufficient numbers of patients to ascertain real but small differences in uncommon side effects. A systematic analysis that combined the results from nine clinical trials of TNF inhibitors has been performed. This analysis found an increased risk of serious infection among patients receiving TNF inhibitors compared with controls (3.6% vs. 1.7%); however, it should be noted that a nonstandard definition of serious infection was used, and no attempt was made to control for the time of exposure,
which was nearly always longer for patients receiving TNF inhibitors. In this same analysis, a nonsignificant trend toward a greater incidence of serious infection was observed with higher doses of TNF inhibitor. In one of the only clinical trials that had a primary outcome of safety, use of a high-dose TNF inhibitor was also associated with a greater incidence of serious infection compared with a lower dose; the lower dose was no different from placebo in this regard. In postmarketing surveillance data, also known as pharmacovigilance, serious infections have certainly been observed among patients receiving TNF inhibitors. The relative impact of potentially confounding factors such as comorbidities and concomitant medications on the rate of serious infection remains incompletely defined. This important question has also been addressed by using registries of RA patients. In a German registry, rates of infection and serious infection among 858 RA patients receiving treatment with TNF inhibitors were compared with those among 601 patients receiving only DMARDs. The relative risk for infection (3.3 to 4.1) as well as serious infection (2.7 to 2.8) was significantly higher among patients receiving TNF inhibitors. However, these patients also had more severe and more active RA, thereby placing them at greater risk of infection. When the investigators used propensity scoring methods to control for severity of disease as a confounder, the relative risks were reduced. The risk of infection decreased to 2.3 to 3.0, and that for serious infection became a nonsignificant trend of 2.1. In data from a British registry, 7644 RA patients treated with TNF inhibitors were compared with 1354 RA patients on DMARDs alone. In this analysis, the crude rate of serious infection was higher among TNF inhibitor–treated patients (1.28; 95% confidence interval [CI], 0.94 to 1.76), although this did not reach statistical significance. Further, when the rates were adjusted for age, sex, severity of RA, use of corticosteroids, and comorbidity, no differences between the groups were noted (relative risk [RR], 1.03; 95% CI, 0.68 to 1.57).

In summary, although treatment with TNF inhibitors can result in increased risk of infection and serious infection, other factors such as the severity of RA, the use of other medications such as corticosteroids, and the presence of comorbidities are important contributors to these outcomes. Clinicians must monitor patients closely for signs and symptoms of infection, and it is worth noting that TNF inhibitor therapy itself can mask the initial signs and symptoms of infection.

Opportunistic infections, particularly disseminated Mycobacterium tuberculosis, are of concern with the use of TNF inhibitors. Of note, more patients treated with TNF inhibitors have extrapulmonary and disseminated tuberculosis (TB), highlighting the specific role of TNF in controlling this infection. Rates of TB associated with the use of TNF antagonists are higher in geographic regions where TB is more prevalent in the general population. Most cases of TB observed in the early years after the introduction of TNF inhibitors arose within the first few months after initiation of therapy and were probably related to the reactivation of latent TB. Very few cases of TB were observed during clinical trials of the TNF inhibitors, highlighting the important role of pharmacovigilance in identifying safety signals with new therapies. For etanercept, no cases of TB occurred in clinical trials, but 38 cases of etanercept-associated TB were reported worldwide among an estimated 150,000 patients exposed through December 2002. For infliximab, 441 cases of TB were reported among approximately 500,000 initially exposed patients; only 6 cases of infliximab-related TB were reported from clinical trials. Ninety-seven percent of infliximab-related cases occurred within 7 months of treatment initiation, with a median time of onset of 12 weeks. The incidence of TB in clinical trials with adalimumab was greater in earlier clinical trials; this was related to lack of screening, the locations of the studies, and the higher doses used in early trials. The incidence dropped to 1% after adalimumab was reduced to its current dose, and after screening for latent TB infection was instituted before therapy (21 cases in 2400 patients). The incidence of TB was even lower in golimumab and certolizumab pegol trials. The incidence of TB was 0.23 in golimumab clinical trials, and most cases occurred in countries with a high incidence rate of TB. In studies with certolizumab pegol, 36 cases of TB occurred among 2367 exposed patients; these cases also occurred in countries with endemic rates of TB. This highlights the benefit of screening for and treating latent TB among patients being considered for TNF inhibitor therapy. However, because treated patients may acquire new cases of TB, and because cases of latent TB may be missed owing to false-negative screening tests, constant vigilance for TB is required during therapy with TNF inhibitors.

The impact of screening for latent TB in patients receiving anti-TNF agents has been assessed in a Spanish registry; the rate of development of active TB among RA patients treated with anti-TNF agents dropped by 83% with use of the recommended guideline. Current U.S. guidelines recommend purified protein derivative (PPD) skin testing and/or ex vivo testing for tuberculosis, as well as a chest radiograph, before anti-TNF therapy is initiated. If the PPD test is positive without evidence of active infection, treatment for latent TB with isoniazid is recommended. The recommended duration of therapy is 9 months. Recommendations concerning the timing of TNF inhibitor therapy and isoniazid prophylaxis for latent TB vary; however, concomitant initiation appears feasible. During anti-TB treatment, alanine aminotransferase (ALT) monitoring is recommended, especially for those who chronically consume alcohol and/or who take potentially hepatotoxic drugs. Treatment should be adjusted according to local guidelines.

Malignancy. Anti-TNF drugs can theoretically affect the host defense against malignancy. To date, the occurrence of malignancies in clinical trials and long-term follow-up of RA patients from clinical trials do not appear to exceed the rate that would be expected in this population. The overall rate of most malignancies in patients with RA is the same as in the normal population. However, risks of lymphoma and lung cancer appear to be increased in patients with RA. Although the actual reason is not known, the severity and duration of the disease and the use of immunomodulatory agents such as MTX seem to play a role in the increased risk of lymphoma in RA patients. Postmarketing analysis of the association between anti-TNF agents and lymphoma is inconclusive. In one population-based analysis, standardized incidence ratios of lymphoma
among patients receiving anti-TNF therapies were somewhat higher than those among RA controls; however, this analysis did not adjust for baseline differences between patients.\textsuperscript{118}

In a more recent analysis, in which age, sex, and disease duration adjustments were made, no increased risk of lymphoma was identified among RA patients treated with anti-TNF agents compared with those treated with other therapies.\textsuperscript{119} The systematic review of clinical trials of anti-TNF monoclonal antibodies demonstrated an increased risk of malignancies, including lymphomas and skin cancers, with treatment, although the longer time of exposure to TNF inhibitors was not accounted for.\textsuperscript{120} As anti-TNF agents are being used more frequently in children, a possible increased risk of malignancies in children with autoimmune disorders, particularly in those receiving TNF inhibitors, has been pointed out by the Adverse Event Reporting System of the FDA. Forty-eight reports of malignancy were identified: 31 following infliximab use, 15 following etanercept use, and 2 following adalimumab use. Half of the malignancies were lymphomas, and most cases involved concomitant use of other immunosuppressants.\textsuperscript{121} Given these uncertainties, caution is indicated when the use of anti-TNF agents is considered for patients with a history of malignancy or for those at high risk of malignancy for other reasons. Longer-term follow-up of larger numbers of patients will provide clinicians with a better idea of the safety of these agents in this regard.

Autoimmune Disorders. Approximately 10% to 15% of patients treated with any TNF inhibitor develop antibodies to double-stranded DNA.\textsuperscript{98,121} However, few patients (0.2% to 0.4%) develop symptoms consistent with drug-induced lupus. The mechanism and the significance of the development of antibodies are uncertain, although this adverse effect seems relatively specific for TNF inhibitors and is not noted with other biologic agents. Of note, patients with TNF inhibitor–related lupus generally do not develop life-threatening lupus involvement (e.g., nephritis, central nervous system lupus) and rarely develop the diversity of other autoantibodies characteristic of idiopathic SLE (e.g., anti-Sm/RNP, anti-Ro/La, anti-ScI70). A few patients have reportedly developed anticytologic antibodies, but they are mostly asymptomatic. Among those few patients who developed lupus-like symptoms while on TNF inhibition therapy, improvement has been seen upon discontinuation of therapy. Although the rare occurrence of autoimmune disorders has not dissuaded most clinicians from using TNF inhibitors in patients with RA, some remain cautious about using these drugs in patients with a history of SLE.

Demyelinating Syndromes. Several cases of multiple sclerosis (MS) or peripheral demyelinating disease have been reported with anti-TNF therapy in patients with RA, PsA, and Crohn’s disease.\textsuperscript{122} In addition, two studies of TNF inhibitors in MS patients showed worsening of MS-related symptoms and exacerbations in the treated group.\textsuperscript{123,124} Although some evidence suggests that the incidence of MS may be increased in patients with RA, the association between anti-TNF therapy and MS remains unclear. The risk of developing a demyelinating disease is very small; however, many clinicians withhold anti-TNF therapy in patients with a history of demyelinating diseases and in those showing signs and symptoms of such disease during anti-TNF therapy.

**Congestive Heart Failure.** Data suggest that TNF may play a role in the pathogenesis of congestive heart failure (CHF), and inhibition of TNF was highly effective in animal models of ischemic cardiomyopathy. However, in trials of TNF inhibitors in patients with stable but severe (class III or IV) CHF, no clinical benefit was observed, and in some treatment arms, higher incidences of mortality and hospitalization for worsening of CHF were reported. Thus, TNF inhibition has been largely abandoned as a therapeutic approach in patients with CHF. In patients with RA, treatment with TNF inhibitors does not appear to result in an increased incidence of CHF.\textsuperscript{125} In fact, TNF inhibitor therapy may actually improve mortality associated with heart disease and overall mortality in RA patients.

**INTERLEUKIN-1**

Members of the IL-1 family include IL-1α, IL-1β, and the naturally occurring IL-1 receptor antagonist (IL-1Ra). Specific cellular proteases process IL-1α and IL-1β to their 17-kD mature forms. Pro–IL-1α precursor is active intracellularly. However, pro–IL-1β is not active before cleavage with IL-1β–converting enzyme. After cleavage, it is secreted and is fully functional. IL-1Ra is a naturally occurring antagonist protein with amino acid sequence homology to IL-1α and IL-1β. Multiple forms of this protein exist. One is secreted and functions as a competitive inhibitor of IL-1α and IL-1β, binding to the same counterreceptor but transducing no signal. The IL-1 polypeptides bind to two cell surface receptors: type I (IL-1RI) and type II (IL-1RII). IL-1RI is found on most cell types, whereas IL-1RII occurs mainly on the surface of neutrophils, monocytes, B cells, and bone marrow progenitor cells. When IL-1 binds to IL-1RI, the signal transduction is mediated through the association of a second receptor unit, IL-1R accessory protein. The three members of the IL-1 family bind to IL-1RI with similar affinities. Binding of IL-1 to IL-1RII does not lead to signal transduction. IL-1RII acts like a decoy receptor and a competitive inhibitor. Soluble forms of IL-1RII inhibit IL-1 activity by competing with IL-1RI for IL-1 binding. IL-18 is another member of the IL-1 family of inflammatory cytokines. It is now recognized as an important regulator of innate and acquired immune responses. IL-18 is expressed at sites of chronic inflammation, in autoimmune diseases, in a variety of cancers, and in the context of numerous infectious diseases. IL-18 likely plays a role in RA, and strategies to block IL-18 activity are under way in clinical trials.\textsuperscript{126}

As with TNF, IL-1 is one of the key mediators of the inflammatory response. Studies in animal models of arthritis have demonstrated the therapeutic potential of IL-1 blockade. IL-1β gene knockout mice show markedly reduced levels of inflammation following immunization with type II collagen. The use of genetically modified mice has helped to confirm the physiologic significance of IL-1Ra: Deletion of this gene in mice results in the spontaneous development of arthritis.
Anakinra

**Structure and Mechanism of Action**

Anakinra is a recombinant, nonglycosylated homolog of IL-1R that differs from native human IL-1R by the addition of a single methionine residue at its amino terminus. Anakinra blocks the activity of IL-1 by competitively inhibiting IL-1 binding to the IL-1RI receptor\(^{128}\) (Figure 63-2). Levels of the naturally occurring IL-1R, which are elevated in the synovium and synovial fluid from RA patients, appear to be insufficient for the excess amount of locally produced IL-1.

**Pharmacokinetics**

In subjects with RA, maximum plasma concentrations of anakinra occur 3 to 7 hours after the subcutaneous administration of clinically relevant doses (1 to 2 mg/kg). The terminal half-life ranges from 4 to 6 hours. In RA patients, no unexpected accumulation of anakinra is observed after daily subcutaneous doses for up to 24 weeks. Estimated anakinra clearance increases with increasing creatinine clearance and body weight.

**Drug Dose**

The recommended dose of anakinra for the treatment of patients with moderately to severely active RA is 100 mg/day administered by subcutaneous injection. Anakinra can be used alone or in combination with MTX. Because of the potential for increased risk of infection, it is not recommended for use in conjunction with TNF inhibitors.

**Efficacy**

**Rheumatoid Arthritis.** Preliminary clinical studies indicated that anakinra could be safely administered by subcutaneous injection.\(^{127}\) The efficacy of anakinra in the treatment of active RA was confirmed in a 24-week, phase II, placebo-controlled study in which 472 patients received daily subcutaneous injections of placebo or one of the three different doses of anakinra (30, 75, or 150 mg).\(^{128}\) Improvements were observed in all individual clinical parameters, including swollen and tender joint counts, pain score, duration of early morning stiffness, and patient and physician assessment of disease activity. More patients on the higher dose of anakinra achieved improvement in ACR20 criteria compared with the placebo group. However, the overall magnitude of the reduction in clinical symptoms and signs (20% to 30%) was relatively modest compared with that seen with TNF-blocking agents (60% to 70%). Patients were allowed to go on to a non–placebo-controlled extension study with three different doses of anakinra after completing the 24-week double-blind study. On completion of the extension study, 55% of patients who had previously received placebo achieved an ACR20 response. Of the patients who continued receiving the same dose of anakinra, 49% maintained an ACR20 response at 48 weeks. Analysis of hand radiographs by two different methods after 24 weeks of treatment showed a statistically significant decrease in the rate of progressive joint damage compared with placebo.

Anakinra is a competitive inhibitor of IL-1 that must be continuously present in great excess to be effective, and it must be administered daily. It was hypothesized that the relatively modest clinical results seen with anakinra in RA patients (compared with those achieved by TNF inhibitors) might be related to the agent rather than to the target. Two lines of evidence appear to refute this hypothesis: the efficacy of anakinra in other inflammatory diseases, and the comparable clinical efficacy of other IL-1 inhibitors in RA.

**Other Inflammatory Diseases.** Autoinflammatory diseases are a group of conditions that include familial cold autoinflammatory syndrome (FCAS), Muckle-Wells syndrome (MWS), and neonatal-onset multisystem inflammatory disease (also known as chronic infantile neurologic, cutaneous, articular [CINCA] syndrome). These conditions, which share some clinical features, are associated with various mutations in the NALP3/CIAS1/PYPAF1 gene, which encodes the protein cryopyrin. Cryopyrin is a key component of the inflammasome; thus, the autoinflammatory syndromes may be related to abnormalities in IL-1 regulation. This was proved by the remarkable responses to anakinra reported in these syndromes.\(^{131,132}\) In addition,
significant and rapid responses have been achieved when anakinra was used to treat patients with adult-onset Still's disease. Improvement in various hematologic, biochemical, and other markers suggests that IL-1 plays a key role in this disease as well.\textsuperscript{133}

### Rilonacept

Rilonacept, previously known as IL-1 Trap, is a fusion protein consisting of the human IL-1 receptor extracellular domains and the Fc portion of human IgG1. It incorporates in a single molecule the extracellular domains of both receptor components required for IL-1 signaling: IL-1RI and the IL-1R accessory protein (Figure 63-3). Rilonacept has a very high binding affinity for IL-1 (dissociation constant \( \approx 1 \) pM), and it is specific for IL-1\( \beta \) and IL-1\( \alpha \). Based on early studies, the subcutaneous administration of rilonacept in subjects with RA provided evidence of its clinical and biologic activity. However, a double-blind, placebo-controlled clinical trial in patients with moderate to severe RA who were randomized to receive weekly injections of placebo or several doses of rilonacept for 12 weeks showed only modest efficacy.\textsuperscript{134} Rilonacept is approved for cryopyrin-associated periodic syndrome (CAPS) disorders: FCAS and MWS.\textsuperscript{135} This agent’s effect on controlling gout attacks is currently being investigated in various clinical trials.

### Canakinumab

Canakinumab is a human monoclonal antibody that selectively targets interleukin-1\( \beta \). It has no cross-reactivity with other members of the interleukin-1 family, including interleukin-1\( \alpha \). It is approved for the treatment of CAPS.\textsuperscript{136} Clinical trials with canakinumab in patients with gout are still ongoing.

### Other Considerations

#### Toxicity

Anakinra is generally well tolerated. Injection site reactions are the most frequently reported adverse event. In a randomized clinical trial, injection site reactions were reported in 25% of patients given placebo and in 50%, 73%, and 81% of patients given anakinra in doses of 30, 75, and 150 mg/day, respectively.\textsuperscript{107} These reactions were generally mild and transient. Infections were uncommon and occurred at a similar rate in the placebo and treatment groups. Infections that required antibiotic therapy occurred in 12% of the placebo-treated group and in 15% to 17% of the treatment group. These infections consisted primarily of bacterial events such as cellulitis and pneumonia. The incidence of pulmonary infection appeared to be higher among patients with underlying asthma. In placebo-controlled studies, up to 8% of patients receiving anakinra showed a reduction in the neutrophil count, compared with 2% of placebo patients. Other adverse events reported were headache, nausea, diarrhea, sinusitis, influenza-like syndrome, and abdominal pain. Malignancy rate and incidences were similar to those expected for the populations studied. Long-term follow-up of patients on anakinra has proved the overall tolerability of therapy over several years.\textsuperscript{137}

In animal studies, the combination of TNF inhibition and IL-1 inhibition achieved synergistic efficacy in arthritis models. However, when this approach was tested in RA patients, the combination did not achieve any additional clinical benefit but did result in greater toxicity—specifically, an increased incidence of infection and serious infection.\textsuperscript{138} Therefore, the combination of biologic therapies targeting TNF and IL-1 is currently not recommended.

---

**Figure 63-3** Rilonacept. Structure: Rilonacept is a fusion protein consisting of the human interleukin (IL)-1 receptor extracellular domains and the Fc portion of human IgG1.
**Monitoring**

Patients should be closely monitored for signs and symptoms of infection. Administration of anakinra must be discontinued if a patient develops serious infection. Neutrophil counts should be assessed before anakinra treatment is initiated, as well as monthly during anakinra therapy for 3 months, and then every 4 months for up to 1 year.

**Pregnancy and Breastfeeding**

Reproductive studies have been performed only on rats and rabbits, and they have not revealed any evidence of harm to the fetus. However, no well-controlled studies have been conducted in pregnant women. Therefore, anakinra should be used during pregnancy only if it is clearly needed. It is not known whether anakinra is secreted in human milk, so it should be discontinued in nursing mothers.

**INTERLEUKIN-6**

Research into the molecular mechanism of IL-6 and other members of the IL-6 cytokine family suggests a potentially critical role in inflammatory and immune responses. IL-6 is a small polypeptide characterized by a four–α-helix bundle structure that is stabilized by intramolecular disulfide bridges. IL-6 is secreted by various cell types, including monocytes, T and B lymphocytes, and fibroblasts. It is detectable at elevated levels in the serum and synovial tissue in inflammatory arthritides, including RA and PsA.\(^{139}\) IL-6 exerts its activity by binding its receptor component, IL-6R, which exists in soluble and membrane-bound forms, and the accessory protein, glycoprotein 130 (gp130). The IL-6R is constitutively expressed on several cell types, including lymphocytes and hepatocytes. However, soluble forms of IL-6R can productively interact with the 130 kD signal transducing component gp130, which is expressed on a wide range of cell types (Figure 63-4). IL-6 has multiple effects on various aspects of the immune system to initiate inflammation. IL-6 stimulates the production of T helper (Th)17 cells. These pathogenic cells secrete IL-17 and are involved in the induction of autoimmune injury.\(^{140}\) IL-6 also plays a role in B cell activation and differentiation. Effects of IL-6 on osteoclast differentiation and activation, including receptor activator of NFκB (RANK) ligand–dependent mechanisms, have been clearly demonstrated.\(^{141}\) Recruitment of neutrophils to the inflammatory sites and stimulation of VEGF synergistically with TNF and IL-1β contribute to pannus formation.\(^{142}\) Levels of IL-6 are directly proportionate with levels of CRP and disease severity. IL-6 knockout mice are resistant to CIA and show reduced levels of serum TNF.\(^{143}\) All these functions make IL-6 blockade an attractive biologic target therapy for the treatment of RA and other autoimmune diseases.

**Tocilizumab**

**Structure and Mechanism of Action**

Previously referred to as myeloma receptor antibody (MRA), tocilizumab is a humanized IgG1 monoclonal antibody that binds with high affinity to soluble and membrane-bound forms of the 80-kD component of the IL-6R. Tocilizumab is a recombinant humanized antihuman interleukin-6 (IL-6) receptor monoclonal antibody of the immunoglobulin IgG1\(^{\kappa}\) (gamma 1, kappa) subclass with a typical H\(_2\)L\(_2\) polypeptide structure. Each light chain and heavy chain consists of 214 and 448 amino acids, respectively. The four polypeptide chains are linked intramolecularly and intermolecularly by disulfide bonds. Tocilizumab has a molecular weight of approximately 148 kD. Treatment with this monoclonal antibody effectively inhibits IL-6–mediated interactions on cells constitutively expressing the IL-6R. In addition, as noted, soluble forms of the IL-6R can productively interact with the 130 kD signal-transducing component gp130, which is expressed on a wide range of cell types; thus, treatment with tocilizumab effectively inhibits a broad array of IL-6–driven processes.

**Pharmacokinetics**

Tocilizumab has a nonlinear pharmacokinetic profile.\(^{144}\) The maximum concentration increases in approximate

---

**Figure 63-4** Interleukin (IL)-6. Mechanism of action: IL-6 binds first to the membrane-bound IL-6 receptor (mIL-6R). The IL-6/mIL-6R complex then associates with the signal-transducing membrane protein, gp130.
proportion to increases in dosage, whereas the area under the concentration-time curve increases disproportionately. As the dosage increases, clearance and the apparent elimination rate constant decrease, and terminal half-life and mean residence times are prolonged. Methotrexate therapy, alcohol consumption, age, and race have not been found to affect the pharmacokinetics of tocilizumab. Tocilizumab binds to soluble IL-6R in a dose-dependent manner and saturates the receptor at approximately 0.1 µg/mL. Tocilizumab also competitively inhibits IL-6 binding to soluble IL-6R; complete inhibition is seen at approximately 4 µg/mL.\(^{145}\)

After intravenous dosing, tocilizumab undergoes biphasic elimination from the circulation. In patients with RA, the central volume of distribution was 3.5 L, and the peripheral volume of distribution was 2.9 L, resulting in a volume distribution at steady state of 6.4 L.

**Drug Dose**

Indications for the use of tocilizumab vary across the globe. In the United States, initial approval for tocilizumab in late 2009 was for it to be used alone or concomitant with methotrexate or other DMARDs in treating adult patients with rheumatoid arthritis in whom one or more anti-TNF agents had failed. Recommended starting dose is 4 mg/kg followed by an increase to 8 mg/kg based on clinical response. In other countries, it is recommended to start therapy at 8 mg/kg, with the possibility of reducing to 4 mg/kg, for example, in the case of tolerability concerns. It is administered once every 4 weeks as a 60-minute single intravenous infusion. Doses exceeding 800 mg per infusion are not recommended. Development of a preparation of tocilizumab for subcutaneous administration is under way.

**Efficacy**

An early double-blind, placebo-controlled, randomized clinical trial revealed statistically significant ACR20 responses with 10 mg/kg and 5 mg/kg single infusions of tocilizumab when given to patients with active RA with no concomitant DMARDs.\(^{144,146}\) Substantial decreases in measures of the acute phase response, including the erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) concentration, were seen with 5 and 10 mg/kg as early as 1 week and were largely sustained through week 4. Larger phase II studies tested tocilizumab monotherapy, given at a dose of 4 mg/kg or 8 mg/kg intravenously every 4 weeks over 3 months, to patients with relatively refractory and active RA. Most individual measures of arthritis activity improved to a statistically significant level by week 4, and the extent of response continued to improve through 12 weeks. Sustained clinical responses were maintained for a considerable number of patients over 5 years of an open-label follow-up period.\(^{147}\) In a large European study known as CHARISMA, patients with active RA despite being on MTX for at least 6 months were randomized to receive tocilizumab 2, 4, or 8 mg/kg every 4 weeks alone or in combination with their existing MTX. Another group received MTX alone. Tocilizumab alone at doses of 4 and 8 mg/kg resulted in improvement in ACR20; however, patients remaining on concomitant MTX, together with tocilizumab, achieved statistically significant improvement compared with those given MTX alone at 16 weeks.\(^ {148}\) The ability of treatment with tocilizumab to alter the progression of joint damage was first assessed in patients with relatively early RA in a study called SAMURAI.\(^ {149}\) Therapy with tocilizumab was shown to have a beneficial effect on the progression of radiographic joint damage defined by total Sharp score, in addition to improvements in clinical and functional status.

Tocilizumab has been studied in series of multinational phase III clinical trials involving more than 4000 patients. In the OPTION trial, moderate to severe RA patients who were methotrexate-inadequate responders received 4 mg/kg or 8 mg/kg tocilizumab or placebo every 4 weeks. Fifty-nine percent and 48% of patients who received tocilizumab 8 mg/kg and 4 mg/kg plus methotrexate, respectively, achieved ACR20 at week 24, compared with 27% of patients who received placebo plus methotrexate. Patients in the placebo and 4 mg/kg groups switched to the 8 mg/kg group (rescue therapy), and by week 24, all patients receiving tocilizumab had significantly better responses in all core set values, including DAS28, HAQ, pain visual analogue scale (VAS), and global VAS.\(^ {150}\) In the TOWARD trial, patients with active disease, despite being on stable doses of DMARDs (the most common being MTX) were randomized to receive monthly 8 mg/kg tocilizumab or placebo infusions. At 24 weeks, the tocilizumab group showed significantly improved ACR20 response rates and DAS28 scores, as well as HAQ and fatigue scores.\(^ {151}\) The AMBITION trial included patients who were methotrexate-naïve or who had discontinued methotrexate for reasons other than lack of efficacy or toxicity. Seventy percent of patients who received tocilizumab 8 mg/kg achieved ACR20 at week 24, compared with 53% of patients receiving methotrexate alone. ACR50 and -70 response rates, HAQ scores, and DAS28 remission were also superior in comparison with the MTX group.\(^ {152}\) The efficacy of tocilizumab for the treatment of patients who had an inadequate clinical response or were intolerant to one or more anti-TNF agents was evaluated in a rheumatoid arthritis study in anti-TNF failures (RADIATE). Fifty percent and 30% of patients who received tocilizumab 8 mg/kg or 4 mg/kg plus methotrexate, respectively, achieved ACR20 at week 24, compared with 10% of patients who received placebo plus methotrexate. Of note, the number and type of TNF antagonists previously failed did not appear to impact the ACR20 response rate. DAS28 remission rates and function assessed by HAQ were also shown to have significantly improved in the tocilizumab groups.\(^ {153}\) The Tocilizumab Safety and the Prevention of Structural Joint Damage (LITHE) trial is a long-term trial designed to evaluate the efficacy of tocilizumab for preventing structural damage in patients who had an inadequate response to prior methotrexate therapy. At week 24, 56% and 51% of patients who received tocilizumab 8 mg/kg or 4 mg/kg plus methotrexate, respectively, achieved ACR20 compared with 27% of patients who received placebo plus methotrexate. Mean joint erosion, joint space narrowing, and total Genant-modified Sharp scores indicated that both tocilizumab doses were associated with significant inhibition of radiographic progression from baseline compared with that of placebo (81%, 85%, and 67% in 8 mg/kg, 4 mg/kg, and placebo groups, respectively).\(^ {154}\)
Clinical, functional, and structural remission in patients with advanced RA was demonstrated in a 52-week trial. Forty-two percent of remarkably severe RA patients with long disease duration and progressive joint damage achieved clinical remission. Long-term treatment with tocilizumab up to 4.2 years revealed increasing numbers of patients achieving ACR50 and -70 scores, DAS28 remission, and high retention rates.

**Safety**

A number of safety concerns are associated with blocking a major regulatory cytokine such as IL-6. These can be grouped as general immunomodulatory effects (e.g., infection), IL-6–related effects (e.g., abnormalities in liver enzymes, lipid profiles), and finally agent-specific effects (e.g., infusion reactions) (Table 63-3).

<table>
<thead>
<tr>
<th>Table 63-3</th>
<th>Adverse Effects Potentially Associated with Tocilizumab</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Immunomodulatory Effects</strong></td>
<td></td>
</tr>
<tr>
<td>Infections/serious infections</td>
<td></td>
</tr>
<tr>
<td><strong>Target-related Effects</strong></td>
<td></td>
</tr>
<tr>
<td>Increased liver enzymes</td>
<td></td>
</tr>
<tr>
<td>Abnormalities in lipid profiles</td>
<td></td>
</tr>
<tr>
<td>Neutropenia</td>
<td></td>
</tr>
<tr>
<td>Low platelet count</td>
<td></td>
</tr>
<tr>
<td>Malignancies</td>
<td></td>
</tr>
<tr>
<td>Demyelinating conditions</td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal perforations</td>
<td></td>
</tr>
<tr>
<td><strong>Agent-related Effects</strong></td>
<td></td>
</tr>
<tr>
<td>Administration reactions</td>
<td></td>
</tr>
<tr>
<td>Hypersensitivity reactions</td>
<td></td>
</tr>
</tbody>
</table>

Infections are a concern, as they are with all immunomodulatory therapies for RA. In clinical trials of tocilizumab, the occurrence of infection appears compatible with that of other approved biologic agents. Although the overall infection rate with tocilizumab monotherapy was comparable with rates in methotrexate monotherapy groups, the incidence of overall infection was slightly higher with concomitant DMARD therapy. The most commonly reported infections (5% to 8%) were upper respiratory infections and nasopharyngitis. A similar incidence of serious adverse events (5%) was generally observed across study groups. Serious infections occurred more often in higher-dose groups as compared with the placebo group. Cellulitis, pneumonia, diverticulitis, gastroenteritis, and herpes zoster were the most common infections noted. Only rare cases of opportunistic infection have been reported.

Transient elevations of alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels were commonly observed in tocilizumab-treated patients. Elevations were more apparent immediately following infusions, potentially reflecting blockade of the antiapoptotic properties of IL-6 in hepatocytes. An increase in ALT levels to three or more times the upper limit of normal was noted in 5% to 6.5% of patients who received tocilizumab and concomitant DMARD therapy. These rates were 1.5% and 2.1% in patients who received only DMARD or only tocilizumab, respectively. To date, elevated transaminases have not been associated with reduced liver function or serious adverse events.

Lipid profiles were altered in the tocilizumab group compared with placebo. Tocilizumab was associated with increases in all lipid levels, including total cholesterol and its fractions, low-density lipoprotein (LDL) cholesterol, and high-density lipoprotein (HDL) cholesterol. Increases in these parameters were noted by the first assessment at 6 weeks and remained elevated through the clinical trials. Despite elevations in these parameters, clinical cardiovascular events have not increased in clinical trials. Mean increases in the 8 mg/kg tocilizumab plus DMARD group were 30.9 mg/dL (total cholesterol), 21.7 mg/dL (LDL cholesterol), 4.3 mg/dL (HDL cholesterol), and 30.1 mg/dL (triglycerides).

A higher proportion of patients receiving tocilizumab had a decrease in neutrophil counts (29%) in comparison with patients who were not receiving this drug (4%). The drop in neutrophil counts was generally mild (grade 1, according to the common toxicity criteria, i.e., 1500 to 2000 cells/mm³ or more) to moderate (grade 2, 1000 to 1500 cells/mm³) in severity and reversed with discontinuation of treatment. To date, no clear association between low neutrophil counts and infection-related adverse events has been noted.

Treatment with tocilizumab was associated with a reduction in platelet counts. The decrease in platelet counts was below the lower limit of normal and was reported in 8% to 9% of patients receiving tocilizumab with or without MTX. However, no serious bleeding incidents occurred. Only isolated cases of epistaxis and hemoptysis were reported in patients with moderate to severe thrombocytopenia.

Rare events of gastrointestinal perforation have been reported in clinical trials, primarily as complications of diverticulitis. The overall rate of gastrointestinal perforation was 0.26 event per 100 patient-years. Most patients who developed gastrointestinal perforations were taking concomitant nonsteroidal anti-inflammatory medications (NSAIDs), corticosteroids, or methotrexate. Patients presenting with new-onset abdominal symptoms therefore should be evaluated promptly for early identification of gastrointestinal perforation.

Because tocilizumab is a humanized antibody, infusion-related adverse events might be expected. Adverse reactions associated with the infusion were reported to be 8% and 7% in patients receiving 4 mg/kg and 8 mg/kg tocilizumab together with MTX. The most commonly reported adverse events were hypertension during the infusion and headache and skin reactions within 24 hours after the infusion. These reactions did not result in termination of treatment. Antibodies to tocilizumab were detected in a small group of patients, and very few were associated with medically significant hypersensitivity reactions that led to withdrawal. The development of antibodies and of infusion reactions may be greater among patients receiving the lower dose of tocilizumab.

Rates of malignancy were similar in the tocilizumab- and DMARD-treated groups. Rare cases of demyelinating disorders were reported in clinical trials. Long-term experience with tocilizumab will determine the actual risk of developing these entities; however, caution needs to be exercised.
Drug Interactions

Inhibition of IL-6 may affect cytochrome P450 substrates. In vivo studies showed that omeprazole and simvastatin levels decreased by 28% and 57% 1 week after tocilizumab infusion. Upon initiation or discontinuation of tocilizumab, drugs with a narrow therapeutic margin such as warfarin or drug concentrations of certain drugs such as cyclosporine should be closely monitored.

Tocilizumab has not been studied, and its use should be avoided, in combination with biologic DMARDs such as TNF blockers, IL-1 blockers, anti-CD20 monoclonal antibodies, and co-stimulation blockers.

It is recommended that live vaccines should not be given concurrently with tocilizumab because clinical safety has not been established. It would be desirable for patients to be brought up to date on all recommended vaccinations before treatment with tocilizumab is initiated.

Monitoring

Patients should be evaluated for tuberculosis and other infections before treatment with tocilizumab is begun. Antituberculosis therapy should be initiated before tocilizumab is started in patients with active tuberculosis.

It is not recommended to initiate tocilizumab treatment in patients who have ALT and AST levels greater than 1.5 times the upper level of normal or any evidence of liver disease. For all patients, liver function tests should be repeated every 4 to 8 weeks during tocilizumab treatment. When the ALT or AST level is between one and three times the upper level of normal, the dose of tocilizumab or concomitant DMARD should be adjusted. For persistent increases in this range, the tocilizumab dose should be modified. If the ALT or AST level is greater than three to five times the upper level of normal, tocilizumab should be interrupted until the level falls below three times the upper level of normal. For ALT or AST elevations greater than five times the upper level of normal, the drug should be discontinued.

Patients receiving tocilizumab should have lipid levels monitored with a goal toward maintaining levels within the target ranges of National Cholesterol Education Program Adult Treatment Panel III or local guidelines. Patients should be managed with lipid-lowering agents if appropriate.

Caution should be exercised when tocilizumab is initiated in patients with a very low neutrophil count at baseline, and all patients should have their absolute neutrophil count (ANC) monitored 4 to 8 weeks after the first infusion. It has been recommended that tocilizumab should not be administered to patients with ANC values less than 2000 cells/mm$^3$. If the ANC falls to between 500 cells/mm$^3$ and 1000 cells/mm$^3$, drug therapy should be discontinued until the ANC reaches above 1000 cells/mm$^3$.

It is not recommended to initiate tocilizumab treatment in patients with a platelet count below 100,000/mm$^3$. Treatment with tocilizumab should be interrupted if the platelet count falls to below 50,000/mm$^3$. Platelets should be monitored every 3 to 8 weeks.

Pregnancy and Breastfeeding

No adequate and well-controlled studies have been conducted in pregnant women. Therefore tocilizumab is a Pregnancy Category C drug and should be used only if potential benefits justify potential risk to the fetus. It is not known whether tocilizumab is excreted in human milk or is absorbed after oral ingestion.

SUMMARY

Treatment of patients with RA with inhibitors of key proinflammatory cytokines TNF, IL-1, and IL-6 is a compelling example of effective targeted biologic therapy. Treatment with these agents, particularly TNF inhibitors, has substantially improved the signs and symptoms of disease; for many patients, quality of life has been improved, the progression of joint damage has been inhibited, and disability has been averted. The success of this approach has "raised the bar" for the goals of treating this pernicious disease and has reinvigorated research aimed at further refinements of therapy. It has also stimulated research into the potential utility of inhibitors of additional cytokines such as IL-6, IL-15, and IL-18 and other components of the immune system relevant to autoimmune disease.

A number of questions remain regarding the optimal use of these drugs. Longer-term safety data will allow clinicians to more fully assess the risk-benefit ratio for individual patients. Given uncertainties regarding the long-term safety of these drugs and the heterogeneity of clinical responses, research defining the populations of patients expected to derive the greatest benefit with the least toxicity is critical. For example, ongoing research assessing genetic polymorphisms or proteomic or glycomic differences among treated patients could optimize efficacy while minimizing toxicity. This is also relevant from a cost standpoint. Although the acquisition costs of these agents are relatively high, data supporting their cost-effectiveness, including gains in employment and reduced hospitalizations, are emerging.

The success observed with these biologic therapies has raised additional clinical questions. For example, can very early treatment with highly effective therapy, such as the combination of TNF inhibitor and MTX, truly alter the disease course? What might the optimal treatment paradigms be for various rheumatic diseases? The success of the TNF inhibitors has also generated substantial interest in targeting this cytokine by alternative approaches, including inhibition of key regulatory factors, such as p38 mitogen-activated protein (MAP) kinase and NFκB. Advances in
biopharmaceuticals could generate agents that possess desirable characteristics in terms of pharmacokinetics, immunogenicity, adverse effects, ease of administration, and cost. These developments should eventually allow clinicians to maximize the use of these novel therapies and achieve clinical benefits that previously were considered unattainable.

Selected References


Full references for this chapter can be found on www.expertconsult.com.
References

80. Strand V, Mease P, Burmester GR, et al: Rapid and sustained improvements in health-related quality of life, fatigue, and other patient-reported outcomes in rheumatoid arthritis patients treated with certolizumab pegol plus methotrexate over 1 year: results from the
145. Mihara M, Kasutani K, Okaazi M, et al: Tocilizumab inhibits signals transduction mediated by both miL-6r and siL-6r, but not by the receptors of other members of IL-6 cytokine family, Int Immunopharmacol 5:1731–1740, 2005.
Cell-Targeted Biologics and Targets: Rituximab, Abatacept, and Other Biologics

PETER C. TAYLOR

KEY POINTS

Rituximab is an effective biologic therapy across the spectrum of rheumatoid arthritis (RA)—but with greatest benefit in seropositive patients. Clinical trials in active, established RA confirm that a single cycle of rituximab produces an enduring clinical response comparable with that observed in tumor necrosis factor (TNF) blockade.

Current data suggest that the most appropriate interval between rituximab courses is 6 to 12 months. Repeat treatment produces responses that equal or exceed those from the first course of treatment, with a comparable duration of effect.

Rituximab has an acceptable safety record in RA trials, but infusion reactions of variable severity can occur; most are mild to moderate. Frequency and severity are reduced by the administration of intravenous methylprednisolone before rituximab infusions.

Clinical trials in active, established RA confirm that abatacept can result in a meaningful clinical response within 16 weeks, although additional improvement may occur for a year or beyond.

Abatacept represents an effective biologic therapy with acceptable safety across the spectrum of RA-patient populations. For a majority of those achieving clinical responses in the first 6 months, sustained clinical responses that follow may be incremental up to 2 years of treatment. A newer subcutaneous formulation is equally effective.

The benefit-to-risk profile of abatacept may be most optimal when introduced earlier in the RA treatment paradigm.

Both rituximab and abatacept slow radiographic progression in RA.

Current uncertainties include the lack of reliable biomarkers to inform the rational choice of a biologic agent.

As appreciation of the gravity of the social and economic burden imposed by rheumatoid arthritis (RA) has grown, so has the recognition that more favorable clinical outcomes are achieved when synovitis is optimally suppressed. The evidence is particularly compelling early in the course of RA, when intervention with disease-modifying combination therapy results in improved remission rates and increased clinical and radiographic benefits. The armamentarium of potential therapeutics has also grown with the identification of relevant disease molecules. Of these, biologic therapeutics targeting tumor necrosis factor (TNF), particularly when used in combination with oral methotrexate, have enjoyed notable success in suppressing inflammation and markedly inhibiting the progression of structural damage previously thought to be an unavoidable characteristic of RA. However, despite the unprecedented clinical and commercial successes of TNF inhibitors, their availability is restricted by high costs. In addition, a substantial proportion of RA patients fail to demonstrate significant clinical responses.

An entirely different treatment approach to the blockade of proinflammatory cytokines is the targeting of cells implicated in the persistence of RA. It is believed that immune responses drive the disease process in RA, and because chronicity is a hallmark of the RA phenotype, it presumably reflects the persistence of immunologic memory, induced and maintained by the adaptive immune system. In particular, T and B cells develop highly specific receptors and, after stimulation, expand enormously in number and then persist for long periods. If this is true of aberrant immune responses that lead to disease, then T and B cells represent rational targets for immune intervention. The focus of this chapter is on biologics with specificity for cellular targets, namely cell surface molecules associated with B cell subsets, most notably CD20, and co-stimulation molecules expressed on antigen-presenting cells that recognize cognate ligands on T cells. The particular emphasis is on two drugs that have become an accepted part of the pharmacologic armamentarium for RA treatment: rituximab and abatacept. Rituximab (Rituxan, Genentech Inc. and Biogen Idec Inc.; MabThera, F-Hoffman LaRoche Ltd.) is an antibody that selectively depletes a B cell subset expressing the CD20 antigen. Other biologics targeting this antigen that have been in clinical trials are the humanized monoclonal antibody ocrelizumab and fully human monoclonal antibody ofatumumab. Abatacept (Orencia, Bristol-Myers Squibb) is a fusion protein that selectively modulates a co-stimulatory signal necessary for T cell activation. Rituximab is approved in the United States and Europe for use in combination with methotrexate to reduce the signs and symptoms of RA in adult patients who have moderately to severely active disease and have failed one or more anti-TNF drugs. Abatacept is the first selective co-stimulation modulator to be approved in the United States and Europe for the treatment of RA patients with an inadequate response to other nonbiologic or biologic disease-modifying antirheumatic drugs (DMARDs).
**TARGETING B CELLS**

**KEY POINT**

B cells are major contributors to RA pathogenesis, but their precise roles in the induction and maintenance of pathogenic immune activation remain poorly understood.

The role of B cells in the pathogenesis of RA is not fully understood. Nonetheless, there are many known B cell functions of likely relevance including their role in antigen presentation, secretion of proinflammatory cytokines, production of rheumatoid factor (RF) and thus immune complex formation, and costimulation of T cells. Of note, immune complexes are an important trigger to the production of TNF and other proinflammatory cytokines. B cells are also implicated in the process of ectopic lymphoid organogenesis in the rheumatoid synovium.

B cells arise from stem cells in the bone marrow, where they acquire an antibody receptor bearing a unique variable region. A number of maturation and activation steps take place as the B cells migrate from the marrow compartment, through blood, and to perifollicular germinal centers and memory compartments in lymphoid tissue, before returning to the marrow as mature plasma cells. Successful maturation and survival of cells are tightly regulated and dependent on a number of trophic signals delivered via cell surface ligands such as vascular cell adhesion molecule-1 (VCAM-1) and soluble factors such as B lymphocyte stimulator (BlyS). \(^\text{7,8}\)

In the late 1990s, Edwards and colleagues \(^\text{9,10}\) suggested that the (assumed) underlying autoreactive response in RA might be driven by self-perpetuating B cells and that the initiation of inflammation results from ligation of the low-affinity immunoglobulin G (IgG) receptor FcγRIIa by immune complexes. An attractive feature of this hypothesis, particularly in seropositive patients, is that it might account for the tissue tropisms of disease expression in the RA syndrome complex because FcγRIIa is expressed in high levels in synovium and other extra-articular tissues that may be involved in RA. RF-producing cells can capture antibodies bound to antigen before antigen internalization by endocytosis and subsequent presentation of peptide fragments to a T cell, with provision of T cell help to the B cell. Edwards and colleagues \(^\text{11}\) also proposed that such RF-producing B cells might become self-perpetuating by an amplification signal arising from co-ligation of the B cell receptor and small immune complexes formed by IgG RF bound to the complement component C3d, providing a survival signal. In contrast, co-ligation of certain other B cell surface receptors with the B cell receptor may provide a negative survival signal. In the rare event that self-perpetuating, autoreactive B cells arise, having escaped normal regulatory mechanisms, this theory predicts that a B cell depletion strategy would remove the autoreactive B cell clones and their antibody products. Because CD20 is not internalized and is highly expressed on a range of B lineage cells including pre-B cells, immature B cells, activated cells, and memory cells, but is not found on stem, dendritic, or plasma cells (Figure 64-1), it is an ideal target for B cell depletion by monoclonal antibodies. The hypothesis that B cells represent a therapeutic target in RA is also being tested in the clinic using other strategies for B cell inhibition, as discussed later.

The CD20 antigen is located in the B cell membrane, with 44 amino acids exposed to the extracellular space. Its function is unknown, although it may have a role in cell signaling or in calcium mobilization. \(^\text{12}\) Interestingly, CD20 knockout mice do not have a clear-cut phenotype or obvious B cell defect. \(^\text{13}\) CD20+ B cells represent a prominent population in the rheumatoid synovial tissue in the majority of patients.

**Rituximab in Rheumatoid Arthritis**

**KEY POINTS**

- Rituximab is a depleting chimeric monoclonal antibody lytic for a population of B cells expressing CD20.
- Rituximab is an effective biologic therapy across the spectrum of RA-patient populations.
- It appears to have greatest benefit in seropositive patients.
- Rituximab given as two infusions of 1 g each slows radiographic progression in RA.
- Rituximab has an acceptable safety record in RA trials, but infusion reactions of variable severity can occur; most are mild to moderate. Frequency and severity are reduced by the administration of intravenous methylprednisolone before rituximab infusions.
- A rare complication in RA patients treated with rituximab is progressive multifocal leukoencephalopathy. Because of the increase in relative risk, patients should be counseled appropriately.
- Repeat treatment produces clinical responses that equal or exceed those from the first course of treatment, with a comparable duration of effect.
- Other approaches to B cell–targeted therapy are under investigation, including antibodies that target other B cell surface antigens, neutralization of B cell stimulating factors, and signaling molecules that participate in B cell activation.

Rituximab is a chimeric mouse-human monoclonal antibody directed against the extracellular domain of the CD20 antigen. It initiates complement-mediated B cell lysis and may permit antibody-dependent, cell-mediated cytotoxicity when the Fc portion of the antibody is recognized by corresponding receptors on cytotoxic cells. Rituximab may also
initiate apoptosis and influence the ability of B cells to respond to antigen or other stimuli. Rituximab initially found a role in the clinic as a single-agent treatment for relapsed or refractory low-grade or follicular CD20+ B cell non-Hodgkin’s lymphoma, for which it was approved. For this reason, there was a wide experience with rituximab in hematologic oncology before clinical trials in RA and its recent approval in the United States and Europe for the treatment of TNF inhibitor–refractory RA patients with active disease.

Following rituximab administration, rapid B cell depletion takes place in peripheral blood and conventional methods for measurement of peripheral blood B cells by means of CD19 expression detect no cells at all in the majority of cases. Investigation of synovial tissue from patients with RA treated with rituximab reveals a decrease in synovial B cells and plasma cells in most but not all patients. These findings raise the possibility that there might be as yet poorly understood rescue mechanisms in the inflamed synovium providing survival niches or, alternatively, that some B cells may have an inherent resistance against depletion.

Analysis of peripheral blood memory B cells before depletion and during and after reconstitution seems to be predictive of clinical outcome, with patients showing early relapses having substantially higher IgD+ and IgD-CD27+ memory B cell numbers and proportion during B cell recovery. In addition to these cellular biomarkers, serologic parameters have been analyzed. Decreases in RF or anti-citrullinated protein antibody serum levels are reported to be associated with B cell depletion, but further studies are necessary to determine the relationships between these serologic changes and clinical response.

Clinical Studies

Clinical trials in active, established RA confirm that a single cycle of rituximab given as two infusions of 1 g each, together with once-weekly oral methotrexate, produces an enduring clinical response. A treatment cycle of two infusions of 500 mg is also efficacious but may result in a lower proportion of patients demonstrating more robust clinical responses.

The findings of early clinical studies of B cell depletion therapy in patients with active RA using rituximab in a number of different treatment regimens suggested an encouraging benefit with an acceptable safety profile and pointed to a possible therapeutic role for rituximab in RA. Confirmation of benefit, however, required a randomized, double-blind, controlled study. In a phase IIa study, the efficacy of rituximab in active RA was tested in 161 patients who had failed to respond adequately to methotrexate at a dose of at least 10 mg a week for a minimum of 16 weeks. Patients were assigned to one of four treatment regimens: 1-g infusion of intravenous rituximab alone on days 1 and 15, methotrexate alone as a comparison arm, intravenous rituximab with cyclophosphamide infusions at a dose of 750 mg on days 3 and 17, or rituximab and methotrexate. All patients received 100 mg methylprednisolone just before each treatment, in addition to prednisolone 60 mg daily on days 2, 4, 5, 6, and 7 and 30 mg daily on days 8 to 14. The primary end point was the proportion of patients achieving an American College of Rheumatology 50% improvement criteria (ACR50) response at week 24, and exploratory analyses were undertaken at week 48. At week 24, a significantly greater proportion of patients achieved an ACR50 in the rituximab and methotrexate combination group (43%; P = .005) and in the rituximab and cyclophosphamide combination group (41%; P = .005) than in the group receiving methotrexate as monotherapy (13%) (Table 64-1). Thirty-three percent of the patients receiving rituximab alone achieved an ACR50 response, but this failed to reach statistical significance compared with methotrexate alone (P = .059). In all the rituximab groups, the mean change from baseline in disease activity score was significant compared with methotrexate alone.

At 48 weeks, exploratory analyses indicated ACR50 and ACR70 responses in 35% and 15%, respectively, of patients in the rituximab and methotrexate group—significantly greater than the 5% and 0% responding at the corresponding levels in the methotrexate group. In the rituximab and cyclophosphamide treatment arm, 27% of patients achieved an ACR50 response.

Rituximab treatment was associated with near-complete peripheral blood B cell depletion, persisting throughout the 24-week period of the primary analysis. Patients in the rituximab groups had a substantial and rapid reduction in the concentration of RF in serum, but despite peripheral B cell depletion, immunoglobulin levels did not change substantially. The overall incidence of infection was similar in the control and rituximab groups at 24 and 48 weeks. By week 24, four patients in the rituximab groups had suffered a serious infection, as well as one in the control group. Two additional serious infections were reported during the extended 48-week period in the rituximab groups, one of which was fatal. Infusion reactions of any type were reported in 36% of patients receiving rituximab and 30% of patients receiving placebo, although most were characterized as mild or moderate. The reactions included hypotension,
hypertension, flushing, pruritus, and rash. In the rare case of severe reactions, it has been suggested that a cytokine release syndrome associated with marked cell lysis following rituximab might be a contributing factor.

In summary, the findings of the phase IIa study indicated that a single course of treatment with rituximab, particularly in combination with methotrexate, produces an enduring response in patients with severe, seropositive, active RA. Further, treatment with rituximab was well tolerated, with a favorable safety profile over 48 weeks of follow-up.

To follow up the phase IIa study, a phase IIb study was undertaken to examine the efficacy and safety of rituximab at different doses, with or without glucocorticoids, in patients with active RA resistant to DMARDs including biologics. The findings of this phase IIb study, known as the DANCER (Dose-ranging Assessment: International Clinical Evaluation of Rituximab in RA) trial, were reported in 2006.24 The researchers recruited 465 patients with active disease who had to have failed to respond to at least one DMARD other than methotrexate, but no more than five, or failed to respond to biologic response modifiers, and they had to have been treated with methotrexate as a single DMARD for at least 12 weeks, with 4 weeks of stable therapy at a dose of at least 10 mg a week. All other DMARDs were withdrawn at least 4 weeks before randomization—8 weeks for infliximab, adalimumab, and leflunomide. Patients were randomized to receive either placebo infusions or rituximab at a dose of 500 mg or 1 g on days 1 and 15, together with one of three glucocorticoid options: glucocorticoid placebo, 100 mg of intravenous methylprednisolone before each rituximab infusion, or 100 mg of methylprednisolone before each infusion in addition to an oral corticosteroid. The results at 24 weeks confirmed the significant efficacy of a single course of rituximab in active RA when combined with methotrexate. This benefit was independent of glucocorticoids, although methylprednisolone on day 1 reduced the incidence and severity of first rituximab infusion reactions by about one-third (Figure 64-3). Both rituximab doses were efficacious. At the lower dose, 55% of recipients achieved ACR20 responses, as did 54% of those at the higher rituximab dose—in both cases, significantly greater than the 28% of those receiving placebo infusions. Similarly, significantly higher proportions of patients achieved ACR50, ACR70 (see Table 64-1), and European League Against Rheumatism (EULAR) good responses at 24 weeks at both rituximab doses compared with patients receiving placebo infusions (Figure 64-3). At the most stringent ACR70 response level, the difference in the percentage of responders in the placebo, lower-dose rituximab, and higher-dose rituximab groups was most marked at the higher rituximab dose of 1 g 2 weeks apart (5%, 13%, and 20%, respectively; P < .05). Adverse events reported up to 24 weeks were largely infusion related, particularly at the time of the first infusion.

A trial, known by the acronym REFLEX (Randomized Evaluation of Long-term Efficacy of Rituximab in RA), was designed to determine the efficacy and safety of rituximab when used in combination with methotrexate in patients with active RA who have an inadequate response to one or more anti-TNF therapies because of either lack of efficacy (90% of patients recruited) or toxicity (10% of patients recruited); in addition, all patients had

---

**Figure 64-2** DANCER Infusion reactions. Chart showing the occurrence of infusion reactions in the DANCER (Dose-Ranging Assessment: International Clinical Evaluation of Rituximab in RA) trial, a phase IIb study undertaken to examine the efficacy and safety of rituximab at two doses, with or without glucocorticoids (GCs), in patients with active rheumatoid arthritis resistant to disease-modifying anti-rheumatic drugs including biologics. Patients were randomized to receive either placebo infusions or rituximab at a dose of 500 mg or 1 g on days 1 and 15, together with one of three glucocorticoid options: glucocorticoid placebo, 100 mg intravenous methylprednisolone before each rituximab infusion, or 100 mg methylprednisolone before each infusion in addition to oral corticosteroid. Pretreatment with methylprednisolone before rituximab infusion reduced the incidence and severity of reactions by about one third. (Based on data in Emery P, Fleischmann R, Filippowicz-Sosnowska A, et al [for the DANCER study group]: The efficacy and safety of rituximab in patients with active rheumatoid arthritis despite methotrexate treatment: results of a phase IIb randomized, double-blind, placebo-controlled, dose-ranging study, Arthritis Rheum 54:1390–1400, 2006.)
Of the patients assigned to rituximab, 82% completed 6 months, compared with only 54% of the patients assigned to placebo. The major reason for study withdrawal was lack of response, reported in 40% of the placebo group and 12% in the rituximab group. At 6 months, significantly more patients receiving rituximab achieved ACR20, ACR50, and ACR70 responses than did those receiving placebo: 51%, 27%, and 12%, respectively, on rituximab, versus 18%, 5%, and 1% of those on placebo (see Table 64-1). In terms of change in disease activity score (DAS28), intention-to-treat analyses showed that in patients administered placebo infusions, the reduction from baseline was 0.34, less than the 0.6-point reduction considered to be clinically meaningful; in contrast, the reduction was 1.83 in the rituximab group.\textsuperscript{25}

The ACR response evaluates RA treatment on the basis of a 20%, 50%, or 70% improvement in five of seven core components. However, from the patient’s perspective, determining the actual benefit of an ACR20 improvement is not straightforward. In the REFLEX study, there were significantly greater improvements in all components of the ACR core measures in the rituximab group. Rituximab demonstrated a clinically meaningful benefit for RA patients in physical function evaluated by a health assessment questionnaire (HAQ) in all nonoverlapping ACR response categories. In the active treatment arm, both clinical and subjective parameters of the ACR core components contributed to the assignment of an ACR20 response, whereas in the placebo group, the subjective parameters dominated.\textsuperscript{25}

In the REFLEX study, following a single treatment course, the maximal clinical response to rituximab plus methotrexate was observed at 24 weeks. After this time, patients were eligible to exit the study and receive further rituximab treatment on the basis of clinical need. Of the patients in the rituximab plus methotrexate group, 37% (114 of 308) remained in the study over 48 weeks, indicating continued radiographic evidence of at least one joint with definite erosion attributable to RA (Figure 64-4). The recruited cohort comprised 520 patients with mean disease duration of 12 years on a background regimen of methotrexate 10 to 25 mg once a week. After a washout period during which other DMARDs and anti-TNF drugs were withdrawn, patients were randomized to receive a single course of 1 g rituximab or placebo infusions on days 1 and 15. All patients were given 100 mg intravenous methylprednisolone before each infusion and a brief course of oral prednisolone between the two doses: 60 mg daily from days 2 to 7 and 30 mg daily from days 8 to 14.\textsuperscript{25}

Of the patients assigned to rituximab, 82% completed 6 months, compared with only 54% of the patients assigned to placebo. The major reason for study withdrawal was lack of response, reported in 40% of the placebo group and 12% in the rituximab group. At 6 months, significantly more patients receiving rituximab achieved ACR20, ACR50, and ACR70 responses than did those receiving placebo: 51%, 27%, and 12%, respectively, on rituximab, versus 18%, 5%, and 1% of those on placebo (see Table 64-1). In terms of change in disease activity score (DAS28), intention-to-treat analyses showed that in patients administered placebo infusions, the reduction from baseline was 0.34, less than the 0.6-point reduction considered to be clinically meaningful; in contrast, the reduction was 1.83 in the rituximab group.\textsuperscript{25}

The ACR response evaluates RA treatment on the basis of a 20%, 50%, or 70% improvement in five of seven core components. However, from the patient’s perspective, determining the actual benefit of an ACR20 improvement is not straightforward. In the REFLEX study, there were significantly greater improvements in all components of the ACR core measures in the rituximab group. Rituximab demonstrated a clinically meaningful benefit for RA patients in physical function evaluated by a health assessment questionnaire (HAQ) in all nonoverlapping ACR response categories. In the active treatment arm, both clinical and subjective parameters of the ACR core components contributed to the assignment of an ACR20 response, whereas in the placebo group, the subjective parameters dominated.\textsuperscript{25}
clinical benefit after the single initial treatment course. The majority of patients who withdrew did so to receive further courses of rituximab between weeks 24 and 48 of the study. In contrast, 89% of the placebo plus methotrexate group (185 of 209) withdrew before week 48.26

Another recently reported phase III study known by the acronym SERENE (Study Evaluating Rituximab’s Efficacy in methotrexate iNadequate rEsponders) confirmed the benefits of rituximab in RA patients receiving concomitant methotrexate who had active disease at baseline despite methotrexate therapy and who had not received prior biologic therapy.27 Patients were randomized to receive either placebo or rituximab at one of two doses, 2 × 500 mg and 2 × 1000 mg. From week 24, in an open-label extension, those patients in the rituximab arms failing to achieve remission as assessed by DAS28 received a second course of rituximab and all placebo patients were treated with the lower rituximab dose. At week 24, a significantly greater proportion of patients receiving rituximab 2 × 500 mg or 2 × 1000 mg + methotrexate achieved the primary end point of an ACR20 response versus patients receiving placebo plus methotrexate (54.5% and 50.6% vs. 23.3%, respectively; P < .0001). By week 48 approximately 90% of patients in all treatment groups had received a second course of treatment. The majority of these repeat treatments (82% to 88%) were given by week 30.

In rituximab-treated patients, efficacy outcomes at week 48 were comparable with those at week 24; additionally, improvement was observed for several clinically important end points including an approximate doubling in the proportion of patients achieving low disease activity (DAS28 < 3.2) in the rituximab (2 × 1000 mg) plus methotrexate dose group from week 24 to week 48. The safety profile of rituximab in the SERENE study was comparable with that observed in earlier trials, with infusion-related reactions—most nonserious—at the time of first infusion being the most common adverse event. Reductions in immunoglobulin levels were observed, predominantly IgM, but with mean levels remaining within normal limits. No relationships between infectious complications and reduced immunoglobulin levels were observed, and in fact the rate of infection observed in patients receiving rituximab plus methotrexate was low and comparable with patients receiving methotrexate alone over the placebo-controlled 24-week period. This low rate of serious infection continued throughout the full 48-week period, with no obvious difference between the rituximab doses.

Results from the IMAGE study, which included 748 patients without prior use of either biologics or methotrexate, were recently reported. In this group, high-risk patients who had baseline high DAS28 scores or high CRP were found to have greater DAS28 improvement at week 52 if they received rituximab in addition to methotrexate.28

Disease Modification

The extraordinary success of TNF blockade in inhibiting structural damage to joints in patients with RA has set a new standard to which all new biologics must aspire. Inhibition of structural joint damage by rituximab in patients with RA and a previous inadequate response to TNF inhibitors from the REFLEX study was first described over a 1-year period29 with a subsequent demonstration that the initial effects of rituximab are maintained over an extended interval of 2 years, with all measures of joint damage significantly improved compared with placebo plus methotrexate.30 At week 56,31 the mean change in Genant-modified Sharp score in the placebo plus methotrexate arm was 2.31, compared with 1.0 in the rituximab plus methotrexate group (P = .0043). Significant differences were also reported for joint space narrowing and bone erosions. At week 104, significantly lower changes in total Genant-modified Sharp score (1.14 vs. 2.81; P < .0001), erosion score (0.72 vs. 1.80; P < .0001), and joint space narrowing scores (0.42 vs. 1.00; P < .0009) were observed with rituximab plus methotrexate versus placebo plus methotrexate.32 Importantly, within the rituximab group, 87% who had no progression of joint damage at 1 year remained nonprogressive at 2 years. Thus these data confirm that rituximab plus methotrexate has the benefit of sustained inhibition of joint damage progression in patients with RA and a previously inadequate response to TNF inhibitors.

Radiographic outcomes were recently reported for a phase III study designed to determine the efficacy of rituximab in the prevention of joint damage and its safety in combination with methotrexate in the context of patients initiating treatment with methotrexate.33 This study, known by the acronym IMAGE (International study in Methotrexate-naïve subjects investiGating rituximab’s Efficacy) was a randomized, controlled, double-blind trial involving 748 methotrexate-naïve patients assigned to receive rituximab at doses of either 2 × 1000 mg or 2 × 500 mg every 24 weeks in combination with methotrexate or methotrexate alone. The primary end point was radiographic progression measured by total modified Sharp score at week 52. In patients treated with 2 × 1000 mg rituximab and methotrexate, significantly smaller change (0.359) in modified total Sharp score (mTSS) was observed compared with patients on methotrexate alone (1.079; P = <.0001). Furthermore, a significantly higher proportion of patients treated with MabThera and methotrexate had no progression in their joint damage over 1 year (64% vs. 53%; P = .0309). By week 52, 65% of these patients achieved a 50% improvement in symptoms (ACR50) while 47% had achieved a 70% improvement (ACR70), compared with 42% and 25% on methotrexate alone (P < .0001 for both ACR50 and ACR70 comparisons).

Safety Issues

The rapidity and magnitude of peripheral blood B cell depletion following anti-CD20 therapy raise concerns about potential adverse sequelae. The peripheral compartment recovers after many months, but repopulation is predominantly with an immature and naïve subset of B cells. But it must also be remembered that the circulation contains less than 2% of total B cells.34 A common concern regarding all therapies directed at B cells is the potential for toxicity related to modulation of humoral immunity. Unlike other newly introduced biologic therapies for RA, rituximab has the considerable advantage of an oncology safety database based on more than 350,000 non-Hodgkin’s lymphoma patient treatments since 1997.35 The overall safety conclusions are that serious adverse events are infrequent and
often associated with well-defined risk factors such as cardiopulmonary disease or a high number of circulating cancer cells. Of note, in the lymphoma population, prolonged peripheral B cell depletion has not been associated with cumulative toxicity or increased occurrence of opportunistic infections.\textsuperscript{23-35} However, it cannot be assumed that the toxicity profile will be identical in distinct disease phenotypes with differing pathogenic processes.

In RA open-label,\textsuperscript{36} phase II,\textsuperscript{23,24} and phase III\textsuperscript{35} studies, although decreases in total serum immunoglobulin levels were observed in patients receiving rituximab, concentrations remained within normal limits. Of note, existing antibody titers against tetanus toxoid appear to be unaffected by a single course of rituximab treatment.\textsuperscript{37} However, there is some anecdotal evidence that total serum immunoglobulin concentrations fall below the normal range in patients receiving multiple cycles of rituximab treatment over a number of years in open-label studies.\textsuperscript{31} It is unclear whether this results in an increased risk of infection. In phase II studies, most adverse events were mild to moderate and associated with infusions including headache, nausea, and rигors. In a recent meta-analysis of randomized clinical trial data from three studies reporting adverse events arising after a single cycle of rituximab treatment in a total of 938 RA patients refractory to nonbiologic DMARDs or biologic anti-TNFs,\textsuperscript{23-25} it was calculated that the incidence of patients experiencing adverse events of all systems was not higher in the rituximab-treated patients than in the placebo groups (relative risk [RR], 1.062; 95% confidence interval [CI], 0.912 to 1.236, \textit{P} = .438).\textsuperscript{38}

In the DANCER trial, adverse events associated with rituximab were largely associated with the first infusion; these occurred in 39% of patients treated with 500 mg (without steroid) and in 46% receiving 1 g, compared with 17% administered placebo infusions.\textsuperscript{24} The corresponding incidence with the second infusion decreased to 5%, 8%, and 10%, respectively. Two serious infusion reactions, hypersensitivity and generalized edema, occurred on day 1. Pretreatment with methylprednisolone reduced the incidence and severity of reactions by about one-third (see Figure 64-2). Infectious adverse events (largely upper respiratory tract infections) were reported in 28% of placebo and 35% of rituximab patients. There were six serious infections: two in the placebo group, four in patients receiving 1000 mg rituximab, and none in patients receiving 500 mg rituximab. No opportunistic infections or tuberculosis reactivations were reported.

Although the overall safety record based on trial data has been favorable, with wider clinical use of biologic B cell depletion, rarer serious complications have come to light. A potential association has been reported between the biologic therapies efalizumab, natalizumab, and rituximab and the rare, progressive, and usually fatal condition progressive multifocal leukoencephalopathy (PML), a rare brain disease caused by reactivation of the JC virus.\textsuperscript{39} PML has been reported in patients receiving rituximab for hematologic conditions and systemic lupus erythematosus (SLE)\textsuperscript{40} and more recently in RA.\textsuperscript{39} The cumulative incidence rate of PML in the RA population has been estimated at 1/100,000 RA admissions in an analysis limited to hospitalized patients with SLE and other rheumatic diseases (including 25 patients with RA), a majority of whom had concomitant risk factors including human immunodeficiency virus (HIV), malignancy, or transplantation of bone marrow or another organ.\textsuperscript{41} The cumulative reporting rate of 2.2 cases of PML per 100,000 RA patients treated with rituximab is more than double the estimated frequency in RA at 2.2 (95% CI, 0.3 to 8.0).\textsuperscript{39} Although the absolute risk is small, the relative risk is such that it emphasizes the importance of providing the prospective patient being considered for B cell–depleting therapy with thoughtful and balanced information about likely benefits, as well as common to rare complications.

The profound and enduring peripheral B cell depletion that accompanies use of rituximab raises a potential safety concern for those patients in clinical practice who fail to derive adequate symptomatic benefit and may then be exposed to biologic DMARDs of alternative mechanism of action at a time point before repopulation of circulating B cells can take place. Relatively little data regarding this circumstance exist to date, but there is preliminary information in 185 of 2578 RA patients who went on to receive an alternative mechanism of action biologic as documented in a safety follow-up period following participation in trials in which they had previously received rituximab.\textsuperscript{42} Of the 185 patients 89% remained peripherally B cell depleted at the point of commencing a new mechanism of action biologic. The rate of serious infectious events reported following rituximab but before second biologic exposure was 6.99 per 100 patient-years, comparable with the reported rate of 5.49 per 100 patient-years after exposure to a second biologic agent, the majority of which were TNF inhibitors. No fatal or opportunistic infections were observed, the nature and course of infectious complications being within expectation for RA patients on biologic therapy. In a population of methotrexate-naïve patients entering the IMAGE study,\textsuperscript{28} safety data were consistent with results from previous rituximab clinical trials and further enhanced the robust safety profile. Rates of serious adverse events and serious infections were similar between the two MabThera groups and the methotrexate-only group.

RA patients, particularly those treated with immunosuppressants, are at an increased risk of infection. For this reason, vaccination is an important aspect of RA clinical management. As to whether B cell–depleting therapy could adversely affect immunization responses by suppressing the antibody response from new vaccination or reducing preexisting antibody from prior vaccination has been investigated, and the findings were recently reported.\textsuperscript{41-45} These studies indicate that vaccine responses to some, but not all, vaccinations may be diminished in rituximab-treated patients, most strikingly in the first 4 to 8 weeks following rituximab administration.\textsuperscript{43} Although prior vaccination and timing of vaccine administration after rituximab infusion may influence the ability to mount a response, there is no straightforward relationship between peripheral B cell reconstitution and response to immunization. Therefore where vaccination is indicated, it should ideally be given before rituximab treatment and avoided immediately after B cell depletion with a delay of several months. Of course, this may not be practical for vaccines with seasonal availability (such as those to influenza variants), and it must be recognized that responses to vaccination do not necessarily correlate with the risk of infection. In general, the timing
of rituximab administration must be determined according to clinical need without requirement to delay until after supplies of a seasonally variable vaccine become available and the patient has been immunized.

In a recent meta-analysis of six studies enrolling 2728 patients, nonmelanotic skin carcinoma occurred more commonly than other carcinomas in the rituximab treatment groups (0.8% in rituximab group vs. none reported in the control group). Moreover, overall incidence of malignancies was higher in the rituximab group (2.1%) compared with the control group (0.6%).

The safety and efficacy profile of rituximab in the treatment of RA discussed so far is based on reports from randomized placebo-controlled trials of 6 to 12 months’ duration. Open-label extension studies have analyzed safety and efficacy results over multiple courses of rituximab. In safety analyses based on 5013 patient-years of rituximab exposure, in a total of 2578 RA patients who received at least 1 course of rituximab, infusion-related reactions were the most common adverse event occurring in a quarter of patients during the first infusion of the initial treatment cycle, although only 1% of infusion reactions in total were considered to be of a serious nature. Importantly and reassuringly, the rates of both adverse events and serious adverse events were stable over infusion cycles, the later reported as 17.85 events per 100 patient-years (95% CI: 16.72, 19.06). Infections and serious infections over time remained stable across five treatment cycles at 4 to 6 events per 100 patient-years. There were no cases of tuberculosis, disseminated fungal infections, or other serious opportunistic infections during the analysis period. Viral reactivation is a potential concern in immunosuppressed patients. The rate of herpes zoster infections in this large series was 0.98 events per 100 patient-years, similar to that reported for other RA populations. The much more serious and rare opportunistic infection, PML occurred in a single patient who also received cancer chemotherapy and radiation. The event occurred about 18 months after the last dose of rituximab and 9 months after receiving chemotherapy and radiation. The causal relationship to rituximab, if any, is thus not entirely clear. There was no increased risk of malignancy by comparison with reference patients with RA and with the general population in the United States. Myocardial infarction was one of the most common serious adverse events reported in the longer-term analyses at a rate of 0.56 per 100 patient-years, but this is consistent with rates reported in epidemiologic studies of patients with RA.

**Duration of Benefit**

Among RA patients achieving clinical responses to rituximab treatment, the time to clinical relapse is heterogeneous. In some patients, relapse is closely correlated to the reappearance of peripheral blood B cells, but in others, it may be delayed by years. Clinical relapse is more closely associated with increases in autoantibody levels, but better biomarkers are necessary to reliably inform optimal management strategies on an individual basis. All B cell populations are depleted following rituximab therapy. Of residual B lineage cells, more than 80% exhibit a memory or plasma cell precursor phenotype. B cell repopulation occurs at a mean of 8 months after rituximab therapy and depends on the formation of naïve B cells of an immature phenotype resembling those found in umbilical cord blood. Peripheral B cell depletion is accompanied by substantial increases in blood BlyS concentrations, which tend to fall with B cell repopulation. BlyS is a naturally occurring protein required for the development of B lymphocytes into mature plasma cells. Elevated levels of BlyS in RA are believed to contribute to the production of autoantibodies. However, in cases of prolonged clinical responses to rituximab, more gradual reductions in BlyS concentrations have been observed, extending beyond the period of B cell depletion. Thus BlyS may contribute to the survival or regeneration of pathogenic, autoreactive B cells. This hypothesis predicts a potential therapeutic role for BlyS blockade in addition to B cell depletion.

Information concerning the efficacy and safety of repeated cycles of rituximab treatment has emerged from experience in clinical practice and from randomized trials. The recently reported phase III MIRROR trial (Methotrexate Inadequate Responders Randomised study Of Rituximab) was a randomized, double-blind, international study to evaluate the efficacy and safety of three dosing regimens of rituximab in combination with methotrexate in 375 patients with active RA and an inadequate response to methotrexate. Patients were randomized to three groups with two courses of rituximab treatment at varying doses: Group A: all courses 500 mg rituximab on days 1 and 15; repeat treatment at 24 weeks; group B: first course 500 mg rituximab on days 1 and 15; second course 1000 mg rituximab; repeat treatment at 24 weeks; group C: all courses 1000 mg on days 1 and 15; repeat treatment at 24 weeks. The primary end point was the proportion of patients achieving ACR20 at week 48. Secondary end points included ACR50, ACR70, and EULAR responses. There was a trend toward better efficacy results with the regular 2 × 1000 mg dose compared with the low 2 × 500 mg dose, and this reached statistical significance for EULAR good/moderate response (2 × 1000 mg = 88% vs. 2 × 500 mg = 72%; P < .05). Other end points, although numerically superior at 48 weeks, did not show any statistically significant difference between the three dosing regimens.

**Current Role**

Rituximab is generally considered as a cost-effective biologic option in RA patients, particularly if seropositive, with inadequate responses to TNF inhibitors. Recent advances in the understanding of the pathogenesis of RA emphasize the critical role of B cells in self-sustaining chronic inflammatory processes. Rituximab is a promising addition to the therapeutic armamentarium for the treatment of RA. In current clinical practice, the major use for rituximab in the treatment of RA is confined to the TNF inhibitor–refractory population. Data from a number of clinical trials (IMAGE, MIRROR, SERENE, REFLEX, and DANCER) suggest that seropositive RA patients (RF and/or anticyclic citrullinated peptide [anti-CCP]) show a higher likelihood of response to B cell–depleting therapy compared with seronegative patients, in particular for improving signs and symptoms and inhibition of radiographic changes. However, it is nevertheless the case in both trials and clinical experience that a proportion of seronegative RA patients
show good clinical responses, although this proportion is less than in the case of seropositive patients. In pooled analyses of data from the MIRROR and SERENE studies, at week 48, odds ratios for seropositive patients versus seronegative patients of achieving ACR20, ACR50, and ACR70 responses were 2.23 (95% CI, 1.38 to 3.58), 2.72 (95% CI, 1.58 to 4.70), and 3.29 (95% CI, 1.40 to 7.82), respectively. These observations generate the hypothesis that other mechanisms may account for lower levels of response in seronegative patients such as antigen presentation, co-stimulation, and cytokine drive, whereas high levels of response to rituximab therapy may be mediated primarily by the suppression of pathogenic antibodies. In contrast to RF, the type 1 interferon (IFN) signature is associated with a negative response to rituximab therapy in RA patients.

The optimal and most cost-effective dosing regimen for rituximab remains a matter of debate. The phase III SERENE study showed equal clinical efficacy for the 500 mg × 2 and 1000 mg × 2 rituximab doses, but the phase III MIRROR trial had differences in some outcomes favoring the higher dosage. Methotrexate-naive patients (not an approved patient population for rituximab) were studied in the IMAGE study with clinical results that were equivalent, but with a radiographic result that favored the higher dosage. Thus a summary of the current randomized controlled trial data on rituximab dosing is that 1000 mg × 2 works well in a clinically meaningful proportion of patients, but not in all. The 500 mg × 2 rituximab dose achieves broadly similar results in the relevant patient populations overall and has the advantages of lower cost and possibly a lower rate of serious adverse events but perhaps with lower probability of high-impact clinical responses and inhibition of structural damage. On the basis of the DANCER study findings, it is recommended that each cycle of 1000 mg × 2 rituximab be given in combination with once-weekly methotrexate, usually at doses of at least 15 mg/week, to optimize efficacy. Further, administration of 100 mg intravenous methylprednisolone is recommended before each rituximab infusion to reduce the frequency and severity of infusion reactions.

Rituximab may also have a role in those patients for whom TNF blockade is relatively contraindicated such as those with connective tissue disease overlap syndromes. At present, there are uncertainties about the implications of long-term peripheral B cell depletion and the timing and need for redosing with rituximab in patients who respond. Current research suggests that restoration of peripheral B cell numbers takes about 8 months after depletion treatment, although retreatment may be necessary earlier. Results have been presented for an open-label study to evaluate the response to repeated courses of rituximab in patients with active RA participating in one of several phase II or III studies and to determine the optimal frequency for repeated treatment. In a series of 155 patients with prior exposure to TNF inhibitors, ACR20, ACR50, and ACR70 scores were 65%, 33%, and 12%, respectively, following the first course; they were 72%, 42%, and 21%, respectively, for the second treatment course, relative to the original baseline. In 82 of these patients who received a third course of rituximab, the median interval between first and second courses was similar to that between second and third courses: 30 to 31 weeks.

Further studies are necessary to identify the optimal regimen for maintenance therapy that will provide efficacy and limit toxicity. Development of biomarkers informative of management decisions that would optimize response is a highly desirable goal, but as yet none are in routine use. The magnitude of clinical response appears to be related to the completeness of peripheral B cell depletion. This holds true whether the lower 500 mg × 2 dose schedule or higher 1000 mg × 2 schedule is administered. By means of sensitive measurements permitting detection of low numbers of preplasma B cells in the circulation, a recent study reported that RA patients whose disease did not respond to an initial cycle of rituximab have higher circulating preplasma cell numbers at baseline and incomplete depletion. Furthermore, an additional cycle of rituximab administered before total B cell repopulation enhances B cell depletion and clinical responses.

Although the available safety data for rituximab in RA are reassuring, they need to be interpreted with caution until larger numbers have been treated and long-term safety and retreatment data become available. Further, there is a substantial body of safety data for rituximab in the treatment of non-Hodgkin's lymphoma, with similarly low infection rates reported. In oncology, some of the associated adverse events are related to circulating tumor loads. Overall, this is reassuring with regard to the RA population, although close monitoring of immunocompetence and for the possibility of rare opportunistic infections is advisable.

**Future Directions and Other Approaches to B Cell–Targeted Therapy**

Rituximab is currently indicated for the treatment of patients with moderate to severe RA who show no response, experience a loss of response with time, or have adverse effects to anti-TNF alpha agents. The findings of recently reported studies indicate that rituximab is also efficacious in a proportion of both treatment-naive and methotrexate patients with RA, particularly if seropositive. Therefore in the face of competitive health economic data compared with TNF inhibitors, there has been interest in its potential as a first-line biologic. However, questions remain about the safety of repeated treatment cycles, although encouraging data are emerging. Although the U.S. Food and Drug Administration (FDA) has received reports of patients who developed fatal PML following rituximab treatment for SLE and RA, this appears to be rare. A key issue determining the future place of B cell depletion therapy will be defining the most effective strategy in early stages of RA to induce a remission and potentially even biologic-free remission, whether this can be achieved safely and effectively with rituximab, and whether any biomarkers can be developed that reliably inform biologic treatment strategy on an individual patient basis.

Clinical trials and safety data for other antibodies targeting CD20 such as ocrelizumab, a humanized version of rituximab, and HuMax-CD20/afatumumab, a fully human anti-CD20, have also been reported. In phase I/II trials in RA, ocrelizumab, in combination with methotrexate, was found to be safe and effective at doses comprising two infusions of 200 mg or higher given 2 weeks apart. Researchers...
Rituximab in Other Rheumatic Conditions

**KEY POINT**

Rituximab has demonstrated benefits in studies of RA and ANCA-associated vasculitis. Clinical trials in SLE have not shown clinical benefit.

Rituximab has also been used in a number of other rheumatic diseases. Theoretic considerations and preliminary data suggested that B cell depletion using rituximab might have efficacy for immune thrombocytopenia, antineutrophil cytoplasmic antibody–associated vasculitis, and SLE. The rationale for the use of rituximab in the treatment of patients with ANCA-associated vasculitis is that elimination of CD20 B-cell precursors could lead to transient removal of pathogenic antibodies and remission, assuming that ANCAs are produced by short-lived B lineage cells rather than long-living plasma cells. Furthermore, in ANCA-associated vasculitis, the number of activated, circulating B lymphocytes correlates with disease activity and tissue involvement. Therefore the hypothesis that rituximab might induce disease remission in patients with severe ANCA-associated vasculitis has been tested in a phase II/III multicenter, randomized, double-blind, placebo-controlled trial known by the acronym RAVE (rituximab for ANCA-associated vasculitis). Findings have recently been reported comparing rituximab (375 mg/m² administered intravenously once weekly for 4 weeks) to cyclophosphamide (2 mg/kg per day administered orally). Sixty-three patients in the rituximab group (64%) reached the primary end point, as compared with 52 patients in the control group (53%), a result that met the criterion for noninferiority (P < .001). The rituximab-based regimen was more efficacious than the cyclophosphamide-based regimen for inducing remission of relapsing disease; 34 of 51 patients in the rituximab group (67%) as compared with 21 of 50 patients in the control group (42%) reached the primary end point (P = .01). Rituximab was also as effective as cyclophosphamide in the treatment of patients with major renal disease or alveolar hemorrhage. There were no significant differences between the treatment groups with respect to rates of adverse events. Despite these encouraging data, the true positive effect of rituximab in ANCA-associated vasculitis is difficult to determine because of the simultaneous administration of high-dose glucocorticoids, which may contribute to a substantial decrease in ANCA titers and the observed remission rates.

Given the large body of evidence implicating abnormalities in the B cell compartment in SLE, a recent therapeutic focus has been to develop interventions that target the B cell compartment by multiple mechanisms, and rituximab has been studied most extensively. The best evidence in ANCA-associated vasculitis, the number of activated, circulating B lymphocytes correlates with disease activity and tissue involvement. Therefore the hypothesis that rituximab might induce disease remission in patients with severe ANCA-associated vasculitis has been tested in a phase II/III multicenter, randomized, double-blind, placebo-controlled trial known by the acronym RAVE (rituximab for ANCA-associated vasculitis). Findings have recently been reported comparing rituximab (375 mg/m² administered intravenously once weekly for 4 weeks) to cyclophosphamide (2 mg/kg per day administered orally). Sixty-three patients in the rituximab group (64%) reached the primary end point, as compared with 52 patients in the control group (53%), a result that met the criterion for noninferiority (P < .001). The rituximab-based regimen was more efficacious than the cyclophosphamide-based regimen for inducing remission of relapsing disease; 34 of 51 patients in the rituximab group (67%) as compared with 21 of 50 patients in the control group (42%) reached the primary end point (P = .01). Rituximab was also as effective as cyclophosphamide in the treatment of patients with major renal disease or alveolar hemorrhage. There were no significant differences between the treatment groups with respect to rates of adverse events. Despite these encouraging data, the true positive effect of rituximab in ANCA-associated vasculitis is difficult to determine because of the simultaneous administration of high-dose glucocorticoids, which may contribute to a substantial decrease in ANCA titers and the observed remission rates.

Given the large body of evidence implicating abnormalities in the B cell compartment in SLE, a recent therapeutic focus has been to develop interventions that target the B cell compartment by multiple mechanisms, and rituximab has been studied most extensively. The best evidence in support of using rituximab for the treatment of patients with ANCA-associated vasculitis or SLE comes from clinical experience, retrospective case series, and small prospective uncontrolled studies, mainly in patients with refractory or frequently relapsing disease. However, recently two moderately sized phase III randomized placebo-controlled trials of rituximab for the treatment of moderately active nonrenal SLE (EXPLORER) or class III/IV lupus nephritis (LUNAR) have failed to demonstrate superiority of this B
Targeting Co-Stimulatory Molecules

**KEY POINTS**

- Activated T cells are implicated in the pathogenesis of RA and co-stimulation is essential in induction of adaptive immune responses.

- Abatacept is a fully human fusion protein comprising the extracellular portion of CTLA-4 and IgG-1 Fc.

- Abatacept is a second-line biologic drug that can be used when conventional synthetic DMARDs and/or other biologic drugs have failed to control inflammatory arthritis.

- The reported risk of serious infections is no higher with abatacept compared to other biologic drugs.

- Abatacept is not associated with cancer risk in clinical reports to date. Larger study groups and long-term study will further clarify this.

- Abatacept has acceptable safety across the spectrum of RA-patient populations.

- It is administered as a 30-minute intravenous infusion that is usually achieved without complications. Subcutaneous administration is equally effective and is now approved for use in the United States.

- Abatacept is generally considered as a biologic option in RA patients with inadequate responses to TNF inhibitors, although recent evidence suggests that the benefit-to-risk profile of abatacept may be most optimal when introduced earlier in the treatment paradigm.

- Clinical efficacy of abatacept in RA validates the importance of co-stimulation in the pathogenesis.

- Co-stimulation inhibition modulates production of a number of proinflammatory cytokines.

- Abatacept has clinical use for RA and polyarticular juvenile idiopathic arthritis.

- Clinical trials on abatacept in psoriatic arthritis have shown benefit.

Co-stimulation is an essential step in the induction of adaptive immune responses. Although the role of T cells in the perpetuation of RA has been debated and remains poorly understood, it has long been believed that T cell activation is a key event in the pathogenesis. Successful T cell activation requires multiple signals. One signal is provided by presentation of an antigen bound to cell surface major histocompatibility complex (MHC) molecules on cell-depleting agent over placebo when added to the standard of care, conventional immunosuppressive therapy. Both trials had a relatively short follow-up period. Preliminary data from LUNAR (A Study to Evaluate the Efficacy and Safety of Rituximab in Subjects with ISN/RPS Class III or IV Lupus Nephritis) have been reported. This multicenter, randomized, double-blind, placebo-controlled trial comprised 144 patients with lupus nephritis (67% of patients had class IV disease) and compared the efficacy and safety of rituximab with placebo. Patients with class III and IV disease and urine protein-to-creatinine ratio greater than 1 were randomly assigned to receive either 1000 mg rituximab or placebo on days 1, 15, 168, and 182, in conjunction with mycophenolate mofetil and corticosteroids. No significant differences were observed in complete or partial renal response or clinical benefit to therapy at week 52, although rituximab administration was associated with significantly reduced titers of antibodies to double-stranded DNA and increased levels of C3 complement component. Serious adverse events such as infection were similar between the two patient groups.

The EXPLORER (A Study to Evaluate the Efficacy and Safety of Rituximab in Patients with Severe Systemic Lupus Erythematosus) trial randomly assigned patients with SLE with moderate to severe disease activity despite treatment with immunosuppressives and corticosteroids to receive either placebo (n = 88) or rituximab infusions (n = 127). Patients with active glomerulonephritis were excluded. The British Isles Lupus Assessment Group Index was used to score treatment response four times per week for 52 weeks after the first infusion. Responses were recorded in 66% of patients in the placebo group, and 75.1% of patients were treated with rituximab. The time to a first moderate or severe flare did not differ between groups, but a trend for a prolonged time to first “a” score flare in the rituximab group was observed. Annual rates of severe and moderate disease activity flares were similar, but the mean annual rate of “a” score flares was significantly lower in the rituximab group than in the placebo group (0.86 vs. 1.41). The numbers of adverse events and overall infections were comparable between groups at 78 weeks, although serious infections were more numerous in the placebo group.

Why the LUNAR and EXPLORER studies failed to prove superiority of rituximab to placebo in patients with SLE is unclear, although design flaws seem likely. Overuse of concomitant steroids and continued immunosuppressive treatment could help mask the possible benefits of rituximab. Moreover, consensus on the optimal dose and administration regimen of rituximab in patients with SLE and adjustments for the organ or system involved is lacking. EXPLORER included patients with very active disease treated aggressively with moderate-to-high-dose glucocorticoids, which made the short-term detection of treatment benefits difficult. Another potential shortcoming for lupus clinical studies in general, including EXPLORER and LUNAR, is that the length of follow-up may have been too short to demonstrate separation between the different treatments. The unexpected failure to show overall clinical benefit of rituximab in SLE trials may also reflect the inadequacy of the clinical outcome instruments employed. This is less likely to be relevant in the LUNAR nephritis study, for which the outcome measurements were more unequivocal. It is of note that in both the EXPLORER and LUNAR studies, outcomes appeared to be more favorable in African American and Hispanic patients.

Rituximab has also been used in primary Sjögren’s syndrome; granulomatosis with polyangiitis (GPA) (formerly Wegener’s granulomatosis); hepatitis C–associated cryoglobulinemia; antineutrophilic cytoplasmic antibody–associated vasculitides other than GPA such as polyarteritis nodosa, dermatomyositis, and polymyositis; antiphospholipid syndrome; and scleroderma.
Abatacept binds to CD80 and CD86 on antigen-presenting cells and helps return the cells to the quiescent state. In the absence of further signals, T cells become unresponsive and may ultimately be eliminated through apoptosis. An important co-stimulatory signal is provided by an interaction between members of the B7 family (either CD80 or CD86) on antigen-presenting cells and CD28 on T cells. After activation, T cells express CTLA-4, which interferes with the B7-CD28 interaction and helps return the cells to the quiescent state.

**Abatacept in Rheumatoid Arthritis**

Abatacept is a novel, fully human fusion protein comprising the extracellular portion of CTLA-4 and the Fc fragment of a human IgG-1 (CTLA4Ig). In December 2005, abatacept (Orencia) became the first co-stimulatory blocker to be approved by the FDA for the treatment of patients with RA who have had an inadequate response to other drugs. Abatacept binds to CD80 and CD86 on antigen-presenting cells, thus preventing these molecules from binding their ligand, CD28, on T cells, with the consequent inhibition of optimal T cell activation. In vitro, abatacept decreases T cell proliferation and inhibits the production of TNF, IFN-γ, and interleukin (IL)-2. CTLA4Ig showed promising activity in rodent collagen-induced arthritis models, prompting its evaluation in several clinical trials in patients with RA.

**Clinical Studies**

Abatacept has been evaluated in several double-blind, placebo-controlled trials in a number of clinical scenarios in adults with active RA. These include an inadequate response to conventional DMARDs such as methotrexate or to TNF inhibitors and, more recently, methotrexate-naïve patients in the early phase of disease. In addition, data have been reported in the context of an exploratory phase II study designed to assess the effect of co-stimulation blockade on progression of undifferentiated, early inflammatory arthritis to fulfillment of classification criteria for RA.

An initial 3-month, phase Ila, double-blind, randomized, placebo-controlled pilot study demonstrated the efficacy of B7 blockade in treating the signs and symptoms in patients with active RA despite treatment with at least one conventional DMARD. In this pilot study, the effect of one of two different biologic co-stimulatory modulators was compared with that of placebo infusions. The two biologic agents used were CTLA4Ig, which binds approximately fourfold less avidly to CD86 than to CD80, and LEA29Y (belatacept), a second-generation CTLA4Ig with two mutated amino acid residues conferring an increased avidity for CD86 over that of the parent molecule. The proportion of patients achieving ACR20 responses on day 85 was dose dependent, suggesting clinical efficacy for both co-stimulatory blocking molecules.

The findings were confirmed in a multicenter phase IIb study of abatacept plus methotrexate in 339 patients with active RA despite methotrexate treatment. In this study, patients were randomized to receive infusions of placebo, abatacept 2 mg/kg, or abatacept 10 mg/kg at baseline, 2 weeks, 4 weeks, and then monthly through 6 months. ACR20 responses were achieved in 60%, 41.9%, and 35.3% of patients receiving abatacept 10 mg/kg, abatacept 2 mg/kg, and placebo, respectively. At the more stringent ACR50 response level, the figures were 36.5%, 22.9%, and 11.8% (Table 64-2). Improvements in the individual components of the ACR response criteria were generally greater in the 10 mg/kg group than the 2 mg/kg group. No deaths, malignancies, or opportunistic infections were reported for

![Figure 64-5](https://example.com/figure64-5.png)  
**Figure 64-5** Interactions between antigen-presenting cells and T cells. Successful T cell activation requires multiple signals. One signal is provided by the presentation of an antigen bound to cell surface major histocompatibility complex (MHC) molecules on antigen-presenting cells to a specific T cell receptor (TCR). In the absence of further signals, T cells become unresponsive and may ultimately be eliminated through apoptosis. An important co-stimulatory signal is provided by an interaction between members of the B7 family (either CD80 or CD86) on antigen-presenting cells and CD28 on T cells. After activation, T cells express CTLA-4, which interferes with the B7-CD28 interaction and helps return the cells to the quiescent state.

---

**Table 64-2  Percentage of Patients Achieving Responses at 24 Weeks in the AIM and ATTAin Studies**

<table>
<thead>
<tr>
<th>Study</th>
<th>Drug Regimen</th>
<th>ACR20</th>
<th>ACR50</th>
<th>ACR70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase Iib AIM</td>
<td>Abatacept 10 mg/kg plus methotrexate</td>
<td>60</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Methotrexate</td>
<td>35</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Phase III AIM</td>
<td>Abatacept 10 mg/kg plus methotrexate</td>
<td>68</td>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Methotrexate</td>
<td>40</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>Phase III ATTAin</td>
<td>Abatacept 10 mg/kg plus methotrexate</td>
<td>50</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Methotrexate</td>
<td>20</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

ACR, American College of Rheumatology; AIM, Abatacept in Inadequate Responders to Methotrexate; ATTAin, Abatacept Trial in Treatment of Anti-TNF Inadequate Responders.
any patient receiving abatacept during the 6 months of therapy. Patients in the phase IIb study continued on blinded therapy for an additional 6 months, during which time response to therapy was maintained. For patients receiving abatacept 10 mg/kg, the ACR70, ACR50, and ACR20 response rates were 21%, 42%, and 63%, respectively, compared with 8%, 20%, and 36% for patients receiving placebo infusions. Further, at the higher dose, statistically significant improvements in physical function and health-related quality of life were maintained over the 1-year period.\(^7\) In the phase IIb study, from day 90 onward, there were statistically significant and progressively rising differences in remission rates between the group receiving methotrexate plus abatacept 10 mg/kg and the group assigned to methotrexate and placebo infusions. By 1 year of treatment, 34.8% of abatacept plus methotrexate patients achieved a DAS28 remission (<2.6), in contrast to 10.1% of the methotrexate plus placebo patients (\(P < .001\)).\(^7\) Patients completing the double-blind phase over 12 months became eligible to enter a long-term extension phase in which all participants received methotrexate plus abatacept 10 mg/kg. At year 3, abatacept-treated patients experienced greater than 70% improvement in swollen and tender joint counts and approximately 50% improvement in pain and physical function.\(^8\) Patients who received placebo infusions during the double-blinded phase and then switched to abatacept during the long-term extension rapidly achieved equivalent efficacy to those treated with abatacept throughout.

The findings of two large phase III studies of abatacept in different RA populations were recently reported. A population of methotrexate-refractory patients was studied in the Abatacept in Inadequate Responders to Methotrexate (AIM) trial. This study was designed to further evaluate the safety and clinical efficacy of abatacept plus methotrexate, as well as the effect of radiographic progression.\(^7\) In the other phase III study, the Abatacept Trial in Treatment of Anti-TNF Inadequate Responders (ATTAIN),\(^8\) the objective was to determine whether abatacept is a safe and effective treatment for RA patients who were unresponsive to previous anti-TNF treatment.

In the AIM study, 652 RA patients with an inadequate response to methotrexate were randomly assigned to receive placebo (219 patients) or a fixed dose of abatacept approximating 10 mg/kg (433 patients) on days 1, 15, and 29 and every 4 weeks thereafter for a year. All patients remained on background methotrexate therapy.\(^7\) Both patient groups exhibited high disease activity at baseline, with a DAS28 of 6.4. Patients receiving abatacept showed greater improvement in all ACR response criteria at 6 and 12 months than did placebo-treated patients (see Table 64-2). In findings similar to the phase IIb study, abatacept plus methotrexate induced DAS28 remission (<2.6) in 14.8% of patients at 6 months and in 23.8% at 12 months, compared with 2.8% and 1.9% of patients receiving methotrexate plus placebo at the corresponding time points (\(P < .001\)). Physical function significantly improved in 63.7% of the abatacept plus methotrexate group, versus 39.3% of the placebo plus methotrexate group (\(P < .001\)). Further, patients on the abatacept and methotrexate combination had a slower progression of mean structural damage (1.2 total Sharp score points over 1 year) compared with methotrexate alone (2.3 Sharp score points).\(^9\) Of interest, when assessed by conventional clinical outcome measures such as EULAR response, the data suggest that a plateau of clinical efficacy is achieved with abatacept between 4 and 6 months of treatment. However, using more stringent measures for analyses, such as time to a low DAS (DAS28 < 3.2) or time to a sustained low DAS, no plateau of efficacy was observed over the first 12 months, suggesting an ongoing recruitment of clinical benefit with abatacept plus methotrexate.\(^9\)

Patients completing the double-blind phase of the AIM study over 12 months became eligible to enter a long-term extension phase in which all participants received methotrexate plus abatacept at a fixed dose approximating 10 mg/kg every 4 weeks (Figure 64-6). Clinically meaningful reductions in disease activity were maintained through 2 years, accompanied by an improved sense of subjective

---

**Figure 64-6** Phase III AIM (Abatacept in Inadequate Responders to Methotrexate) study design. ACR, American College of Rheumatology; DMARD, disease-modifying antirheumatic drug; MTX, methotrexate.

- MTX inadequate responders with active disease* receiving MTX (\(n = 652\))
- MTX stabilization†
- Randomization‡
- Randomization§
- Placebo + MTX (\(n = 219\))
- Abatacept 10 mg/kg + MTX (\(n = 433\))
- Patients completing 12 months were eligible for LTE phase
- Long-term extension phase

*≥swollen joints (66 joint count), ≥12 tender joints (68 count) and CRP ≥1.0 mg/dL and ≥3 months on MTX ≥15 mg/week.
†Stable MTX dose for >1 month before enrollment.
‡2:1 randomization to study arm.
§Sulfasalazine, hydroxychloroquine, gold.
well-being assessed by patient-reported outcomes. Further, inhibition of structural damage to joints was sustained, as evaluated by plain radiography; the effect after 2 years of abatacept was significantly greater than that at 1 year, with minimal radiographic progression observed over the second year of treatment.

A third trial in methotrexate-inadequate responders provided the opportunity to evaluate two biologics in a single study. The placebo- and active comparator-controlled ATTEST (abatacept or infliximab vs. placebo, a trial for tolerability, efficacy and safety in treating RA) study, although not powered to detect superiority, provided information on the relative efficacy and safety profiles of abatacept and infliximab versus placebo in the same population. Patients with an inadequate response to methotrexate were randomized (3:3:2) to abatacept (approved dose, n = 156), infliximab (3 mg/week, n = 165), or placebo (n = 110), with background methotrexate. At month 6, patients in the placebo group were switched to abatacept, and infliximab and abatacept groups continued to year 1, with blinding maintained. The primary end point of this trial, reduction in DAS28 (erythrocyte sedimentation rate [ESR]) at month 6 for abatacept versus placebo, was met, with mean reductions of −2.53 versus −1.48 (P < .001), respectively. The proportion of patients achieving low disease activity and DAS28 remission was also greater with abatacept. Improvements in ACR20, ACR50, and ACR70 responses at month 6 were significantly greater versus placebo for both abatacept and infliximab. The onset of ACR20 responses was generally more rapid for infliximab than abatacept, but responses were similar by month 3. By year 1, DAS28 (ESR) reductions of −2.88 and −2.25 were seen for abatacept- and infliximab-treated patients, respectively, and ACR responses were maintained from month 6 with abatacept but not with infliximab treatment.

In the ATTAIN phase III study, abatacept therapy was evaluated in 391 patients with active disease receiving conventional DMARDs or anakinra who failed to respond adequately to at least 3 months of therapy with etanercept, infliximab, or both agents at the approved doses. Anti-TNF therapy was discontinued at the time of enrollment if it had not been stopped previously. Following a washout period, patients were randomly assigned in a 2:1 ratio to receive either the same fixed dose of abatacept (approximately 10 mg/kg) or placebo (Figure 64-7). Patients receiving abatacept showed significantly greater improvement in all ACR response criteria through 6 months (see Table 64-2) than did placebo-treated patients (ACR20 response, 50.4% vs. 19.5%, P < .001; ACR50 response, 20.3% vs. 3.8%, P < .001; and ACR70 response, 10.2% vs. 1.5%, P = .003) (Figure 64-8). Further, a DAS28 remission was achieved in 10% of patients receiving abatacept, versus only 1% of patients receiving placebo infusions plus DMARDs. ACR20 responses were seen irrespective of whether patients had previously been exposed to etanercept, infliximab, or both anti-TNF therapies without an adequate response. Improvement in physical function was also significantly increased in the abatacept group (47% vs. 23%). The incidence of infection was slightly higher in the abatacept group than in the placebo group, although no specific infection was clearly more frequent, and the intensity of infections was similar in the two groups. There were no significant differences in the number of patients discontinuing treatment as a result of infection or in the incidence of serious infection.

All patients who completed the 6-month double-blind phase of the ATTAIN study were eligible to enter a 1-year, long-term extension phase during which all patients...
received a once-monthly fixed dose of abatacept in addition to at least one conventional DMARD.83 Of 258 patients randomized to abatacept during the double-blind phase, 223 completed 6 months of treatment and 218 entered the long-term extension. Of these, 168 completed 18 months’ treatment. The ACR20 responses observed at the end of the double-blind phase were sustained throughout the 1-year extension phase, with the proportion of patients achieving the more stringent ACR50 and ACR70 responses rising to 35% and 18%, respectively, at 18 months. Further, the proportion of patients meeting DAS28 remission criteria doubled to 22.5% by the end of the extension period. Similarly, among all patients initially treated with abatacept and DMARDs who entered the long-term extension phase, the mean reduction in DAS28 from baseline to the end of the double-blind phase was −1.99; by the end of 18 months, the mean reduction from baseline was −2.81. In the double-blind phase, patients assigned to placebo infusions together with DMARDs had a mean reduction in DAS28 of −0.93; at the end of the long-term extension, after crossing over to abatacept infusions, the reduction from baseline was −2.72.84 These data again emphasize the sustained but relatively slow and incremental clinical responses observed following abatacept therapy.

The second trial conducted in TNF-inadequate responders was a Phase IIIb/IV, 6-month, open-label study. The ARRIVE (Abatacept Researched in RA patients with an inadequate anti-TNF response to Validate Effectiveness) trial was the first to assess the safety of abatacept in patients who switched directly from TNF antagonist therapy without undergoing washout, an approach that may be more clinically relevant for everyday practice.85 Enrolled patients in this study had high disease activity levels at entry and an inadequate response to up to three TNF antagonists that they could have failed for efficacy, safety, or tolerability reasons. Patients were eligible even if they had a positive purified protein-derivative test result. Abatacept could be administered as monotherapy (for those recruited in the United States only), and patients were not limited to a particular background DMARD. Similar, clinically meaningful improvements were seen in disease activity, physical function, and health-related quality of life, regardless of whether there was a washout period or not. Post hoc analyses revealed that numerically more patients who had previously failed one TNF antagonist achieved DAS28-defined remission and low disease activity than those who had failed two or more TNF inhibitors.

Because T cell activation is believed to be an initiating event in an immunologic cascade observed in RA, co-stimulation blockade might be predicted to have benefits from an early stage of evolution of the syndrome independently of any driving antigen or antigens. The 2-year abatacept study to gauge remission and joint damage progression in methotrexate-naïve patients with early erosive RA (AGREE) study consisted of a 12-month double-blind phase followed by a 12-month open-label period in methotrexate-naïve patients with early RA.86 Eligibility included short disease duration, poor prognostic factors including high C-reactive protein (CRP) levels, radiographic evidence of erosions, and seropositivity for RF or anti-CCP2. Patients were randomized 1:1 to receive abatacept plus methotrexate (n = 256) or methotrexate alone (n = 253) over the first 12-month period.85 From year 1 onwards, all patients received open-label abatacept plus methotrexate. The co-primary end points were 28-joint DAS (DAS28)-defined remission and radiographic joint damage progression at year 1. Significantly more patients treated with abatacept plus MTX achieved DAS28 (CRP)-defined remission and ACR50 and ACR70 responses at year 1, and the difference between treatment arms was significant by month 2. Over 1 year, 27.3% versus 11.9% of abatacept plus methotrexate versus methotrexate alone–treated patients (P < .001) achieved a major clinical response (ACR70 maintained for 56 consecutive months).86 Significant improvements were also seen in physical function at year 1, for abatacept plus methotrexate over methotrexate alone–treated patients.86 Improvements in disease activity and ACR responses were sustained or improved over the second year for patients remaining on abatacept plus methotrexate therapy, with 55.2% achieving remission at year 2.87 Changes from baseline to year 1 in the Genant-modified total Sharp score and erosion score were significantly lower for methotrexate-naïve patients randomized to the abatacept plus methotrexate arm.88 Furthermore, there was an increasing degree of inhibition of progression in year 2 relative to year 1 for patients originally randomized to abatacept.88 For patients originally receiving methotrexate alone, structural damage progression was reduced over year 2 relative to year 1, following the addition of abatacept.87 However, overall structural damage progression at year 2 remained greater for these patients compared with patients who received abatacept from baseline.87

The potential for early treatment with abatacept to delay the time for progression to fulfillment of classification criteria for RA in patients with anti-CCP2-positive, undifferentiated, early inflammatory arthritis with clinical synovitis of two or more joints was investigated in an exploratory, phase II, 2-year study known by the acronym ADJUST (Abatacept Study to Determine the Effectiveness in Preventing the Development of RA in Patients with Undifferentiated Inflammatory Arthritis).89 Following 6 months of double-blind, randomized (1:1) treatment with either abatacept at the approved dose of approximately 10 mg/kg (n = 28) or placebo (n = 28), abatacept treatment was terminated. The proportion of patients who developed RA according to ACR 1987 criteria or discontinued due to lack of efficacy at year 1 was assessed. When abatacept was stopped after 6 months, 22 and 17 patients treated with abatacept and placebo, respectively, remained in the study inasmuch as they had not fulfilled classification criteria for RA; by year 2, 7 and 4 patients remained in the study. Numerically more placebo than abatacept patients developed RA over 1 year (66.7% vs. 46.2%), although confidence intervals overlapped. Radiographic assessments demonstrated an inhibitory effect on structural damage progression at month 6, which was maintained for 6 months following therapy cessation, with similar trends observed for magnetic resonance imaging (MRI)-assessed osteitis, erosion, and synovitis.89 Abatacept treatment was also associated with reductions in anti-CCP2 levels that persisted beyond cessation of active drug. At enrollment, patients had short symptom duration, and although they did not meet ACR 1987 criteria for RA at the point of
recruitment, more than half already had evidence of one or more erosions and were thus likely to have had early RA. The findings must be interpreted in light of this.

Safety Issues

Safety assessments from abatacept clinical trials have in general demonstrated a comparable overall incidence of adverse events and serious adverse events for abatacept and placebo-treated patients. The safety of long-term abatacept treatment is reported to be consistent, with the incidence of overall adverse events and serious adverse events remaining stable up to 7 years.90

A safety analysis, which included 4150 patients who were exposed to abatacept, has been undertaken on pooled data from abatacept clinical trials through December 2007. This represents 10,365 patient-years of exposure, with an average exposure period of 2.5 years.90 The incidence of serious infections was generally low, although it was higher for abatacept-treated compared with placebo-treated patients over 1 year (serious infections: 3.47 vs. 2.41 events/100 patient-years, respectively). Records of annual incidence rates for serious infections did not appear to show an increase in risk over time. Pneumonia, bronchitis, cellulitis, and urinary tract infections were the most common causes of hospital admission for infections. Opportunistic infections were rarely observed in this pooled cohort including the following events per 100 patients-years of treatment: Mycobacterium tuberculosis, 0.06 events; aspergillosis, 0.02; blastomycosis, 0.01; and systemic candida, 0.01.90 The incidence of malignancies (excluding nonmelanoma skin cancer) during the double-blind treatment periods was reported to be 0.59 events per 100 patient-years for abatacept-treated versus 0.63 events per 100 patient-years for placebo-treated patients. This low incidence rate did not rise with increasing exposure. In particular, for lung cancer and lymphoma, the incidence rates were 0.24 lung cancers and 0.06 lymphomas per 100 patient-years over the double-blind period and 0.16 and 0.07 lymphomas per 100 patient-years, respectively, over the cumulative period. In order to better interpret malignancy risk data, the incidence rates observed in abatacept studies were compared with those documented in five observational cohorts of biologic-naïve RA patients on nonbiologic DMARDs. Standardized incidence rates for the observational cohorts ranged from 0.4 to 1.06, and for abatacept-treated patients the risk of lung cancer did not appear to be increased with standardized incidence rates ranging from 0.65 to 1.84. The lymphoma risk in abatacept-treated patients appeared to be comparable with that in biologic-naïve RA with standardized incidence rates ranging from 0.60 to 1.23.90 In the pooled trial data from double treatment periods, autoimmune events were reported in 1.4% of abatacept-treated and 0.8% of placebo-treated patients. Most events were mild or moderate in intensity, and the most frequently reported was psoriasis with rates of 0.53 and 0.56 events per 100 patient-years, over the double-blind and cumulative periods, respectively.

In clinical practice, it is common to use conventional DMARDs in combination regimens on the basis of the belief that there are additive benefits in terms of efficacy without the downside of unacceptable toxicity. Whether these same principles apply to the use of abatacept was addressed in the ASSURE trial (Abatacept Study of Safety in Use with Other RA Therapies).91 This multicenter, randomized, double-blind study investigated the safety of adding abatacept or placebo infusions to a background treatment regimen of at least one of the traditional nonbiologic or biologic DMARDs currently approved for RA treatment for at least 3 months. A total of 1436 patients were randomized 2:1 to receive abatacept at a fixed dose approximating 10 mg/kg by weight range or placebo. A number of interesting observations arose from this study. In the group as a whole, the proportion of serious adverse events occurring in each treatment arm was similar: 13% for abatacept and 12% for placebo. The discontinuations rate due to adverse events was 5% in the abatacept group and 4% in the placebo group. As expected on the basis of prior studies, serious infections occurred more frequently in the abatacept group (2.9%) than in the placebo group (1.9%). There were five deaths in the abatacept group and four in the placebo group; all but one death in each group was thought unlikely to be related to the study drug. All the deaths occurred in patients without concomitant biologic background therapy. However, a subanalysis of the data, based on whether patients were receiving biologic or nonbiologic background therapy, revealed that serious adverse events occurred almost twice as frequently in the subgroup receiving abatacept plus another biologic agent (22.3%) as in the other subgroups (12.5%). A particularly important observation in this study was the increased number of serious infections observed when abatacept was combined with other biologic therapies (5.8% vs. 1.6% for the subgroup on background biologic therapy plus placebo infusions). Further, the clinical benefits of abatacept tended to be less in the patients receiving background biologic therapy than in those with a background of nonbiologic DMARDs. There were no reported cases of lymphoma, demyelinating disorders, or tuberculosis.

The ASSURE trial findings mirrored those of a smaller randomized, placebo-controlled, double-blind pilot study. This phase IIb trial investigated the efficacy and safety of the addition of abatacept infusions at 2 mg/kg over 1 year in patients with at least 8 of 66 swollen joints and 10 of 68 tender joints despite at least 3 months of treatment with twice-weekly 25 mg subcutaneous etanercept.92 The biologic combination had limited clinical benefit over etanercept and placebo infusions but was associated with an increase in the proportion of patients experiencing serious adverse events (16.5% vs. 2.8%) and serious infections (3.5% vs. 0%). On the basis of these observations, the use of abatacept is not advised in combination with other biologic therapies.

Emerging data concern the comparative efficacy, safety, and kinetics of response for the anti–TNF antibody infliximab and abatacept.93 In a 1-year double-blind study, RA patients with an inadequate response to methotrexate (mean baseline DAS28 of 6.8) and no prior anti-TNF therapy were randomized to receive abatacept at a dose approximating 10 mg/kg every 4 weeks (156 patients), infliximab 3 mg/kg every 8 weeks (165 patients), or placebo every 4 weeks (110 patients). Patients randomized to placebo were switched to abatacept after 6 months but were not included in the 1-year analyses. At the end of the first...
6 months, the frequency of serious adverse events was 5.1%, 11.5%, and 11.8% for abatacept, infliximab, and placebo, respectively. In the same order, the frequency of acute infusion-related adverse events was 5.1%, 18.2%, and 10%. Over the 1-year period, infections reported as serious adverse events were more frequent with infliximab (8.5%) than with abatacept (1.9%). These included two cases of tuberculosis, both in infliximab-treated patients. When considered in the light of clinical response data discussed earlier, this and other studies emphasize the relatively slow time to peak clinical response with abatacept in comparison with TNF blockade, with increasing efficacy beyond 6 months. They also point to the possibility of a favorable benefit-risk profile over 1 year. It will be critical to see whether these encouraging early safety data are maintained over the longer term.

Of note, the incidence rate of serious infections observed with abatacept is at the lower end of the range reported in RA patients treated with other biologics. In summary, the long-term integrated safety data from up to eight abatacept trials, representing more than 10,000 patient-years of exposure, confirm that, overall, abatacept has a favorable safety profile that is consistent with observations from the short-term experience in all RA populations studied, with no new clinically important safety issues identified with long-term exposure, a conclusion supported by a recent Cochrane Review.93

**Current Role**

Abatacept may be used as monotherapy or concomitantly with DMARDs other than TNF antagonists. It is not recommended for use concomitantly with IL-1 or TNF antagonists. The encouraging clinical trial data confirm that abatacept, like rituximab, represents a valuable agent in the therapeutic armamentarium for patients with RA who have not responded adequately to TNF blockade. Abatacept has particular advantages among intravenously delivered biologics for RA, which is the formulation in which it is currently available, in that it is well tolerated and quick to administer and infusion-related problems are rare. However, the clinical responses and radiographic benefits observed with abatacept appear to be greater in methotrexate-naïve patients compared with patients who have failed methotrexate or other DMARDs. Furthermore, patients who previously failed methotrexate treatment appear to demonstrate higher clinical responses than patients who have failed TNF antagonists. These considerations based on recent clinical trial data suggest that the most favorable clinical outcomes with abatacept may be achieved if it is used earlier in the treatment paradigm than has generally been the case since it became available for use in the clinic. Other factors that influence thinking in positioning a biologic therapy are that the comparative effects of abatacept, rituximab, and TNF blockade on structural damage are still unknown. It is clear that when clinical responses are unsatisfactory, combination therapy with methotrexate and an anti-TNF agent may confer significant joint protection compared with methotrexate alone. Thus the merits of switching a patient from a TNF inhibitor to abatacept or rituximab on the basis of a clinical response that is incomplete are not yet clear-cut with respect to disease modification. Other factors likely to inform the future use and relative positioning of biologics in the clinic include additional long-term safety data, comparative cost-effectiveness analyses, and the perceived convenience of intravenous administration. The popularity among patients of the subcutaneous delivery route out of available choices of parenteral delivery has prompted studies to look at effectiveness and safety of subcutaneously delivered abatacept. Preliminary data look encouraging. In a noninferiority trial in 1457 moderate to severely active RA patients on background methotrexate, weekly subcutaneous injection of an investigational formulation of abatacept, following a single intravenous loading dose provided an improvement in disease activity similar to that observed in patients assigned to monthly intravenous administration.94 Supporting studies evaluated safety and efficacy when switching patients from intravenous to subcutaneous abatacept therapy. They investigated immunogenicity, safety, and efficacy when withdrawing patients from and reintroducing them to the subcutaneous formulation. Preliminary data report that subcutaneous abatacept appears to be well tolerated after a 3-month interruption in patients with RA. Subsequent reintroduction is not associated with any impairment of efficacy or safety, compared with continuous dosing.95

**Implications for Understanding the Pathogenesis of Rheumatoid Arthritis**

The clinical efficacy of abatacept in a proportion of RA patients implicates co-stimulatory events in disease pathogenesis. However, there are a number of different mechanisms by which abatacept might mediate immunosuppressive effects in RA. In the RA joint, blocking access of CD28 to CD80 or CD86 might be of little importance because memory T cells, which predominate in inflamed synovium, are much less dependent on this pathway.96 A more important mechanism of action in the synovium might be the induction of tolerogenic antigen-presenting cells. Binding of CD80 or CD86 on antigen-presenting cells by CTLA4Ig initiates a “reverse signal,” with induction of tryptophan catabolism and inhibition of antigen presentation. Naïve T cells are located predominantly in lymphoid tissue, and blockade of the CD28-CD80/CD86 interaction in lymph nodes may reduce T cell priming and the production of autoreactive T cells. Interestingly, patients with RA are reported to have higher frequencies of CD28-null T cells.97 Further, the level of expression of CD28 is significantly reduced on all naïve T cells and memory CD4+ CD28+ T cells, a phenomenon related to overproduction of TNF.98 A likely consequence of low-density cell surface expression of CD28 is that, for these T cells, abatacept can more readily block the interaction with CD80 and CD86.

The clinical and radiographic benefits of abatacept administration in RA illustrate the importance of the co-stimulation pathway in T cell activation and subsequent amplification of the inflammatory cascade including pathways that promote tissue destruction.99 The modulatory effect of co-stimulation blockade on the expression of a range of inflammatory genes in synovial tissue has been demonstrated by quantitative polymerase chain reaction studies and evaluation of synovial biopsies in patients with active RA who received abatacept treatment having
previously failed to respond to TNF inhibitors. Furthermore, a small, largely nonsignificant reduction in cellular content was observed in biopsy samples following abatacept treatment, suggesting that co-stimulation inhibition reduces the inflammatory status of the synovium without disrupting cellular homeostasis.

**Abatacept in Other Rheumatic Conditions**

In addition to the use of abatacept in the established phase of RA, its use in combination with methotrexate is indicated for the treatment of moderate to severe active polyarticular juvenile idiopathic arthritis in pediatric patients 6 years of age and older who have had an insufficient response to other DMARDs including at least one TNF inhibitor. Results from a 12-month multicenter clinical trial did not show therapeutic benefit of abatacept over placebo in patients with non–life-threatening SLE. In particular, abatacept failed to prevent new disease flares in SLE patients tapered from corticosteroids in an analysis where mild, moderate, and severe disease flares were evaluated together. Serious adverse events were higher in the abatacept group (19.8% vs. 6.8%). Although the primary and secondary end points were not met, there were improvements in certain exploratory measures. There are ongoing studies in type 1 diabetes and inflammatory bowel disease. Abatacept was also used in a 26-week, phase I, open-label dose-escalation study of psoriasis vulgaris. Sustained improvements of at least 50% in clinical disease activity were reported following four infusions of abatacept in 20 of 43 patients with stable psoriasis vulgaris. Clinical improvement was associated with quantitative reduction in epidermal hyperplasia, which correlated with a quantitative reduction in skin-infiltrating T cells. However, no clear-cut increase in the rate of intraleisonal T cell apoptosis was identified. It may be that the observed reduction in lesional T cell numbers was due to inhibition of T cell proliferation, T cell recruitment, or apoptosis of antigen-specific T cells at extraleisonal sites. Altered antibody responses to T cell–dependent neocarriers were observed, but immunologic tolerance to these antigens was not demonstrated. This study illustrates the importance of the CD28-CD152 pathway in the pathogenesis of psoriatic skin disease. The findings of a 6-month, double-blind, placebo-controlled phase II trial in psoriatic arthritis have recently been reported. The study included 170 adults meeting the Classification Criteria for Psoriatic Arthritis (CASPAR) with active arthritis (≥3 swollen joints and ≥3 tender joints), active plaque psoriasis (at least one qualifying target lesion ≥2 cm in diameter), and a disease duration of at least 3 months. All participants were required to have an inadequate response to DMARDs including, but not limited to, methotrexate or anti-TNF therapies. Those who were either intolerant of, or had inadequate response to, TNF inhibitors discontinued these therapies at screening and, following a washout period of 28 or more days, were assessed for arthritis and psoriasis before randomization. Methotrexate was the only DMARD permitted to be continued at a stable dose during the study (in = 60% of participants). The participants were randomized to receive placebo or abatacept at a dose of 3, 10, or 30/10 mg/kg (two initial doses of 30 mg/kg followed by 10 mg/kg) on days 1, 15, and 29 and then every 28 days. The primary end point was the ACR20 response at the end of 6 months. Key secondary end points included Investigator’s Global Assessment (IGA) of psoriasis (ratings ranging from 1 [clear] to 4 [severe]) and the scores for target lesion (TL; based on the rating of skin erythema, induration, and scaling on a scale of 0 to 4 each), the HAQ-DI, and SF-36 General Health Survey. A total of 19%, 33%, 48%, and 42% of patients achieved an ACR20 with placebo, abatacept 3, 10, and 30/10 mg/kg, respectively. Compared with placebo, improvements were statistically significantly higher for the abatacept 10 mg/kg (P = .006) and 30/10 mg/kg (P = .0022) dose groups, but not for 3 mg/kg (P = .121). An IGA response (rating of psoriatic lesions as “clear or almost clear”) was seen in 21% of the 30/10 mg/kg dose group, 25% of the 10 mg/kg group, and 38% of the 3 mg/kg—the latter was the only one that was statistically significantly different to placebo (26%; no statistical data presented). All doses of abatacept showed improvements in scores on magnetic resonance imaging for joint erosion, osteitis, and synovitis; HAQ; and the SF-36 General Health Survey. These encouraging early findings suggest that abatacept 10 mg/kg may be a treatment option for patients with active psoriatic arthritis previously exposed to DMARDs including anti-TNF therapies. Longer-term follow-up of these patients is ongoing, and further studies are planned.

**TARGETING T CELLS**

**KEY POINTS**

A number of approaches to targeting T cells independently from co-stimulation pathways have not shown clear-cut benefit in clinical trials.

Novel approaches to targeting T cells are being tested including T cell vaccination.

**Clinical Studies**

In the early years of investigating the potential of biologic therapies in RA, T cells were among the first targets to be explored. Data from several different preclinical animal models of inflammatory arthritis suggested a pathogenic role for CD4+ T cells in response to various arthritogenic antigens presented in the context of MHC class II molecules. These observations led to a number of experimental protocols designed to investigate the effect of depleting and non-depleting antibodies directed at CD4, as well as other T cell–associated molecules. Early randomized, placebo-controlled clinical studies exploring the potential of biologic therapies targeting T cells in the treatment of RA have generally had disappointing results. Some anti–T cell agents were not efficacious; other preliminary trials demonstrating some clinical efficacy were terminated owing to adverse events, particularly prolonged and profound T cell depletion. However, the primatized monoclonal anti-CD4 antibody keliximab results in a dose-dependent clinical response when administered once weekly over 4 consecutive weeks, and the clinical response correlates with CD4+ T cell coating with keliximab rather than T cell depletion. In two consecutive randomized, double-blind trials with comparable populations, keliximab treatment was
associated with CD4+ T cell counts below 250 cells/mm³ in 12% in one study and 47% in the other.103

Examples of biologic therapies targeting other T cell–associated molecules include Campath-1H, a monoclonal antibody directed against CD52; a monoclonal anti-CD5 antibody linked to ricin toxin; and a fusion protein comprising an IL-2 receptor–binding domain coupled to diphtheria toxin (DABβ6IL-2 fusion toxin). CD52 is a polypeptide expressed on all lymphocytes. Campath-1H was tested as a treatment for refractory RA in two small trials, and although a single intravenous dose of between 1 and 100 mg resulted in significant CD4+ T cell depletion and clinical improvement in more than half of patients, there was poor correlation between biologic action and clinical response.106,107 Further, therapy was associated with significant acute toxicity, presumed to reflect a cytokine release syndrome including headache, nausea, and hypotension. Arthritis activity returned over time despite prolonged suppression of peripheral blood CD4+ T cell numbers.

CD5 is a transmembrane glycoprotein expressed on 70% of T cells. CD5-1C, a monoclonal antibody linked to ricin, a plant toxin that inhibits protein synthesis, was used to treat RA in a double-blind, placebo-controlled trial.108 At the doses tested, only modest and transient T cell depletion was observed and there was no clinical benefit.

In a strategy designed to selectively deplete activated T cells expressing IL-2 receptor, DABβ6IL-2 fusion toxin was given by intravenous infusion in open-label and placebo-controlled studies.109,110 Although a small percentage of patients (18%) exhibited clinical responses in the placebo-controlled study, there was a significant incidence of adverse events including nausea, fever, and raised plasma transaminases. Further, nearly all patients developed antibodies against diphtheria toxin.

Other approaches to targeting T cells that have been tested include efforts to directly interfere with the trimolecular complex comprising HLA class II, antigenic peptide, and T cell receptor (TCR) by means of a DR4-DR1 peptide vaccine, T cell receptor Vβ peptide vaccine, collagen, or cartilage glycoprotein 39.111 Despite a rationale for these approaches based on promising preclinical animal model data, all have been abandoned because of borderline or absent clinical benefits in human disease.

Future Directions

The importance of immune regulation in maintenance of the healthy state is perhaps best illustrated by the consequences of immune dysregulation, a phenomenon common to a wide range of chronic inflammatory disease phenotypes. Emerging evidence points to the importance of certain CD4+ T cell subsets in the negative regulation of the adaptive immune system. Best characterized of these subsets are the so-called naturally occurring CD4+CD25+ regulatory T cells and IL-10–producing Tr1 cells. Recent advances in the understanding of the molecular basis of CD4+CD25+ regulatory T cell generation include the observation that the X-linked forkhead–winged helix transcription factor Foxp3 is required for CD4+CD25+ T cell development and function. Although further progress in understanding the pathophysiologic role of regulatory T cells will require the identification of more specific markers for distinct regulatory T cell populations, there is already evidence of the feasibility of enhancing regulatory T cell function in vivo by either T cell receptor modulation, using antibodies to CD3, or co-stimulatory signals, using a CD28 superagonist.112

Where a chronic inflammatory disease reflects an antigen-driven process, an attractive goal is to modulate T cell function in such a way as to generate antigen-specific unresponsiveness in the absence of long-term generalized immunosuppression. Some, but not all, studies report a relative deficiency of regulatory T cells in RA.28 Of note, antibodies to CD3 and TNF appear to enhance regulatory T cell function or number in RA patients.113-116 This raises the possibility that combination treatment with anti-CD3 and anti-TNF might be beneficial in more completely restoring immune regulation in RA. In fact, chronic inflammation and overproduction of TNF perturb T cell antigen receptor–dependent signaling,117 suggesting that active inflammation may attenuate tolerogenic signals expected to be induced by a nondepleting anti-CD3 antibody. Thus pretreatment with TNF blockade might restore tolerogenic signals transduced by the T cell receptor in response to drugs such as anti-CD3. Although the use of anti-CD3 in the clinic has been limited by the occurrence of drug-induced cytokine release syndrome,118 the effects can be modulated by anti-TNF agents, as demonstrated in patients treated for acute allograft rejection.119 Such a combination therapy approach has yet to be tested in RA.

A potentially beneficial immunomodulatory response has been demonstrated in a small open pilot study by vaccinating 16 RA patients with expanded, activated, and irradiated autologous synovial fluid T cells.120 Vaccination was associated with expansion of CD4+ and CD8+ T cells, many of which expressed the Vβ2 T cell receptor chain. Some were anti-idiotypic, responding specifically to vaccine T cells with the production of IL-10 (CD4+ cells) or granzyme B (CD8+ cells). A broader regulatory response, however, was directed toward activated T cells in general, specifically against peptides derived from the IL-2 receptor α-chain. This broader response may be important to generate in a syndrome such as RA where the precise autoantigen and pathogenic T cell clones are not readily identifiable. This might also explain why earlier attempts at T cell vaccination using TCR-derived peptides have not been pursued.111 However, the efficacy of such a T cell vaccination approach needs to be validated in further trials and the safety and durability of response further investigated before the potential feasibility of such an approach to therapy in routine clinical use can be properly assessed.

SUMMARY

**KEY POINT**

Rituximab and abatacept have both established a role in the pharmacologic management of RA.

Despite recent advances in understanding the optimal use of nonbiologic DMARD therapies and the considerable success of biologic therapies targeting TNF, a substantial proportion of RA patients remains refractory to or intolerant of these therapeutic modalities. The data discussed in
this chapter illustrate that there is a role in clinical practice for newer biologics with specificity for cellular targets. In particular, depletion of B cells with the monoclonal antibody rituximab results in sustained improvement in the signs and symptoms of RA after just two doses (one treatment cycle), with little evidence of drug-related toxicity despite the profound and lasting depletion of B cells. Similarly, inhibition of T cell co-stimulation by abatacept demonstrates clear clinical efficacy within 16 weeks and, in some cases, additional improvement for 1 year or beyond, with acceptable safety. This observation is in marked contrast to the previously observed unfavorable risk-to-benefit ratio associated with a T cell–depleting strategy using Campath-1H. The response of RA patients to this wide spectrum of therapeutic strategies attests to the complexity and heterogeneity of the syndrome and provides further impetus for studies that use these therapies to enhance our understanding of disease pathogenesis.

Although both rituximab and abatacept have been shown to benefit a significant proportion of RA patients whether naïve to DMARDs or DMARD experienced, with either biologic, nonbiologic DMARDs or both, to date, little is known about the comparative effects of these treatment approaches on symptoms and signs of disease or on inhibition of structural damage. Furthermore, despite some pointers to biomarkers that might optimize treatment outcomes with a particular therapeutic approach at the cohort level, such as seropositivity in RA patients treated with rituximab, there are as yet no biomarkers that reliably inform the best choice of therapy on an individual basis. Rituximab and abatacept are a welcome addition to the biologic armamentarium for RA, and both agents may also have a role in the pharmacologic management of rheumatoid disorders beyond RA alone. It is anticipated that further clinical research and experience in the use of these biologic therapies will help better inform optimal treatment strategies.

References


The references for this chapter can also be found on www.expertconsult.com.
Antihyperuricemic Agents
TED R. MIKULS

KEY POINTS

Urate-lowering therapy (ULT) is central to the management of hyperuricemia in gout.

The goal of ULT administration in gout is to reduce the frequency of acute flares and prevent progressive joint destruction and tophaceous deposition; this goal can be achieved by lowering and maintaining serum uric acid (UA) below 5 to 6 mg/dL.

Optimal ULT requires careful patient selection with attention to comorbid illness, ongoing education, and effective anti-inflammatory prophylaxis with treatment initiation.

Gout is among the most common forms of inflammatory arthritis, affecting between 1% and 3% of most populations with prevalence rates as high as 6% to 7% in older men. The health burden posed by gout is likely to grow with reports suggesting up to a twofold increase in prevalence in recent decades alone. The rapid growth in gout incidence appears to be largely attributable to a near endemic increase in hyperuricemia. Hyperuricemia is the serum concentration above which uric acid (UA) precipitates into monosodium urate crystals typically defined as serum levels of 6.8 mg/dL or 405 µmol/L or greater, resulting from renal “underexcretion” in most cases (≈80%) with the remainder due to “overproduction.” UA is the end-product of purine degradation, and circulating concentrations reflect the complex interactions of dietary purine intake, endogenous production, and elimination.

Treatment of chronic gout is based primarily on the use of urate-lowering therapy (ULT). Available ULTs include (1) xanthine oxidase (XO) inhibitors (allopurinol, febuxostat), (2) uricosurics (probenecid, benz bromarone, sulfinpyrazone), and (3) uricases (pegloticase) (Table 65-1). This chapter focuses primarily on the most common rheumatic indication for ULT—treatment of hyperuricemia in gout. Optimal use of ULT in gout, regardless of the specific agent used, requires careful consideration of several factors, as detailed in the following sections.

NONPHARMACOLOGIC TREATMENT OF HYPERURICEMIA

The importance of education and lifestyle advice pertaining to weight loss, select dietary restrictions, and reduced alcohol intake has been emphasized in recent gout management guidelines. Despite a growing list of dietary factors implicated in hyperuricemia and gout, investigations of the effects of dietary interventions on health outcomes in gout are lacking. Furthermore, evidence suggests that such interventions in isolation yield only modest results and suffer from lack of widespread patient acceptance. In addition to dietary modifications that include reducing the intake of dietary purines, fructose, beer, and liquor, weight loss represents an important goal for overweight gout patients. Although weight loss may lead to reductions in serum UA, its impact on serum UA appears to be modest and may be insufficient alone in many gout patients.

Patient Selection and Timing of Treatment Initiation

Consensus indicates that hyperuricemia should be treated in gout patients with recurrent and frequent flares, tophi, and/or radiographic changes consistent with gout. For patients with acute gout, ULT should be initiated only after the acute inflammation has resolved because ULT initiation during a flare can amplify the duration and magnitude of symptoms. Conversely, the development of a gout flare complicating ongoing ULT is not an indication to discontinue or hold ULT. Results from a cost-effectiveness analysis in nontophaceous gout, involving a hypothetical patient cohort, suggest that the institution of allopurinol is cost-effective in patients presenting with two or more acute gout flares within a 1-year period. Available ULTs are not approved for the treatment of asymptomatic hyperuricemia in the absence of gout.

Duration of Urate-Lowering Therapy

In asymptomatic gout patients successfully treated with prior ULT, withdrawal of therapy often results in an abrupt increase in serum UA, and recurrent attacks occur in approximately one-third of patients within 2 years. Similarly, reductions from continuous ULT to an “intermittent” regimen in previously stable gout patients lead to significantly higher flare rates, and ULT discontinuation in the setting of tophaceous gout leads to recurrent gout flares in a vast majority and to recurrent tophi in nearly half of patients. Taken together, these reports suggest that ULT administration should be unabated and “lifelong” in a majority of gout patients.

Target Serum Urate Goals

Evidence suggests that lowering and sustaining serum UA below 6.0 mg/dL (<360 µmol/L), a treatment goal advocated in recent gout treatment guidelines, leads to improved long-term outcomes in gout. ULT reduces the long-term risk of recurrent flare by approximately 60% for each 1 mg/dL
| Table 65-1  Dosing and Safety Information for Currently Available Urate-Lowering Therapies in the Management of Gout |
|---|---|---|---|---|---|---|
| **Xanthine Oxidase Inhibitors** | **Dose** | **Route/Schedule** | **Half-Life** | **Primary Site of Metabolism/Elimination** | **Adverse Effects** | **Contraindications (C)/Drug Interactions (DI)** |
| Allopurinol | 100-800 mg | Orally daily | 1-2 hr (half-life of active metabolite oxypurinol, 15-30 hr) | Met: hepatic xanthine oxidase (into oxypurinol) Elim: renal (renal dosing may be required) | Common: gout flare, skin rash, nausea, diarrhea, LFT abnormalities Rare: allopurinol hypersensitivity syndrome (AHS) (more common in HLA-B*5801+), cytopenias | C: concomitant azathioprine, 6-MP, theophylline, prior hypersensitivity DI: azathioprine, 6-MP, theophylline, ampicillin/amoxicillin, uricosurics, thiazides, cyclosporine, warfarin, ACE inhibitors (possible), Dilantin, cyclophosphamide, vidarabine |
| Febuxostat | 40-120 mg | Orally daily | 6-8 hr | Met: hepatic (glucuronyl conjugation and oxidation via cytochrome P450) Elim: hepatic and renal | Common: gout flare, skin rash, nausea, arthralgias, LFT abnormalities Rare: cardiovascular events (unclear association), cytopenias | C: concomitant azathioprine, 6-MP, theophylline, prior hypersensitivity, severe hepatic impairment DI: azathioprine, 6-MP, theophylline |
| **Uricosurics** | | | | | Adverse effects, contraindications, drug interactions similar for available uricosurics—grouped together below. |
| Probenecid | 500-2000 mg | Orally twice daily | 3-8 hr (500 mg), 6-12 hr (larger doses) | Met: hepatic (hydroxylation) Elim: hepatic and renal | Common: gout flare, nephrolithiasis, rash, flushing, nausea, loss of appetite Rare: cytopenias, nephrotic syndrome, anaphylaxis, back pain (rare reports of hepatotoxicity with benzbromarone) | C: prior hypersensitivity, nephrolithiasis, UA overexcretion, concomitantly with other cancer therapies, known blood dyscrasias, active peptic ulcer disease; sulfinpyrazone should be avoided in patients with phenylbutazone/pyrazole allergy DI (more extensive for probenecid/sulfinpyrazone than for benzbromarone): warfarin, allopurinol, NSAIDs, salicylates, penicillins, cephalosporins, fluoroquinolones, imipenem, rifampin, nitrofurantoin, sulfonamides, heparin, dapsone, acyclovir, ganciclovir, zidovudine, alcohol, diazoxide, mecamylamine, pyrazinamide, antineoplastic agents, clofibrate, dyphylline, diuretics, benzodiazepines, methotrexate, riboflavin, thiopental |
| Sulfinpyrazone | 200-800 mg | Orally twice daily | 3-12 hr | Met: hepatic (CYP2C9) Elim: hepatic and renal | Common: gout flare, nephrolithiasis, rash, flushing, nausea, loss of appetite Rare: cytopenias, nephrotic syndrome, anaphylaxis, back pain (rare reports of hepatotoxicity with benzbromarone) | C: prior hypersensitivity, nephrolithiasis, UA overexcretion, concomitantly with other cancer therapies, known blood dyscrasias, active peptic ulcer disease; sulfinpyrazone should be avoided in patients with phenylbutazone/pyrazole allergy DI (more extensive for probenecid/sulfinpyrazone than for benzbromarone): warfarin, allopurinol, NSAIDs, salicylates, penicillins, cephalosporins, fluoroquinolones, imipenem, rifampin, nitrofurantoin, sulfonamides, heparin, dapsone, acyclovir, ganciclovir, zidovudine, alcohol, diazoxide, mecamylamine, pyrazinamide, antineoplastic agents, clofibrate, dyphylline, diuretics, benzodiazepines, methotrexate, riboflavin, thiopental |
| Benzbromarone | 50-200 mg | Orally daily | 3 hr (half-life of active metabolite 6-hydroxybenzbromarone =30 hr) | Met: hepatic (CYP2C9) Elim: hepatic and renal | Common: gout flare, nephrolithiasis, rash, flushing, nausea, loss of appetite Rare: cytopenias, nephrotic syndrome, anaphylaxis, back pain (rare reports of hepatotoxicity with benzbromarone) | C: prior hypersensitivity, nephrolithiasis, UA overexcretion, concomitantly with other cancer therapies, known blood dyscrasias, active peptic ulcer disease; sulfinpyrazone should be avoided in patients with phenylbutazone/pyrazole allergy DI (more extensive for probenecid/sulfinpyrazone than for benzbromarone): warfarin, allopurinol, NSAIDs, salicylates, penicillins, cephalosporins, fluoroquinolones, imipenem, rifampin, nitrofurantoin, sulfonamides, heparin, dapsone, acyclovir, ganciclovir, zidovudine, alcohol, diazoxide, mecamylamine, pyrazinamide, antineoplastic agents, clofibrate, dyphylline, diuretics, benzodiazepines, methotrexate, riboflavin, thiopental |
| **Uricases** | | | | | | |
| Pegloticase | 8 mg* | IV every 2 weeks | Highly variable (days to weeks) | Not well defined | Common: gout flare, allergic reactions, anaphylaxis (≈7%), infusion reactions (urticaria, dyspnea, chest discomfort, pruritus) Rare: CHF exacerbation (unclear association) | C: allergic reactions to medication or loss of effect (serum UA >6.0 mg/dL indicates development of anti-pegloticase antibody) DI: other PEGylated agents (possible) |

*Predosing with fexofenadine 60 mg, acetaminophen 1000 mg, and hydrocortisone 200 mg IV on the day of the infusion.

ACE, angiotensin-converting enzyme; CHF, congestive heart failure; Elim, elimination; IV, intravenously; LFT, liver function testing; Met, metabolism; MP, mercaptopurine; NSAIDs, nonsteroidal anti-inflammatory drugs; UA, uric acid.
Anti-inflammatory Prophylaxis with Urate-Lowering Therapy

Rebound gout is the most common adverse effect with ULT regardless of the agent used, rendering anti-inflammatory prophylaxis a key component of successful gout treatment. Gout flares complicating ULT are thought to be due to reduced serum UA concentrations that result in mobilization of urate from tissue deposits. In fact, the frequency of treatment-related gout flares appears to be greater with more rapid and potent urate-lowering interventions. Both colchicine and naproxen are effective in reducing gout flares during ULT initiation (this is covered in greater detail in Chapter 95). Results from Borstad and associates suggest that anti-inflammatory prophylaxis with low-dose oral colchicine (0.6 mg twice daily) protects against rebound gout flares and may need to be continued for at least 6 months following ULT initiation. Although frequently used in gout flare prevention, the efficacy of other anti-inflammatory agents in this setting, including low-dose glucocorticoids and alternative nonsteroidal anti-inflammatory drugs (NSAIDs), has not been defined. Under active investigation for the treatment of acute gout, interleukin-1 inhibition may also represent an alternative means of prophylaxis with ULT initiation.

Xanthine Oxidase Inhibition

Allopurinol

KEY POINTS

- Advantages of allopurinol over the most commonly used uricosurics include once-daily dosing, effectiveness in both “underexcretors” and “over producers,” and potential effectiveness in patients with renal insufficiency.
- Allopurinol hypersensitivity syndrome (AHS) is a relatively uncommon albeit potentially serious adverse event associated with its use.
- Allopurinol doses greater than 300 mg/day are frequently required to achieve target serum UA goals; evidence and consensus-based gout treatment guidelines recommend the initial use of low-dose allopurinol (≤100 mg/day) with gradual increases in dosing to achieve target serum uric acid (UA) goals.

Available for more than 40 years, allopurinol accounts for a vast majority of ULT prescriptions. In addition to its established track record in gout care, allopurinol offers several potential advantages: (1) relatively low cost, (2) once-daily oral administration in most cases, (3) effectiveness in patients who “underexcrete” and in those who “overproduce” UA, (4) a favorable safety profile, and (5) potential effectiveness in patients with renal impairment. Allopurinol treatment results in significant declines in serum UA, decreased gout flare rates, and declines in tophus area.

Role in Rheumatic Disease and Indications

Approved indications for allopurinol include (1) treatment of hyperuricemia in gout, (2) management of malignancy (most often leukemia or lymphoma) in patients undergoing cancer treatment that results in marked increases in serum/urinary UA (allopurinol is stopped once UA overproduction is absent), and (3) management of calcium oxalate–associated nephrolithiasis with daily urinary UA excretion greater than 800 mg/day in men and greater than 700 mg/day in women. Although not approved for the treatment of asymptomatic hyperuricemia in the absence of gout, evidence suggests that allopurinol use may lead to other health benefits. Hyperuricemia is independently associated with cardiovascular morbidity and mortality, giving rise to speculation that ULT could be cardioprotective. In a placebo-controlled study of pediatric essential hypertension, allopurinol use resulted in significant, albeit modest, declines in blood pressure. XO inhibition via allopurinol also has been shown to improve endothelial function, improving measures of local and systemic blood flow, and has been associated with improvement in renal function in at-risk populations.

Chemical Structure and Mechanism of Action

An antimetabolite in simple organisms, allopurinol inhibits XO, a key enzyme in purine catabolism, but does not inhibit the biosynthesis of purines in humans (Figures 65-1 and 65-2).

Pharmacology

Pharmacologic characteristics of the ULTs are summarized in Table 65-1. Allopurinol is approximately 90% absorbed from the gastrointestinal (GI) tract and is metabolized into oxypurinol, its active metabolite. Peak serum concentrations for allopurinol and oxypurinol are achieved within approximately 1 to 2 hours and 4 to 5 hours, respectively. The plasma half-life of allopurinol is relatively brief (1 to 2 hours), although the half-life of oxypurinol is substantially longer (~15 hours or longer), allowing for once-daily dosing. Allopurinol is eliminated primarily via glomerular filtration, whereas oxypurinol undergoes some degree of renal tubular reabsorption. With renal mechanisms primarily responsible for drug elimination, the plasma half-life of allopurinol and, to a greater degree, oxypurinol is increased with renal injury.

Dose and Drug Administration

Allopurinol is available as 100-mg and 300-mg pills for once-daily oral administration (see Table 65-1), with recognition that split dosing has been advocated for daily
of allopurinol-treated gout patients achieved a final study urate level of less than 6.0 mg/dL.\textsuperscript{18,19} Recent gout treatment guidelines recommend that allopurinol should be started at low doses (e.g., 100 mg daily) and increased by 100 mg every 2 to 4 weeks as required to achieve a target urate concentration.\textsuperscript{4} Combined data from two studies suggest that each 100-mg increment in allopurinol is associated with an additional decline in serum urate of approximately 1.0 mg/dL.\textsuperscript{38,39} Although no data directly support the “start low and go slow” approach over the “fixed-dose” approach, it has been suggested that the former strategy could reduce the incidence of rebound gout flares and mitigate treatment-related toxicity.\textsuperscript{4}

General consensus indicates that initial allopurinol dosing should be adjusted for diminished renal function.\textsuperscript{4,7,14}
which prolongs the plasma half-life of oxypurinol. With age-related increases in gout incidence, the issue of renal dosing has added relevance for the elderly with age-associated declines in kidney function. Commonly cited dosing algorithms suggest administering an initial daily dose of 100 mg or less for patients with a glomerular filtration rate (GFR) below 20 mL/min, and even lower doses for those with more severe renal impairment. Whether existing dosing guidelines preclude the use of incremental dosing beyond these recommended renal thresholds is controversial. Renal dosing guidelines that have been promulgated are not evidence based and are founded largely on a single retrospective case series showing that patients who developed allopurinol hypersensitivity syndrome (AHS) were more likely to have renal insufficiency. Indeed, many patients with chronic kidney disease (CKD) have developed AHS even with “appropriately” dosed allopurinol. In their review of 120 gout patients receiving allopurinol, more than half (57%) required daily doses above the “renal threshold” recommended by Hande and co-workers—a strategy reported to be well tolerated in a majority of patients.

Toxicity

AHS, an uncommon but potentially fatal treatment complication, is a febrile illness characterized by the presence of an erythematous desquamating rash (similar to Stevens-Johnson), eosinophilia, and end-organ damage, including hepatitis and renal failure. It has been estimated that AHS complications between 0.1% and 0.4% of allopurinol treatment courses, although rates have been estimated to be as low as 1 in 56,000 allopurinol users. AHS appears to be substantially more common in individuals positive for HLA-B*5801. In a small case-control study, all patients who developed AHS were positive for HLA-B*5801 compared with just 13% of allopurinol-treated patients without AHS. This risk allele is seen in approximately 2% to 7% of whites, 7% of blacks, and 8% of Asian Indians. Given the potential severity of AHS, patients should be educated about the remote possibility of this adverse event and cautioned to discontinue allopurinol with development of rash, particularly if this is accompanied by fever or mucocutaneous lesions.

Given the rare occurrence of AHS, allopurinol is generally well tolerated, with estimates suggesting that less than 5% to 10% of those exposed are intolerant to the drug. Rebound flares are among the most common adverse event accompanying allopurinol and other ULTs—an issue that is most prominent in the early phases of drug initiation and can be mitigated by anti-inflammatory prophylaxis. Isolated maculopapular skin rash can occur outside the context of AHS and is estimated to complicate approximately 1% to 3% of allopurinol treatment courses. Other common adverse events associated with allopurinol use are summarized in Table 65-1. Liver function abnormalities can be seen in approximately 6% to 7% of allopurinol users, although rates of severe liver injury appear to be exceedingly rare. The role and recommended frequency for laboratory surveillance in toxicity monitoring have not been well defined.

Fertility, Pregnancy, and Lactation

Although no human studies have investigated its use in pregnancy, allopurinol is classified as a Pregnancy Category C agent (animal reproduction studies have shown an adverse effect on the fetus without adequate human studies). Both allopurinol and oxypurinol are expressed in breast milk, and because drug effects on the developing infant are largely unknown, it should be administered to a nursing mother with caution.

Drug Interactions and Contraindications

Allopurinol drug interactions have been well characterized. Azathioprine and 6-mercaptopurine (6-MP) are metabolized primarily by XO, hence co-administration of allopurinol results in marked increases in circulating drug levels that can lead to bone marrow suppression. Theophylline is also metabolized by XO; therefore co-administration of this agent with allopurinol can lead to increased theophylline levels and can potentiate toxicity. Co-administration of allopurinol with ampicillin/amoxicillin has been associated with a higher incidence of drug-related rash. Thiazide diuretics may also reduce the renal excretion of allopurinol and oxypurinol, and it has been suggested that this could potentiate drug-related toxicity. Uricosurics increase the renal excretion of oxypurinol, thus offsetting to some degree the urate-lowering effect of allopurinol treatment. Allopurinol co-administration may increase drug levels of cyclosporine and warfarin, mandating close monitoring of drug levels and bleeding parameters, respectively. Other allopurinol-associated drug interactions are summarized in Table 65-1. Although regimens for desensitization have been described, allopurinol should be avoided in patients with known allergies such as AHS.

Febuxostat

KEY POINTS

Febuxostat is a potent inhibitor of XO with a chemical structure that is distinct from allopurinol.

Febuxostat represents an alternative for patients for whom allopurinol is not effective because of intolerance or lack of efficacy.

Febuxostat, a potent and selective inhibitor of XO, represents the first ULT approved for gout treatment in more than 40 years.

Role in Rheumatic Disease and Indications

Febuxostat is approved for the treatment of hyperuricemia in patients with gout; similar to all other ULTs, febuxostat is not indicated for the treatment of asymptomatic hyperuricemia. Given its unique structure, febuxostat represents an important alternative means of XO inhibition, particularly in gout patients intolerant to allopurinol. The role of febuxostat in gout management was recently addressed by an international panel, which provided guidance specific to
indications, contraindications, monitoring, and issues requiring future research.

Chemical Structure and Mechanism of Action

In contrast to allopurinol, febuxostat is a nonpurine analogue that reduces serum and urinary urate concentrations through potent and selective XO inhibition (see Figures 65-1 and 65-2). In contrast to allopurinol, which may inhibit other enzymes involved in purine and pyrimidine synthesis, febuxostat demonstrates significant enzymatic inhibition only for XO at therapeutic concentrations.

Pharmacology

Following oral administration, febuxostat is rapidly absorbed from the GI tract with approximately 50% absorption, reaching peak plasma concentrations within a few hours with near complete plasma protein binding (see Table 65-1). Febuxostat displays linear pharmacokinetics that are not time dependent with drug metabolism occurring primarily in the liver through conjugation via uridine diphosphate glucuronosyltransferase (UGT) and oxidation via cytochrome P450 enzymes. Peak urate-lowering effects with febuxostat generally occur during the first 5 to 7 days of treatment. Drug elimination occurs via both hepatic and renal pathways. Although active metabolites are produced via oxidation, these are present in much lower plasma concentrations.

Dose and Drug Administration

Febuxostat is available in 40-mg (United States), 80-mg (United States and Europe), and 120-mg (Europe) tablets for oral administration, with usual dosing ranging from 40 mg to 120 mg daily (see Table 65-1). Febuxostat should be initiated at a lower dose (40 mg to 80 mg daily) and increased to higher doses (80 mg to 120 mg daily) if serum UA remains greater than 6.0 mg/dL after 2 weeks. In a phase II study, serum UA less than 6.0 mg/dL was obtained by 56%, 76%, and 94% of patients receiving 40 mg, 80 mg, and 120 mg per day of febuxostat, respectively, compared with 0% in the placebo-treated group. Mean serum UA reductions were greater with higher daily doses, ranging from a 37% reduction in the 40 mg/day group to 59% in the 120 mg/day group. Two subsequent trials lasting 28 weeks (n = 1067) and 52 weeks (n = 762) compared febuxostat (80 mg to 240 mg per day) with fixed-dose allopurinol (300 mg daily); both trials used the primary outcome of obtaining a serum UA less than 6.0 mg/dL at the last three consecutive monthly observations. The primary endpoint was achieved by 48% to 53%, 62% to 65%, and 69% of gout patients receiving daily doses of 80 mg, 120 mg, and 240 mg, respectively, compared with 21% and 22% of patients receiving fixed-dose allopurinol. Secondary outcomes in the trial by Becker and associates showed a decline in gout flare rates and an approximately 70% to 80% reduction in gout tophus area over follow-up—differences that were not significantly different from those observed with allopurinol. It is noteworthy that several studies of febuxostat employed fixed-dose allopurinol as an active comparator. As detailed previously, current gout treatment guidelines recommend initiating low-dose allopurinol (i.e., 100 mg/day) with escalations in dosing as needed to achieve target urate thresholds. Because optimal allopurinol dosing strategies were not used in these studies, these results likely overestimate the effectiveness of febuxostat relative to optimally dosed allopurinol.

Metabolized primarily in the liver, febuxostat may not require renal dosing. This receives support from a few small short-term pharmacokinetic studies that included a small number of patients with renal impairment. Only limited data have been obtained from longer-term studies of febuxostat in patients with moderate renal impairment (serum creatinine, 1.6 to 2.0 mg/dL), and essentially no data are available from patients with more severe renal dysfunction (serum creatinine >2.0 mg/dL). Available data, albeit limited, suggest that those with mild or moderate renal impairment (estimated creatinine clearance [CrCl] between 30 and 90 mL/min) experience similar efficacy and similar rates of toxicity compared with those with preserved renal function receiving equivalent doses of febuxostat.

Toxicity

Similar to other ULTs, including allopurinol, rebound gout flares are the most common complication of febuxostat administration, underscoring the importance of anti-inflammatory prophylaxis. Other adverse reactions observed with febuxostat are summarized in Table 65-1; adverse effects appear to occur at similar rates as those observed with allopurinol. In initial studies comparing febuxostat with allopurinol, a slightly higher rate of cardiovascular events was observed in patients receiving febuxostat (0.74 events per 100 patient-years; 95% confidence interval [CI], 0.36 to 1.37) compared with those randomized to allopurinol (0.60 events per 100 patient-years; 95% CI, 0.16 to 1.53). In the large 6-month CONFIRMS trial, investigators found no differences in rates of cardiovascular events between febuxostat (40 mg/day and 80 mg/day) and allopurinol (200 mg/day to 300 mg/day), given the limited power of this study, with only six adjudicated events occurring during follow-up. In this study, withdrawal rates due to adverse events were similar across treatment groups: 6.4% for febuxostat 40 mg/day, 8.1% for febuxostat 80 mg/day, and 8.5% for allopurinol.

Owing to its unique structure and select XO inhibition, it has been suggested that febuxostat may represent a rational alternative for gout patients with a history of allopurinol hypersensitivity. Whether febuxostat can be effectively administered to such patients is not clear, given early reports of hypersensitivity in patients taking febuxostat. In a small retrospective study involving 13 gout patients with a history of severe allopurinol-related adverse reactions, 12 were subsequently treated safely with febuxostat (with 10 achieving target serum urate goals). A single patient developed hypersensitivity vasculitis of the skin.

Fertility, Pregnancy, and Lactation

The effects of febuxostat on fertility are not known. Likewise, no studies have examined the use of febuxostat in pregnant women, although results from animal studies have not suggested a significant risk of teratogenicity. In the
absence of appropriate human studies, febuxostat is labeled as Pregnancy Category C. It is not known whether the drug is excreted in human milk, and febuxostat should be used only with caution in nursing women because its effects on developing infants are unknown.

Drug Interactions and Contraindications

Although formal drug interaction studies have not been performed, febuxostat should be used with caution with drugs that are metabolized by XO (azathioprine, 6-MP, theophylline) (see Table 65-1). Its use should be avoided in patients with prior hypersensitivity to febuxostat. Owing to hepatic metabolism, febuxostat should not be used in patients with moderate to severe liver impairment.35

**URICOSURIC AGENTS**

<table>
<thead>
<tr>
<th>KEY POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probenecid, sulfinpyrazone, and benzbromarone are the most common uricosuric agents used in gout treatment worldwide.</td>
</tr>
<tr>
<td>Uricosurics are potentially effective in lowering serum urate levels in patients who underexcrete uric acid, the most common pathologic defect leading to hyperuricemia.</td>
</tr>
<tr>
<td>Probenecid and sulfinpyrazone have limited efficacy in the context of renal insufficiency.</td>
</tr>
</tbody>
</table>

Uricosurics, which were introduced more than 50 years ago, represent the first class of ULT to be used in gout treatment.62 A complex system of renal handling of urate has been proposed that sequentially includes four components:63

1. near complete glomerular filtration,
2. proximal reabsorption,
3. tubular urate secretion occurring more distally, and
4. reabsorption a second time more distally. Available uricosurics diminish the postsecretory reabsorption of UA, therefore promoting its elimination and reducing circulating urate concentrations. Uricosurics address the most common physiologic defect in gout—UA underexcretion. In addition to promoting renal UA excretion, uricosurics may inhibit the tubular secretion of a number of other compounds, including penicillins. Although many agents display uricosuric properties, the most commonly used uricosurics in gout worldwide include probenecid, sulfinpyrazone, and benzbromarone. Probenecid is the most widely used uricosuric in gout treatment. Owing primarily to concerns of treatment-related toxicity, sulfinpyrazone and benzbromarone are less widely available; neither is currently available in the United States.

Although the focus of this section is on primary uricosurics, several medications have been approved for the treatment of nongouty conditions that have been shown to display uricosuric properties. Secondary uricosurics and their indications are summarized in Table 65-2.64-69 Salicylates exert a paradoxical effect on renal urate elimination, with inhibition of active secretion with lower doses (e.g., <1 g/day) and “uricosuric-like” inhibition of UA reabsorption at higher doses (>4 to 5 g/day). Along with salicylates, losartan has perhaps been most closely scrutinized for its hypouricemic properties—an effect that is mediated via inhibition of URAT1 and appears to be specific to this angiotensin receptor blocker (ARB).65 The magnitude of UA lowering that can be achieved with these secondary hypouricemic agents is modest. Losartan used at an antihypertensive dose (50 mg/day) resulted in a mean approximately 9% decline in serum UA concentrations.66 It is important to recognize that the urate-lowering effect of losartan appears to be negated when co-administered with hydrochlorothiazide,67 a combination used in hypertension treatment.

Role in Rheumatic Disease and Indications

Probenecid, sulfinpyrazone, and benzbromarone are used to treat hyperuricemia associated with gout and are indicated in gout patients who underexcrete UA (24-hour urine UA, <700 mg). Probenecid is also approved as an adjuvant to penicillin therapy, increasing plasma concentrations and prolonging the terminal half-life of penicillin and other penicillin derivatives. Although potentially effective for most gout patients, uricosurics are much less commonly used than allopurinol. In a population-based study from the United Kingdom, uricosuric therapy accounted for less than 5% of all ULT prescriptions.1

Chemical Structure and Mechanism of Action

The chemical structures of probenecid, sulfinpyrazone, and benzbromarone are shown in Figure 65-2. Uricosurics work primarily through inhibition of the renal tubular urate-anion exchanger, URAT1 (SLC22A12), in addition to GLUT9 (SLC2A9). Thus, uricosurics reduce UA reabsorption, promote renal elimination, and decrease circulating urate concentrations (Figure 65-3).71-75

Pharmacology

Administered orally, the uricosurics undergo GI absorption and are extensively protein bound in the serum. The half-lives of probenecid66 and sulfinpyrazone77 are relatively brief, ranging from approximately 3 to 12 hours (see Table 65-1). Although benzbromarone also has a relatively short half-life of approximately 3 hours, its active metabolite, 6-hydroxybenzbromarone, has a much longer half-life, allowing for effective once-daily administration.78 Elimination of the uricosurics occurs primarily via hepatic metabolism followed by variable excretion of metabolites in the urine, bile, and/or feces. The metabolism of probenecid is limited to side-chain hydroxylation, N-depropylation, and glucuronic conjugation of its carboxyl group.79,81

<table>
<thead>
<tr>
<th>Table 65-2</th>
<th>Agents Not Approved for Gout Treatment That Display Urate-Lowering Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Agent</strong></td>
<td><strong>Usual Indication</strong></td>
</tr>
<tr>
<td>Losartan</td>
<td>Hypertension, congestive heart failure</td>
</tr>
<tr>
<td>Fenofibrate</td>
<td>Hyperlipidemia, hypertriglyceridemia</td>
</tr>
<tr>
<td>Atorvastatin</td>
<td>Hyperlipidemia</td>
</tr>
<tr>
<td>Rosuvastatin</td>
<td>Hyperlipidemia</td>
</tr>
<tr>
<td>Guainifensin</td>
<td>Upper respiratory airway congestion</td>
</tr>
<tr>
<td>Leflunomide</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Salicylates (high-dose)</td>
<td>Analgesia, fever, anti-inflammatory</td>
</tr>
</tbody>
</table>
Both sulfinpyrazone and benzbromarone are metabolized via cytochrome P450 (CYP), with CYP2C9 representing a principal mediator. Warfarin is also metabolized primarily via CYP2C9, explaining its drug-drug interactions with both sulfinpyrazone and benzbromarone.

**Dose and Drug Administration**

Usual probenecid dosing is 500 to 2000 mg daily in divided doses. Probenecid (500 mg) is also formulated with colchicine (0.5 mg) as a combination tablet (colbenemid). In a 12-week randomized study involving patients with preserved renal function, probenecid 1500 mg/day was associated with a 32% reduction in plasma urate concentration, and older studies have documented other qualitative benefits of therapy, including softening of tophi, functional improvement, and improved pain symptoms. Sulfinpyrazone is given in divided doses at an initial daily dose of 200 to 400 mg, increasing to 800 mg daily, if necessary, to achieve target serum urate goals. Although there is no level of renal function below which probenecid and sulfinpyrazone are known to completely lose efficacy, these drugs appear to be substantially less effective in patients with moderate to severe renal insufficiency (CrCl <60 mL/min). Loss of hypouricemic effect with these agents with advancing renal impairment limits their utility in gout, given the relatively high prevalence of impaired renal function in this population.

In contrast to probenecid and sulfinpyrazone, benzbromarone is administered once daily and may be effective even in patients with moderate renal impairment. Benzbromarone treatment in usual doses (50 to 200 mg/day) leads to between 25% and 50% reductions in circulating UA, in addition to decreases in gout flare rates and dissolution of tophi. In a randomized controlled trial, benzbromarone maintained its hypouricemic effect even among patients with a CrCl as low as 20 to 40 mL/min. Among patients intolerant to or experiencing treatment failure with allopurinol 300 mg/day, 92% of those administered benzbromarone (200 mg/day) achieved a serum urate level less than 5.0 mg/dL compared with 65% of those receiving probenecid (2000 mg/day).

Uricosurics may also serve as effective adjuvants to XO inhibition. Although uricosurics enhance excretion of oxypurinol, the active metabolite of allopurinol, their effect on oxypurinol clearance does not appear to negate the additive urate-lowering effect gained by using these medications in combination. For instance, among patients failing allopurinol monotherapy (200 to 300 mg/day), 86% achieved a target serum urate goal of less than 5 mg/dL after the addition of probenecid (1000 mg/day). Likewise, the combination of benzbromarone and allopurinol has been shown to...
have a more potent urate-lowering effect than allopurinol alone, and the combination has been shown to yield more rapid resolution of tophi than allopurinol monotherapy.104

**Toxicity**

As with other ULTs, rebound gout flares represent a common complication of uricosurics. Because urinary UA acts as a potential nidus for stone formation, uricosuric therapy is associated with increased risk of nephrolithiasis. In a recent longitudinal study including more than 780 patient-years of benzbromarone exposure, approximately 10% (21 of 216) of patients developed nephrolithiasis.97 Of the 21 patients with incident nephrolithiasis, 7 developed oxalate stones and 14 developed stones composed of UA or a combination of UA and calcium. In addition to limiting uricosurics to patients underexcreting UA, the risk of treatment-related nephrolithiasis can be mitigated by optimizing fluid intake and alkalinizing the urine with a goal of maintaining a urine pH greater than 6.0.98 Other adverse effects observed with uricosurics are summarized in Table 65-1. Rare side effects associated with uricosurics have included anaphylaxis, anemia (including aplastic and hemolytic anemia), other cytopenias, fever, nephrotic syndrome, and back pain. Rare reports of severe liver injury led to the withdrawal of benzbromarone by its primary manufacturer in 2003, a decision recently challenged in systematic risk-benefit analyses of the drug.99 Owing to its NSAID-like properties, sulfinpyrazone has been associated with an increased frequency of blood dyscrasias (rare), in addition to upper GI disturbances including peptic ulcer disease.100

**Fertility, Pregnancy, and Lactation**

Limited data are available regarding the impact of uricosurics on fertility, fetal development, and use in nursing infants; thus, these agents should be used with caution in such patients only when the potential benefits of treatment outweigh its potential risks.

**Drug Interactions and Contraindications**

Given that tubular secretion plays a central role in the renal clearance of numerous drugs, drug-drug interactions are well recognized with the use of uricosurics, particularly probenecid (see Table 65-1). Inhibition of tubular drug secretion appears to be greater for probenecid and sulfinpyrazone than for benzbromarone owing to increased URAT1 “specificity” of the latter agent. Drug interactions observed with uricosurics are summarized in Table 65-1.

As with all ULTs, uricosurics should not be initiated during an acute gout flare and should not be used in patients with known allergies to these agents. Because of the potential for cross-reactivity, sulfinpyrazone should be avoided in patients with allergies to phenylbutazone or other pyrazole compounds. Except under special circumstances, these agents should be avoided in patients with nephrolithiasis or evidence of UA overexcretion, and during cancer treatment, including chemotherapy or radiation. Caution should also be used in patients with known blood dyscrasias, active peptic ulcer disease, or significant hepatic or renal disease.

---

**URICASES**

**Pegloticase**

**KEY POINTS**

- Pegloticase facilitates the conversion of UA into allantoin, which is far more soluble.
- Intravenous administration of pegloticase is associated with rapid and marked declines in serum UA concentration.
- Drug-related antigenicity represents a major limitation in the repeated dosing of pegloticase.

Pegloticase, a modified mammalian uricase, is a biologic parenterally administered agent that represents the most recently approved ULT in gout. Unlike other mammalian species, humans have lost the ability to synthesize functional uricase that converts UA into allantoin, the latter being 5 to 10 times more soluble. Use of alternative uricases (e.g., rasburicase—a recombinant uricase from *Aspergillus flavus*100) in the treatment of tumor lysis syndrome has been substantially limited by drug-related antigenicity and prohibitively high rates of anaphylaxis with repeat drug administration. In contrast to older-generation uricases, pegloticase appears to be less allergenic and has been administered successfully to many patients in repeated intravenous infusions.

**Role in Rheumatic Disease and Indications**

Pegloticase is approved for the treatment of hyperuricemia in patients with treatment-refractory gout. Approved with an orphan drug status in the United States, pegloticase is indicated in a small subset of gout patients. Treatment-refractory disease is characterized by severe disabling gout, often accompanied by significant comorbid illness, in which conventional ULT may be contraindicated or ineffective.101 Pegloticase administration is associated with rapid and marked declines in serum UA; in clinical trials, pegloticase use resulted in serum UA reduction with nadirs as low as 0.5 to 1 mg/dL within 24 hours of an initial dose.102,103 As a consequence, pegloticase administration has been associated with dramatic regression of tophi and depletion of urate stores, raising speculation that pegloticase or other uricase formulations could play a role as an “induction” therapy in select patients with severe tophaceous gout.

**Chemical Structure and Mechanism of Action**

Pegloticase is a recombinant mammalian uricase linked to polyethylene glycol (PEG). Pegloticase facilitates the conversion of UA into allantoin. The conversion of allantoin from UA generates hydrogen peroxide (H$_2$O$_2$) that is scavenged by erythrocytes and thus does not appear to increase levels of oxidative stress.104

**Pharmacology**

The pharmacokinetics of pegloticase follows a one-compartment linear model. Maximum serum concentrations and the magnitude of the urate-lowering effect
following intravenous pegloticase administration increase in a dose-dependent fashion. Pegloticase pharmacokinetics are not affected by age, sex, weight, or underlying renal function. A single dose of pegloticase (≥2 mg) has been shown to have a highly variable duration of effect, suppressing serum urate concentrations below 7.0 mg/dL for periods ranging from 1 to 8 days. The high degree of variability in drug elimination and treatment durability appears to be related at least in part to the presence and concentration of circulating anti-pegloticase antibody (see later).

**Dose and Drug Administration**

Pegloticase is approved for intravenous infusion at a dose of 8 mg every 2 weeks (see Table 65-1). The agent is administered as a 2-hour infusion. Regulatory approval of pegloticase is based on results from two 6-month replicate randomized placebo-controlled studies, in addition to open-label follow-up extensions (N = 212 total). Patients in these clinical trials had pretreatment serum urate concentrations exceeding 8.0 mg/dL and had symptomatic gout with failure of prior allopurinol therapy based on reported intolerance or ineffectiveness at a “maximum medically appropriate dose.” Patients received prophylaxis against infusion reactions (see Table 65-1) and against rebound gout flares (colchicine, NSAID, or glucocorticoid). The primary outcome for both randomized studies was a serum UA less than 6.0 mg/dL for 80% of more of the sampling period from month 3 to month 6 of follow-up. This outcome was met by 47% and 38% of subjects receiving pegloticase 8 mg intravenously (IV) every 2 weeks in the two randomized studies. In pooled analyses, 40% of subjects with baseline tophi who received pegloticase 8 mg every 2 weeks (vs. 7% of those receiving placebo) had complete resolution of their tophi by the time of the final study visit.

Although nonresponders and responders generally achieved similarly striking plasma urate reductions following their initial pegloticase infusion, nonresponders generally “lost” treatment effect within the first 3 months of therapy. Efficacy loss, reflected in serum UA rising above 6.0 mg/dL during treatment, appears to be most striking at antibody titers exceeding 1:2430. It is recommended that serum urate levels, rather than drug antibody titers, be monitored closely during treatment, with discontinuation of pegloticase if serum UA increases above 6.0 mg/dL.

**Toxicity**

The most common serious adverse event observed with pegloticase therapy is anaphylaxis, observed in approximately 7% of patients (see Table 65-1). Anti-pegloticase antibody can be detected in a vast majority of patients (≥90%) receiving treatment, although “clinically meaningful” antibody titers are encountered less often. In clinical studies, anaphylaxis occurred despite prophylaxis that included antihistamine and glucocorticoid therapy with onset of symptoms typically seen within 2 hours of drug administration. Treatment-related anaphylaxis is far more common among patients with treatment failure; thus, close serum UA surveillance and discontinuation of pegloticase for patients with serum UA greater than 6.0 mg/dL are essential elements in risk mitigation. Given overlap with symptoms of anaphylaxis, infusion reactions occur in up to 25% of patients receiving pegloticase with manifestations that include urticaria, dyspnea, chest discomfort/pain, erythema, and pruritus. Infusion reactions can occur at any time in the course of therapy, and rare reports have described delayed hypersensitivity reactions.

As with other ULTs, rebound gout flare is common following administration of pegloticase. Other adverse events observed with pegloticase are summarized in Table 65-1. In clinical studies and in open-label follow-up, exacerbations of congestive heart failure were also more common with pegloticase than placebo, although a causal association with active drug therapy has not been established.

**Fertility, Pregnancy, and Lactation**

No studies have examined the impact of pegloticase on fertility or pregnancy in humans, and it is unknown whether pegloticase is excreted in human milk. In the absence of appropriate human studies, pegloticase is classified as Pregnancy Category C, and it is not recommended for use in nursing mothers.

**Drug Interactions and Contraindications**

Given that anti-pegloticase antibodies bind the PEG portion of the agent, pegloticase should be used with caution in patients receiving other PEG-containing therapies. Whether the formation of anti-pegloticase antibodies precludes or impacts future treatment with other PEGylated molecules is unknown. Pegloticase is contraindicated in patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency owing to increased risks of hemolysis and methemoglobinemia; patients at increased risk should be screened for G6PD deficiency before initiation of treatment.

**URATE-LOWERING DRUGS IN DEVELOPMENT**

The treatment armamentarium in chronic gout will likely grow in the future with several novel ULTs under development, including uricosurics, uricases, and agents inhibiting endogenous UA synthesis.

**Inhibitors of UA Synthesis**

BCX4208 represents a potentially novel mechanism inhibiting endogenous UA synthesis via inhibition of purine nucleoside phosphorylase (PNP). Along with adenosine deaminase, PNP plays a role in purine degradation, acting “proximally” to XO by metabolizing inosine into hypoxanthine. In a 3-week study of 60 gout patients (all with baseline serum UA >8.0 mg/dL), 31% to 36% of patients given BCX4208 (40 mg, 80 mg, and 120 mg daily) achieved a
Uricosuric Agents

Lesinurad (formerly RDEA 594) is a metabolite of RDEA 806, a non-nucleoside reverse transcriptase inhibitor. Lesinurad appears to specifically inhibit URAT1, lacking significant effects on other organic anion transporters (see Figure 65-3). Selective URAT1 inhibition may limit drug interactions that complicate probenecid use. In contrast to probenecid, the co-administration of lesinurad does not appear to increase the renal elimination of oxpurinol or febuxostat, suggesting that this may be an ideal uricosuric for use in combination. In a recent study, the addition of lesinurad to allopurinol (300 mg/day) led to dose-dependent declines in serum UA.

In addition to displaying urate-lowering properties, tranilast has been studied for possible therapeutic effects in a number of conditions, including allergy, malignancy, and conditions characterized by excessive tissue fibrosis. Similar to those of currently available uricosurics, the urate-lowering effects of tranilast are mediated through inhibition of URAT1 and GLUT9 transporters (see Figure 65-3). In a recent study, tranilast administration (300 mg/day) was associated with a 14% decline in serum UA, with 28% of patients achieving a serum UA less than 6.0 mg/dL.

Uricases

Pegsiticase is a recombinant uricase derived from Candida utilis, conjugated to PEG. Pegsiticase has been examined in an open-label, single-dose (0.05 to 0.4 mg/kg IV) escalation study in 20 gout patients. In this study, pegsiticase was well tolerated, with gout flare representing the most common adverse event. Rapid and marked declines in serum UA were observed in all patients, and UA concentrations decreased to below 2.0 mg/dL for 7 to 24 days in a dose-dependent fashion. The degree to which immunogenicity decreases in serum UA for 7 to 24 days in a dose-dependent fashion. The degree to which immunogenicity decreases to below 2.0 mg/dL for 7 to 24 days in a dose-dependent fashion. The degree to which immunogenicity decreases to below 2.0 mg/dL for 7 to 24 days in a dose-dependent fashion.

References


52. Ulicor [package insert], Deerfield, Ill, 2009, Takeda Pharmaceuticals America, Inc.


Pain is the most common reason why patients seek medical attention, yet undertreatment of acute and chronic pain persists despite decades of efforts to provide clinicians with information about analgesics. A major consequence of undertreating the patient who initially presents with pain as a result of rheumatic disease is the development of chronic pain. Chronic pain has been demonstrated to have deleterious effects on many aspects of the patient's daily life. These effects include deterioration in physical functioning, the development of psychologic distress and psychiatric disorders, and impairments in interpersonal functioning. In addition to the personal suffering it causes, chronic pain imposes a burden on society in increased health care costs, disability, and lost workdays.

**PHYSIOLOGY OF PAIN PERCEPTION (THE PAIN EXPERIENCE)**

The “pain experience” involves more than just the sensation of pain. Pain activates many areas of the brain that interact, resulting in the pain experience, which will differ among individual patients. The three components of the pain experience include (1) sensory/discriminative, (2) affective/emotional, and (3) evaluative/cognitive.

The sensory/discriminative component of pain provides information on the intensity, location, and quality of the pain. These pathways consist of peripheral receptor activation, axon depolarization, and ascending pathways to the cortex for processing.

The ascending pathways that carry impulses from the nociceptor to the sensory cortex also give off fibers to brain stem structures and deep brain structures. Activation of these structures in the brain stem and deep brain will stimulate emotional and sympathetic responses from the individual, leading to the affective/emotional component of pain.

Finally, ascending pain pathways also send projections to the forebrain structures where the pain is processed on a cognitive and evaluative level, explaining why patients respond differently to pain based on culture, gender, and past experiences.

The main neurotransmitter in primary afferents is the excitatory amino acid glutamate. Activation of nociceptors causes the release of glutamate from presynaptic terminals in the spinal cord dorsal horn; this release acts on the ionotropic glutamate receptor amino-3-hydroxy-5-methylisoxazole-4-propionic acid postsynaptically to cause rapid depolarization of dorsal horn neurons and, if threshold is reached, action potential discharge.

**PAIN CLASSIFICATION**

Pain can be mechanistically divided into four classifications: nociceptive, inflammatory, functional, and neuropathic. Nociceptive pain is transient pain in response to a noxious stimulus that activates high threshold afferents. Nociceptive pain serves a protective function. Inflammatory pain is the spontaneous hypersensitivity to pain that occurs in response to tissue damage and inflammation (e.g., postoperative pain, trauma, arthritis). Functional pain is hypersensitivity to pain resulting from abnormal central processing of normal input (e.g., pathologic irritable bowel syndrome,
fibromyalgia). Neuropathic pain is spontaneous pain and hypersensitivity to pain that occurs in association with damage to or lesions of the nervous system (e.g., peripheral neuropathy, postherpetic neuralgia). Nociceptive pain and inflammatory pain are prevalent in rheumatic disease. Functional pain and neuropathic pain are probably less prevalent in rheumatic disease; however, both should always be considered because poorly controlled pain can lead to nervous system dysfunction and functional and neuropathic pain.\(^\text{11}\)

**PHARMACOLOGIC TREATMENT OF CHRONIC PAIN**

The efficacy of analgesics is dependent on the pain mechanism. Primary analgesics are more efficacious in nociceptive and inflammatory pain, whereas adjuvant agents are more efficacious in neuropathic and functional pain. Each pain classification involves different pain mechanisms, and within each classification are multiple different pain mechanisms.\(^\text{12,13}\)

Pain is mediated through both peripheral and central mechanisms, and often more than one mechanism of pain is active in a given patient. Thus, the use of two or more agents with differing mechanisms increases the likelihood of interrupting pain signals and relieving pain.

Pharmacologic agents for the management of chronic pain are divided into primary analgesics, which have intrinsic analgesic properties, and adjunct analgesics, which may have primary analgesic properties in neuropathic pain but usually enhance the analgesic effects of primary analgesics when used in non-neuropathic pain syndromes. Table 66-1 summarizes these agents. Nonsteroidal anti-inflammatory drugs (NSAIDs)/cyclooxygenase (COX)-2 inhibitors will be discussed elsewhere (see Chapter 59).

**OPIOIDS**

Numerous studies have demonstrated the efficacy of the opioids in a variety of chronic pain states, including neuropathic and non-neuropathic. However, studies to document long-term efficacy have not been conducted. Long-term opioid therapy to treat chronic pain remains controversial. Over the last century, the pendulum has swung back and forth with regard to the use of the opioids to treat pain. In the mid 20th century, opioids were limited because of fears of addiction and diversion. In the late 20th century, the pendulum went to the other side with liberal use of opioids to treat chronic pain. With the turn of the 21st century, the pendulum is moving back toward the middle with recognition of the importance of opioids in chronic pain management but an understanding of the need to balance these benefits with risks. Recently published guidelines acknowledge that opioid analgesics have an important role in pain management, and that underuse of these agents may contribute to suboptimal pain management.\(^\text{14,15}\) However, these guidelines also acknowledge that abuse of prescription opioids has become an epidemic, with dramatic increases seen in the United States. Therefore, a set of Universal Precautions have been developed as a guide to help the physician who prescribes opioids\(^\text{15}\) (Table 66-2).

After conservative treatments, including nonopioids, have failed to control the patient’s pain, an opioid should be considered. In general, a short-acting weak opioid such as hydrocodone or codeine should be used first. If the patient is requiring more than three or four short-acting weak opioids per day, consider converting to a long-acting opioid. Controlled-release or long-acting opioids not only provide convenience to patients by reducing the number of daily doses required, they also provide a pharmacokinetic profile that results in reduced serum level peaks and troughs, and thereby an improvement in the consistency of effective analgesia and a potential reduction in opioid-related side effects that are often correlated with high peak serum levels. When converting to a long-acting opioid, access to a short-acting opioid for breakthrough pain can be continued, but monitored and controlled. Evaluation by a pain specialist may be considered when morphine equianalgesic dosages exceed 90 mg/day. The benefits of levels higher than 180 mg/day have not been established, and recent evidence suggests a significant increase in morbidity and mortality when doses exceed 100 mg/day.\(^\text{16}\)

**Opiate Receptor Classes**

Early work by Martin and associates led to the postulation of three opiate receptors: mu, kappa, and sigma.\(^\text{17,18}\) Later studies by Kosterlitz and colleagues led to the identification of the delta opioid receptor.\(^\text{19}\) With the exception of the sigma receptor, all of the opiate receptors are responsible for opioid-induced analgesia. Opioid receptors are mediated through a G protein, leading to a cascade of events that typically inhibit neuron activation.\(^\text{20}\) Persistent activation of G protein–coupled receptors typically results in progressive loss of effect, known as tolerance (discussed later).

---

**Table 66-1** Analgesic Options to Treat Chronic Pain

<table>
<thead>
<tr>
<th>Primary Analgesics</th>
<th>Adjunct Analgesics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetaminophen</td>
<td>Tricyclic antidepressants</td>
</tr>
<tr>
<td>Nonsteroidal anti-inflammatory drugs/cyclooxygenase-2 inhibitors</td>
<td>Serotonin-norepinephrine reuptake inhibitors</td>
</tr>
<tr>
<td>Opioids</td>
<td>Anticonvulsants</td>
</tr>
<tr>
<td></td>
<td>Muscle relaxants</td>
</tr>
<tr>
<td></td>
<td>Topical agents</td>
</tr>
</tbody>
</table>
Opiate Receptor Distribution and Mechanisms of Opioid-Induced Analgesia

The main sites of action of the opiates are believed to be located in the brain and spinal cord; however, under some circumstances, peripheral mechanisms are involved. Responses of an individual patient may vary dramatically with different mu-opioid receptor (MOR) agonists. If problems are encountered with one drug, another should be tried. Mechanisms underlying variations in individual responses to morphine-like agonists are poorly understood; however, they are thought to be due to MOR polymorphisms.

Numerous sites in the brain have MORs. Activation of receptors located in the mesencephalic periaqueductal gray (PAG) appears to be the most important cause of opioid-induced analgesia. MOR agonists block release of the inhibitory transmitter γ-aminobutyric acid (GABA) from tonically active PAG systems that regulate activity in projections from the PAG to the medulla. This results in an increase in PAG outflow to the medulla, leading to activation of medullospinal projections and release of noradrenaline or serotonin at the level of the spinal dorsal horn. This release can attenuate dorsal horn excitability and analgesia. PAG activation can also increase the excitability of the dorsal raphe and the locus coeruleus, from which ascending serotonergic and noradrenergic projections originate to project to the limbic forebrain, leading to the euphoric effects sometimes experienced with systemic MOR agonists.21

In the spinal cord, MOR agonists are limited for the most part to the substantia gelatinosa of the superficial dorsal horn, the region in which small, high-threshold sensory afferents terminate. Most of these opiate receptors are located presynaptically and postsynaptically on small peptidergic primary afferent C fibers. Presynaptic activation of the MOR prevents the opening of voltage-sensitive Ca2+ channels, thereby preventing transmitter release. Postsynaptic activation of the MOR increases potassium conductance, resulting in hyperpolarization and reduced excitation induced by the presynaptic release of glutamate.22 The ability of spinal opiates to reduce the release of excitatory neurotransmitters from C fibers presynaptically and to decrease the excitability of dorsal horn neurons postsynaptically is believed to account for powerful and selective effects upon spinal nociceptive processing. In humans, an extensive literature indicates that a variety of opiates delivered spinaly (intrathecally or epidurally) can induce a powerful analgesia.22

Systemic delivery of the opioids will reduce nociceptive pain through a central mechanism located in the brain and spinal cord, as described previously, whereas the peripheral application has no effect. However, under conditions of inflammation, which result in an exaggerated pain response (hyperalgesia), the peripheral application of the opioids will reduce the hyperalgesia. This action is believed to be mediated by opiate receptors on the peripheral terminals of small primary afferents that become active under inflammatory conditions. Whether the effects are uniquely on the afferent terminal or on inflammatory cells that release products that sensitize the nerve terminal is not known.23

Tolerance

Over time, a given dose of an opioid shows less effect and an increased dose is required to produce the same physiologic response. Tolerance to different effects of opioids occurs at different rates. For example, tolerance to sedation and nausea occurs earlier than analgesic tolerance. Some effects, such as constipation, never show tolerance.

The mechanism of opioid tolerance is controversial. Apparent opioid tolerance may be an indication of disease progression with a resultant increase in pain intensity; this is the first event that should be ruled out before true tolerance is assumed. Many cellular mechanisms can lead to tolerance. First, long-term opioid exposure can lead to receptor internalization, dephosphorylation, and desensitization. Second, exposure to high doses of opioids can lead to an increase in intracellular cyclic adenosine monophosphate (cAMP), activation of bulbospinal pathways, and glutamate receptor phosphorylation, producing an excitatory state (opioid-induced hyperalgesia).24,25 An incomplete cross-tolerance occurs between the various opioids, and when tolerance develops to one opioid, switching to another opioid can result in an increased effect.

Physical Dependence

Physical dependence is not addiction (and the terms should not be used interchangeably). Physical dependence is a pharmacologic effect characteristic of a number of different types of medications. Physical dependence is defined as the occurrence of an abstinence syndrome (withdrawal reaction) following abrupt discontinuation of the drug, substantial dose reduction, or administration of an antagonist. Physical dependence is generally assumed to occur with regular opioid use for as brief a period as a few days.

Opioid withdrawal can be manifested by significant somatomotor and autonomic outflow (reflected by agitation, hyperalgesia, hyperthermia, hypertension, diarrhea, pupillary dilation, and release of virtually all pituitary and adrenal medullary hormones) and by affective symptoms (dysphoria, anxiety, and depression).26 Opioid withdrawal can be minimized by slowly tapering the opioid. These phenomena are considered to be highly aversive and motivate the drug recipient to make robust efforts to avoid the withdrawal state. Initially, drug addicts are driven to repeated doses of opioids owing to the euphoric effects. However, over time, the euphoric effects are lessened, and addicts are driven to continue use to avoid withdrawal.

Addiction

Identification of the disease of addiction is important for safe and effective clinical management of pain in individuals with addictive disorders. The disease of addiction affects approximately 10% of the general population, and its prevalence may be higher in subpopulations of patients with pain. Active addiction is a contraindication to the use of the opioids; a past history is a relative contraindication, but opioids can be used successfully with appropriate monitoring. A persistent misunderstanding is prevalent among health care providers, regulators, and the general population regarding the nature and manifestations of addiction,
Opioid Pharmacology

Morphine

Morphine is the gold standard against which all other opioids are measured. Morphine can be administered by oral, rectal, subcutaneous, intravenous, intramuscular, and intraspinal routes. Oral bioavailability of morphine is about 25% (range, 10% to 40%). Peak plasma concentrations occur 0.5 to 1 hour after ingestion with a half-life of around 3 to 4 hours. To increase the dosing interval of oral morphine, several sustained- and controlled-release preparations have been developed (Table 66-4). The average protein binding of morphine is around 35%, but this decreases with renal and hepatic dysfunction. Because of the hydrophilicity of morphine, there is poor central nervous system (CNS) penetration and little tissue accumulation with repeated dosing. Morphine is metabolized primarily in the liver to two main metabolites: morphine-3-glucuronide (M3G) and morphine-6-glucuronide (M6G). M3G is the primary metabolite with potent CNS excitatory properties. Because of its polarity, CNS penetration is poor; however, in renal failure, M3G plasma concentrations can be high enough to drive this metabolite into the CNS, leading to excitation. M6G possesses potent MOR agonism, but because of its high polarity, CNS penetration is poor. However, like M3G, M6G can accumulate in renal impairment, leading to exaggerated opioid effects. Only about 10% of unmetabolized morphine is excreted renally, whereas 90% is excreted renally as morphine glucuronide (70% to 80%) and normorphine (5% to 10%) conjugates.21

Methadone

Methadone is a long-acting MOR agonist with potency similar to that of morphine but with two important differences: (1) It has a long half-life, and (2) it has high oral bioavailability. Routes of administration are oral and intravenous, with oral delivery being by far the most common. Intravenous routes have been used as a single loading dose for postoperative pain control. Because of its low hepatic which may result in undertreatment of pain and stigmatization of patients using opioids for pain control.

Evaluating for addiction in a patient who is prescribed long-term opioids for pain control is often problematic. This risk of opioid addiction and misuse can be assessed with validated questionnaires (Table 66-3).27 Although the concept of addiction may include the symptoms of physical dependence and tolerance, physical dependence or tolerance alone does not equate with addiction. In the chronic pain patient taking long-term opioids, physical dependence and tolerance should be expected, but the maladaptive behavior changes associated with addiction are not expected.28,29

Addiction is a behavioral pattern characterized by compulsive use of a drug and overwhelming involvement with its procurement and use in spite of potential harm. For patients with continuous pain, inadequate pain management (e.g., PRN dosing schedule, use of drugs with inadequate potency, use of dosing intervals that are too long) can lead to behavioral symptoms that mimic those seen with psychological dependence and can be mistaken for addiction termed pseudoaddiction. In the case of pseudoaddiction, problem behaviors resolve after sufficient pain relief is established. However, behaviors related to true addiction may also resolve after dose escalation. Thus, it can be difficult to distinguish pseudoaddiction from true addiction; careful evaluation and management may be required until the circumstances are sorted out28 (see Table 66-3).

Table 66-3  Opioid Risk Tool

<table>
<thead>
<tr>
<th>Personal History</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol abuse</td>
<td>(3)</td>
<td>(1)</td>
</tr>
<tr>
<td>Illegal drug use</td>
<td>(3)</td>
<td>(2)</td>
</tr>
<tr>
<td>Prescription drug abuse</td>
<td>(4)</td>
<td>(4)</td>
</tr>
<tr>
<td>Mental Health</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis of ADD, OCD, bipolar, schizophrenia</td>
<td>(2)</td>
<td>(2)</td>
</tr>
<tr>
<td>Diagnosis of depression</td>
<td>(1)</td>
<td>(1)</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 16-45 years</td>
<td>(1)</td>
<td>(1)</td>
</tr>
<tr>
<td>History of preadolescent sexual abuse</td>
<td>(0)</td>
<td>(3)</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Numbers in parentheses indicate points. Scoring:
0-3: low risk: 6% chance of developing problematic behaviors.
4-7: moderate risk: 28% chance of developing problematic behaviors.
8: high risk: >90% chance of developing problematic behaviors.
ADD, attention deficit disorder; OCD, obsessive-compulsive disorder.


clearance and therefore low first-pass effect, oral bioavailability is around 80%. Methadone has higher lipid solubility than morphine, which results in faster CNS penetration and onset of action; however, its lipid solubility leads to a higher volume of distribution and a shorter duration of action with initial dosing. Low hepatic clearance leads to accumulation, a longer duration of action, and possible overdose with repeated dosing.

Because of the long half-life of methadone (average, 15 to 30 hours, with published reports ranging from 8 to 59 hours) with drug accumulation over several days, methadone needs to be administered with caution with long intervals between dose adjustments (5 to 7 days). Deaths have been reported with the use of methadone for chronic pain, and in November 2006, the U.S. Food and Drug Administration (FDA) issued a public health advisory for methadone with a black box warning. Although details are often unclear, many of these deaths may be due to conversion of patients from other opioids to methadone. Methadone has been associated with QRS prolongation, which can lead to sudden cardiac death. If unfamiliar with the use of methadone, it is advisable to seek consultation with a pain specialist, especially when considering prescribing methadone at greater than a low dose (20 to 30 mg/day).

The L-isomer of methadone possesses opioid activity, whereas the D-isomer is weak or inactive as an opioid. Evidence suggests that D-methadone is antinociceptive as a result of its N-methyl-D-aspartate (NMDA) receptor antagonist activity, making methadone attractive in treating neuropathic pain.

Methadone is metabolized primarily by CYP3A4, secondarily by CYP2D6, and to a smaller extent by CYP1A2 and additional enzymes that are under study. CYP3A4, the most abundant metabolic enzyme in the body, can vary 30-fold between individuals in terms of its presence and activity in the liver. In addition, drugs that inhibit this enzyme (several antiretrovirals, clarithromycin, itraconazole, ketoconazole, nefazadone, telithromycin, aprepitant, diltiazem, erythromycin, fluconazole, grapefruit juice, verapamil, cimetidine, and others) will increase the effect of methadone, possibly leading to overdose. This enzyme is also found in the gastrointestinal tract, so methadone metabolism actually starts before the drug enters the circulatory system. The amount of this enzyme in the intestine can vary up to 11-fold, partially accounting for the variable breakdown of methadone.

Because the half-life of methadone is so long, it has been used extensively to treat opioid dependence and withdrawal. The use of methadone to treat addiction requires a separate clinic and physician license. However, methadone can be used to treat pain under the routine medical Drug Enforcement Administration (DEA) license.

**Fentanyl**

Fentanyl is a potent opioid with very high lipid solubility. This high lipid solubility leads to fast onset, a short duration of action, high protein binding (80%), and a high volume of distribution. The high volume of distribution accounts for the short duration of action owing to the high concentration gradient from the plasma to fat and muscle. However, with repeated dosing, the duration of action increases as fat and muscle stores become saturated with fentanyl. Although the analgesic half-life is around 1 to 2 hours, the terminal half-life is about 3 to 4 hours.

Routes of delivery of fentanyl include transdermal, transmucosal, intravenous, and intraspinal. Subcutaneous delivery is also used in terminal cancer patients. High potency and lipid solubility make fentanyl ideal for transdermal and transmucosal delivery. Two transdermal fentanyl patches are available on the market: reservoir and matrix. The reservoir patch can be accessed and the fentanyl extracted, whereas the matrix patch is tamper-proof. Each system delivers fentanyl over 72 hours; however, some patients will deplete the patch within 48 hours, requiring more frequent patch changes. This variability can result from skin perspiration, fat stores, skin temperature, and muscle bulk. After application, the skin serves as a depot, and systemic levels rise for the next 12 to 24 hours, then remaining stable until 72 hours. Peak concentration occurs somewhere between 27 and 36 hours (25 mg 27 hours, 100 μg 36 hours). A steady state is reached after several applications. After removal of the system, plasma concentrations fall by 50% over 17 hours.

Three transmucosal products are currently available. The fentanyl orale contains fentanyl in a base of food starch, confectioners’ sugar (2 g vs. 30 g in a Snickers bar), edible glue, citric acid, and artificial berry flavor. The orale is placed between the gum and the cheek and is allowed to dissolve over 15 minutes with bioavailability of about 50%. Onset is fast, with a peak effect at about 35 minutes. The fentanyl buccal tablets use an oravescent delivery system that generates a reaction that releases carbon dioxide when the tablet comes in contact with saliva. Transient pH changes accompanying the reaction optimize the dissolution (at a lower pH) of fentanyl through the buccal mucosa. Onset is fast with a peak effect at about 25 minutes. The Bio-Erodable Muco Adhesive (BEMA) fentanyl delivery system is composed of water-soluble polymeric films. This system consists of a bioadhesive layer bonded onto an inactive layer. The active ingredient, fentanyl citrate, is incorporated into the bioadhesive layer, which adheres to the moist buccal mucosa. The amount of fentanyl delivered transmucosally is proportional to the film surface area. It is believed that the inactive layer isolates the bioadhesive layer from the saliva, which may optimize delivery of fentanyl across the buccal mucosa, resulting in higher bioavailability (71%). Onset is fast with a slightly longer peak effect (60 minutes) as compared with the other transmucosal systems; however, there may be a longer duration. Chewing and swallowing any of the transmucosal fentanyl products results in lower bioavailability and peak effect because swallowed fentanyl is poorly absorbed from the gastrointestinal tract.

Fentanyl is metabolized primarily by CYP3A4 to the inactive metabolite norfentanyl; therefore, drugs that inhibit this enzyme will increase the drug effect (see earlier under “Methadone”). Only opioid-tolerant patients (>60 mg/day of oral morphine equivalent) should be started on fentanyl products owing to risks of severe respiratory depression.
**Oxycodone and Oxymorphone**

Oxycodone is administered by the oral or rectal route. Oral bioavailability is 60% and protein binding, 45%. Onset and duration of action are similar to morphine. Oxycodone is metabolized by the CYP3A4 enzyme primarily to noroxycodone, which has about 25% potency of the parent compound but also has neuroexcitatory effects. Minor metabolites include oxymorphone, which is more potent than oxycodone and has a longer half-life, and noroxymorphone, which has no analgesic properties.

Oxycodone is available as an immediate-release tablet or solution and as a controlled-release preparation. The controlled-release preparation has an immediate release effect of up to 40% of the contents, followed by a sustained 12-hour release component. The immediate release effect results in rapid onset followed by a sustained effect. This can be a disadvantage because the patient may perceive the duration as short when coming down from the high peak plasma concentration caused by the immediate release.

Oxymorphone is a metabolite of oxycodone and can be delivered via the oral and rectal routes. Both bioavailability and protein binding are low (10%), but terminal half-life is long (10 to 12 hours). Most of the oxymorphone is metabolized to oxymorphene-3-glucuronide, which is an inactive metabolite. Oxymorphone is available in immediate-release and sustained-release preparations.

**Hydromorphone**

Hydromorphone is administered by the oral, intravenous, intramuscular, and subcutaneous routes. Oral bioavailability is low (ranging from 25% to 50%), and protein binding is less than 20%. The kinetics are very similar to morphine. Hydromorphone has an extensive first-pass effect, and 95% is metabolized to the inactive hydromorphone-3-glucuronide. A 24-hour-release preparation of hydromorphone has been approved by the FDA. This preparation uses an osmotic piston-driven system in pill form that slowly delivers hydromorphone over 24 hours.

**Meperidine**

Meperidine can be delivered by the oral, rectal, intravenous, intramuscular, subcutaneous, and intraspinal routes. Meperidine use has decreased over time owing to side effects and risks associated with the parent compound and metabolites (see later). Oral bioavailability is about 50%, protein binding is about 60%, and peak concentrations in plasma usually are observed in 1 to 2 hours. Onset and duration are similar to morphine. Large doses of meperidine repeated at short intervals may produce an excitatory syndrome that includes hallucinations, tremors, muscle twitches, dilated pupils, hyperactive reflexes, and convulsions. These excitatory symptoms are caused by accumulation of the metabolite, normeperidine, which has a half-life of 15 to 20 hours compared with 3 hours for meperidine. Because normeperidine is eliminated by the kidney and the liver, decreased renal or hepatic function increases the likelihood of such toxicity. As a result of these properties, meperidine is not recommended for the treatment of chronic pain because of concerns over metabolite toxicity. It should not be used for longer than 48 hours or in doses greater than 600 mg/day.

Severe reactions may follow the administration of meperidine to patients being treated with monoamine oxidase (MAO) inhibitors. Two basic types of interactions can be observed. The most prominent is an excitatory reaction ("serotonin syndrome"). This reaction may be due to the ability of meperidine to block neuronal reuptake of serotonin and resulting serotonergic overactivity. Another type of interaction, a potentiation of opioid effect due to inhibition of hepatic CYPs, also can be observed in patients taking MAO inhibitors, necessitating a reduction in the doses of opioids.

**Hydrocodone**

Hydrocodone is a weak MOR agonist. It is delivered only by the oral route and is available only in acetaminophen-containing compounds, thus making it a Schedule III opioid. This has led to overutilization with reports of acetaminophen toxicity. Hydrocodone has an oral bioavailability of about 25% with low protein binding. It is metabolized to hydromorphone, but its analgesic activity is not dependent upon metabolism to hydromorphone. However, rapid metabolizers may experience a faster onset of action owing to the production of hydromorphone. Onset and duration of effect are similar to morphine.

**Codeine**

Codeine has an exceptionally low affinity for the MOR, and the analgesic effect of codeine is due to its conversion to morphine. The oral bioavailability is about 60% and is poorly protein bound. The onset and duration of effect are similar to morphine. It is administered via the oral route.

The conversion of codeine to morphine is effected by CYP2D6. Well-characterized genetic polymorphisms in CYP2D6 lead to the inability to convert codeine to morphine, thus making codeine ineffective as an analgesic for about 10% of the white population. Other polymorphisms can lead to enhanced metabolism and thus to increased sensitivity to the effects of codeine. Variation in metabolic efficiency is evident among ethnic groups. For example, Chinese individuals produce less morphine from codeine than do whites and are less sensitive to the effects of morphine.

**Tramadol**

Tramadol is a synthetic codeine analogue with a dual mechanism of action. Analgesia results through weak MOR agonism and inhibition of uptake of norepinephrine and serotonin.

Tramadol is 68% bioavailable after a single oral dose with about 20% protein binding. Its affinity for the opioid receptor is only 1/1000 that of morphine. However, the primary O-demethylated metabolite of tramadol is two to four times as potent as the parent drug and may account for part of the analgesic effect. Tramadol is supplied as a racemic mixture, which is more effective than either enantiomer alone. The
(+)-enantiomer binds to the receptor and inhibits serotonin uptake. The (−)-enantiomer binds norepinephrine uptake and stimulates α2-adrenergic receptors.\textsuperscript{46} The compound undergoes hepatic metabolism and renal excretion, with an elimination half-life of 6 hours for tramadol and 7.5 hours for its active metabolite. Analgesia begins within an hour of oral dosing and peaks within 2 to 3 hours. The duration of analgesia is about 6 hours. The maximum recommended daily dose is 400 mg. Tramadol is also available in an extended 24-hour release preparation.

Physical dependence on and abuse of tramadol have been reported. Although its abuse potential is unclear, tramadol probably should be avoided in patients with a history of addiction. Because of its inhibitory effect on serotonin uptake, tramadol should not be used in patients taking MAO inhibitors and triptans. Seizures have been reported with concomitant use of selective serotonin receptor inhibitors (SSRIs), serotonin-norepinephrine reuptake inhibitors (SNRIs), tricyclic antidepressants (TCAs), and neuroleptics.\textsuperscript{46}

**Tapentadol**

Tapentadol is a strong opioid with a dual mechanism of action, similar to tramadol. It is a strong MOR agonist and a serotonin and norepinephrine reuptake inhibitor. Because of its first-pass metabolism, its bioavailability is low (32%) and protein binding is low (20%). The metabolites of tapentadol are inactive. The half-life is approximately 4 hours.

Similar to tramadol, tapentadol is contraindicated in patients taking MAO inhibitors, triptans, SSRIs, SNRIs, TCAs, and neuroleptics, owing to risk of serotonin syndrome. There does not appear to be a risk of seizures, as is seen with tramadol. Tapentadol is currently available as an oral immediate-release drug; however, an extended-release preparation is in development (Figure 66-1).

**Toxicity**

**Respiration**

Although effects on respiration are readily demonstrated, clinically significant respiratory depression rarely occurs with standard analgesic doses in the absence of other contributing comorbidities or concomitant use of sedatives. In addition, the respiratory depressant effect of the opioid is significantly reduced with continued opioid use owing to tolerance. It should be stressed, however, that respiratory depression represents the primary cause of morbidity secondary to opiate therapy.\textsuperscript{51} In humans, death from opiate poisoning is nearly always due to respiratory arrest or obstruction.\textsuperscript{48} For example, in November of 2006, the FDA notified health care professionals of reports of death and life-threatening adverse events, such as respiratory depression and cardiac arrhythmias, in patients receiving methadone (FDA ALERT [11/2006]: Death, narcotic overdose, and serious cardiac arrhythmias, www.fda.gov/cder/drug/infopage/methadone). Methadone appears to be involved in approximately one-third of all prescription opioid-related deaths, exceeding hydrocodone and oxycodone despite being prescribed one-tenth as often. This has led to revisions in methadone conversion tables.

At therapeutic doses, opiates depress all phases of respiration (rate, minute volume, and tidal exchange). High doses can produce irregular and agonal breathing.\textsuperscript{48} A number of factors can increase the risk of opioid-induced respiratory depression, even at therapeutic doses. These include (1) concomitant use of sedatives such as alcohol, benzodiazepines, and tranquillizers, (2) obstructive and central sleep apnea, (3) extremes in age (both newborns and elderly), (4) comorbidities such as pulmonary disease and renal disease, and (5) removal of the painful stimulus. Pain serves as a respiratory stimulant, and removal of pain (e.g., neurolytic block with severe cancer pain) will reduce the ventilatory drive, leading to respiratory depression.

**Sedation**

Opiates can produce drowsiness and cognitive impairment, which can increase respiratory depression. These effects most typically are noted following initiation of opiate therapy or after a dose increase but usually resolve with continued opioid use. If the sedation does not resolve, other causes of the sedation should be investigated, such as concomitant use of other sedatives or the presence of sleep apnea. In the absence of these variables, opioid rotation may result in less sedation.\textsuperscript{49}

**Neuroendocrine Effect**

Regulation of the release of hormones and factors from the pituitary is under complex regulation by opiate receptors in the hypothalamic-pituitary-adrenal (HPA) axis. The opioids block the release of a large number of many HPA hormones, including sex hormones, prolactin, oxytocin, growth hormone, and antidiuretic hormone.

In males, the opioids inhibit adrenal function, resulting in reduced cortisol production and reduced adrenal androgens. In females, the opioids lower luteinizing hormone (LH) and follicle-stimulating hormone (FSH) release. In both males and females, long-term opioid therapy can result in endocrinopathies, including hypogonadotropic hypogonadism, leading to decreased libido, and females may develop menstrual cycle irregularities. These changes are reversible with removal of the opiate. Mechanisms of neuroendocrine effects of the opioids are thought to be due to a direct effect on the pituitary and an indirect action on the hypothalamus, blocking the release of gonadotropin-releasing hormone (GnRH) and corticotropin-releasing hormone (CRH). Secretion of thyrotropin is relatively unaffected.

Opioids inhibit the release of dopamine from neurons of the tuberoinfundibulum of the arcuate nucleus. Prolactin release from lactotrope cells in the anterior pituitary is under inhibitory control by dopamine; therefore the reduced dopamine caused by the opioids results in an increase in plasma prolactin. Prolactin counteracts the effects of dopamine, which is responsible for sexual arousal, and high levels of prolactin can result in impotence and loss of libido. Long-term opioid use can increase growth hormone by inhibiting somatostatin release; this regulates GH-releasing hormone secretion.\textsuperscript{50}

Antidiuretic hormone (ADH) and oxytocin are synthesized in the perikarya of the magnocellular neurons in the
Figure 66-1  Opioid structures.
paraventricular and supraoptic nuclei of the hypothalamus and are released from the posterior pituitary. Kappa opioid receptor agonists inhibit the release of oxytocin and antidiuretic hormone (and cause prominent diuresis). MOR agonists have minimal effect or tend to produce antidiuretic effects in humans, and reduce oxytocin secretion. Some of the opioids (i.e., morphine) stimulate histamine release, resulting in hypotension and secondary ADH release. The effects of the opioids on vasopressin and oxytocin release may reflect both a direct effect upon terminal secretion and indirect effects upon dopaminergic and noradrenergic modulatory projections into the paraventricular and supraoptic hypothalamus.

Miosis

Lumination of the pupil activates a reflex arc, which, through local circuitry in the Edinger-Westphal nucleus, activates parasympathetic outflow through the ciliary ganglion to the pupil, producing constriction. The parasympathetic outflow is locally regulated by GABAergic interneurons, and the opiates are believed to block GABAergic interneuron–mediated inhibition, resulting in miosis. At high doses of agonists, the miosis is marked, and pinpoint pupils suggest opioid intoxication. Although some tolerance to the miotic effect develops with therapeutic doses of the opioids, high circulating concentrations of opioids continue to result in constricted pupils.

Myoclonus and Seizure

Myoclonus and seizures have been reported in patients receiving high doses of opiates. In addition, seizure-like activity can occur with extremely high doses of the opioids. Seizures may also occur at lower doses of meperidine owing to the normeperidine metabolite, which lowers the seizure threshold. Myoclonus and seizure activity is thought to be due to inhibition of GABA from inhibitory interneurons in the hippocampal pyramidal cells and dorsal horn cells. The opiates also have a direct stimulatory effect through interaction with G inhibitory and stimulatory coupled receptors. It has been suggested that in addition to the normeperidine metabolite, which is seizurogenic, morphine-3-glucuronide (from morphine) may induce myoclonus and seizures at high doses.

Nausea and Vomiting

The gastrointestinal tract is under the control of the vomiting center located in the brain stem. The vomiting center is activated by the chemoreceptor trigger zone (CRTZ), the vestibular system, and the gastrointestinal tract. The opioids activate the CRTZ, sensitize the vestibular system, and reduce gastric emptying time, all of which will lead to activation of the vomiting center. Drugs that reduce CRTZ activity (antidopaminergics, 5-hydroxytryptamine, (5-HT) serotonin receptor antagonists) and vestibular sensitization (anticholinergics, antihistamines) will reduce opioid-induced nausea. Drugs that enhance gastrointestinal motility are also effective (e.g., metoclopramide).

Constipation

It is estimated that 40% to 95% of patients treated with opioids show constipation, and that changes in bowel function can be demonstrated even with acute dosing. Opioid receptors are densely distributed in enteric neurons between the myenteric and submucosal plexuses and on a variety of secretory cells. Stimulation of MOR in the intestines reduces propulsatile and diminishes intestinal secretions. The prolonged transit time of the intestinal contents, along with reduced intestinal secretion, leads to increased water absorption, increasing the viscosity of bowel contents and constipation. In addition, anal sphincter tone is increased, and reflex relaxation in response to rectal distension is reduced. All of these effects combine to contribute to morphine-induced constipation. Tolerance to opioid-induced constipation does not occur.

Biliary Spasm

Relaxation of the sphincter of Oddi is suppressed by the opioids, which can result in an increase in the common bile duct pressure, leading to symptoms of biliary colic. Therefore, treatment of the pain of biliary colic with the opioids can lead to an exacerbation of the pain rather than relief.

Urinary Retention

MOR agonists inhibit the urinary voiding reflex, increase the tone of the external sphincter, and induce bladder relaxation, which can result in urinary retention. This effect is mediated by MOR and delta-opioid receptor activation in the brain and spinal cord. This will result in higher bladder volumes and sometimes requires catheterization.

Pruritus

Pruritus can occur with all of the MOR agonists. The mechanism is thought to be due to disinhibition of itch-specific neurons, which have been identified in the spinal dorsal horn. It can occur with both systemic and spinally administered doses but is more common with spinally administered doses. It tends to be focused in the trunk and face.

Immunosuppression

The effect of the opioids on immune function is controversial. The opioids suppress immune function through a direct effect on immune system cells and indirectly by activating sympathetic outflow and modulation of the hypothalamic-pituitary-adrenal axis. A proposed mechanism for the immune suppressive effects of morphine on neutrophils is through nitric oxide–dependent inhibition of nuclear factor κB (NFκB) activation. Others have proposed that the induction and activation of mitogen-activated protein (MAP) kinase may play a role. Pain itself is immunosuppressive; therefore, a reduction in pain will counteract and probably outweigh the direct immunosuppressive effects of the opioids. In addition, it appears that acute delivery of the opioids is more immunosuppressive than long-term delivery, suggesting tolerance to this effect.
**Sweating**

The opioids exert a wide range of effects on thermoregulation with high doses leading to hyperthermia and low doses leading to hypothermia.\(^7\) Excessive sweating has been reported to occur in as many as 45% of patients taking methadone. The mechanism appears to be related to release of histamine.\(^7\)\(^5\)\(^7\)

**ANTIDEPRESSANTS**

Antidepressants have long been used for the treatment of chronic pain. In the past, they were more often chosen to improve mood rather than to treat pain because of coexisting anxiety and depression. After separate studies showed improved pain control in patients without depression, as well as in patients with depression without improvement in mood, it was realized that these medications have independent analgesic actions.\(^7\)\(^8\)\(^9\) Further, improvement in pain is seen earlier and at lower doses than improvements in depression, also demonstrating this analgesic efficacy.\(^8\) A recent meta-analysis of 18 randomized, placebo-controlled studies with multiple antidepressants concluded that there was strong evidence for efficacy of antidepressants for pain relief, fatigue, and sleep disturbance, and in improving health-related quality of life.\(^9\)\(^1\) The FDA in the United States has now approved selective SNRIs for fibromyalgia, diabetic peripheral neuropathy, and chronic musculoskeletal pain. Nevertheless, not all antidepressants are effective pain medications. For example, the SSRIs have minimal analgesic properties and therefore have a limited role in the treatment of chronic pain.\(^8\) Trazadone, used mainly for sleep, is another antidepressant without analgesic actions.

The main site of analgesic action is thought to be the reuptake inhibition of norepinephrine and serotonin at the level of the spinal cord and higher. This inhibition increases the extracellular concentration of these two monoamines, resulting in activation of descending inhibitory pain pathways, and ultimately decreases pain.\(^8\)\(^3\) A peripheral action has also been proposed because topical antidepressants were shown to produce analgesia in animal models of pain.\(^8\)\(^1\)\(^5\) Current investigation has shown that topical application can result in the inhibition of NMDA receptors and the blockade of sodium and calcium channels, which alone or in combination could explain their peripheral analgesic properties.\(^8\)\(^3\)\(^8\)\(^6\)\(^8\)\(^8\) Because TCAs are effective in blocking sodium channels, and the proliferation of sodium channels plays a key role in the pathogenesis of neuropathic pain, this may be one of the main sites of action for these drugs.\(^8\)\(^9\)\(^1\)\(^0\) Finally, TCAs have been shown to potentiate endogenous opioids.\(^9\) (Figure 66-2).

**Tricyclic Antidepressants**

TCAs are considered first-line agents in the treatment of chronic neuropathic pain and fibromyalgia.\(^9\)\(^2\)\(^9\) Two systematic reviews with 17 randomized controlled trials (RCTs) using 10 antidepressants have shown numbers needed to treat (NNTs) of approximately 2.5 for neuropathic pain (Table 66-5).\(^9\)\(^4\)\(^9\)\(^5\) All TCAs have shown fairly equal efficacy within the class.

TCAs are usually divided into tertiary amines (amitriptyline, imipramine, and doxepin) and secondary amines (nortriptyline and desipramine). Nortriptyline and desipramine are demethylated in the liver from amitriptyline and imipramine, respectively. The tertiary amines tend to block the reuptake of serotonin more than norepinephrine, and the secondary amines are more selective in their inhibition of norepinephrine uptake.

It is recommended that tricyclics should be started at the lowest dose possible and titrated slowly. A typical starting dose for amitriptyline, nortriptyline, or desipramine is as low as 5 to 10 mg before bedtime. Doses can be increased by the same amount as the starting dose approximately every 7 days (Table 66-6). Studies have shown improved analgesia with amitriptyline in the range of 25 to 50 mg, but some studies have gone as high as 200 mg. Response should be seen within 3 to 4 weeks (see Table 66-6).

**Table 66-5 Analgesic Efficacy of Antidepressants**

<table>
<thead>
<tr>
<th>Antidepressant</th>
<th>Number of Studies</th>
<th>Positive Results</th>
<th>Humans*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Number of Studies</td>
</tr>
<tr>
<td>TCAs</td>
<td>126</td>
<td>Acute pain tests 81%</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chronic pain models 95%</td>
<td></td>
</tr>
<tr>
<td>SSRIs*</td>
<td>39</td>
<td>Acute pain tests 44%</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chronic pain models 33%</td>
<td></td>
</tr>
<tr>
<td>SNRIs*</td>
<td>10</td>
<td>Acute pain tests 100%</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chronic pain models 100%</td>
<td></td>
</tr>
<tr>
<td>Others†</td>
<td>7</td>
<td>Acute pain tests 100%</td>
<td>1</td>
</tr>
</tbody>
</table>

---


*Number of patients treated to improve the health of one patient (at least 50% decrease in pain intensity).

*For example, fluoxetine, fluvoxamine, sertraline, paroxetine, and citalopram.

*For example, venlafaxine, milnicipran, and duloxetine.

*NTT, number needed to treat; SSRIs, selective serotonin reuptake inhibitors; TCAs, tricyclic antidepressants.

Side Effects

Unfortunately, side effects with all of the TCAs are common and can limit use, especially in the elderly and in patients with hepatic impairment. Side effects tend to be less with the secondary amines. Sedation is common owing to its antihistamine effects but can be beneficial if there are sleep complaints. Anticholinergic side effects, which include dry mouth, constipation, urinary retention, and blurred vision, are also prevalent. Both anticholinergic and antihistamine side effects are dose dependent and can decrease with time; slow titration can improve compliance. Weight gain and sexual dysfunction are frequently reported. Cardiac side effects include orthostatic hypotension and dysrhythmias. Cardiac history should be reviewed and a baseline electrocardiogram (ECG) should be considered for patients over the age of 40. Seizure thresholds are lowered. Risk of suicidal thinking and behavior is increased in children, adolescents, and young adults up to the age of 24. Further, if these medications are taken intentionally or accidentally in overdose, they are dangerous and can be extremely lethal at relatively low doses. When these medications are stopped, it is recommended that they be reduced by 25% per week to decrease side effects.

Serotonin-Norepinephrine Reuptake Inhibitors (SNRIs)

The SNRIs, venlafaxine, duloxetine, and milnacipran, have also shown analgesic properties. Together they have shown benefit in painful diabetic peripheral neuropathy, neuropathic pain, fibromyalgia, and, most recently, chronic musculoskeletal pain. Venlafaxine, the first drug in this class, is an inhibitor of serotonin reuptake at low doses, and at higher doses it inhibits norepinephrine reuptake. This means that it acts more like an SSRI at lower doses and

Figure 66-2  Chemical structures of antidepressants.
Table 66-6  Tricyclic Antidepressants: Pharmacodynamics

<table>
<thead>
<tr>
<th>Drug</th>
<th>Amitriptyline</th>
<th>Desipramine*</th>
<th>Doxepin</th>
<th>Imipramine</th>
<th>Nortriptyline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half-life</td>
<td>9-27 hr</td>
<td>7-60 hr</td>
<td>6-8 hr</td>
<td>6-18 hr</td>
<td>28-31 hr</td>
</tr>
<tr>
<td>Metabolism</td>
<td>Hepatic</td>
<td>Hepatic</td>
<td>Hepatic</td>
<td>Hepatic</td>
<td>Hepatic</td>
</tr>
<tr>
<td>Excretion</td>
<td>Urine</td>
<td>Urine</td>
<td>Urine</td>
<td>Urine</td>
<td>Urine</td>
</tr>
<tr>
<td>Protein binding</td>
<td>≥90%</td>
<td>90%-92%</td>
<td>80%-85%</td>
<td>60%-95%</td>
<td>93%-95%</td>
</tr>
<tr>
<td>Therapeutic dose/day</td>
<td>10-150 mg</td>
<td>10-150 mg PO</td>
<td>qhs or 2 times/day</td>
<td>qhs or 2 times/day</td>
<td>≥2 days/dose change</td>
</tr>
<tr>
<td>Administration</td>
<td>≥3 days/dose change</td>
<td>≥2 days/dose change</td>
<td>≥7 days/dose change</td>
<td>≥7 days/dose change</td>
<td>≥3 days/dose change</td>
</tr>
<tr>
<td>Titration</td>
<td>≥27 days/dose change</td>
<td>≥3 days/dose change</td>
<td>≥7 days/dose change</td>
<td>≥7 days/dose change</td>
<td>≥3 days/dose change</td>
</tr>
</tbody>
</table>

Metabolism/Transport Effects

<table>
<thead>
<tr>
<th>Substrate CYP1A2</th>
<th>Minor</th>
<th>Minor</th>
<th>Minor</th>
<th>Minor</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate CYP2C9</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
<td>Major</td>
<td>Minor</td>
</tr>
<tr>
<td>Substrate CYP2D6</td>
<td>Major</td>
<td>Major</td>
<td>Major</td>
<td>Major</td>
<td>Major</td>
</tr>
<tr>
<td>Substrate CYP3A4</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
</tr>
<tr>
<td>Substrate CYP2B6</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
</tr>
<tr>
<td>Inhibits CYP1A2</td>
<td>Weak</td>
<td>Weak</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
</tr>
<tr>
<td>Inhibits CYP2A6</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
</tr>
<tr>
<td>Inhibits CYP2B6</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
</tr>
<tr>
<td>Inhibits CYP2C9</td>
<td>Weak</td>
<td>Weak</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
</tr>
<tr>
<td>Inhibits CYP2C19</td>
<td>Weak</td>
<td>Weak</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
</tr>
<tr>
<td>Inhibits CYP2D6</td>
<td>Moderate</td>
<td>Weak</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
</tr>
<tr>
<td>Inhibits CYP2E1</td>
<td>Moderate</td>
<td>Weak</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
</tr>
<tr>
<td>Inhibits CYP3A4</td>
<td>Moderate</td>
<td>Weak</td>
<td>Minor</td>
<td>Minor</td>
<td>Minor</td>
</tr>
</tbody>
</table>


Side Effects

Side effects of all three of these SNRIs include nausea, dry mouth, and constipation. Sexual side effects tend to be less than with SSRIs. Duloxetine should not be used with coexisting hepatic insufficiency. Milnacipran is not recommended for patients with end-stage renal disease but is unique in that it is not metabolized by cytochrome P450 isoenzymes. Serotonin syndrome, caused by iatrogenic overstimulation of central and peripheral serotonin receptors, presents with neuromuscular hyperactivity, autonomic hyperactivity, and altered mental status. It may occur abruptly and progress rapidly when these drugs are used at high doses or combined with other medications that stimulate serotonin. These drugs can also impair platelet aggregation, particularly if used concomitantly with aspirin or NSAIDs. Dose-related increases in blood pressure have been reported and should be followed. SNRIs carry a black box warning in the United States for increased risk of suicidal thinking and behavior in children, adolescents, and young adults under the age of 25. When these medications are discontinued, withdrawal symptoms have been reported. It is recommended that the dose be decreased by 25% per week to minimize these symptoms (see Tables 66-3 and 66-5 through 66-9).

ANTICONVULSANTS

Neuropathic pain (NeP) is chronic pain initiated by nervous system lesions or dysfunction and maintained by a number of mechanisms. Excess stimulation of nociceptive pathways or damage to non-nociceptive pathways alters the balance between painful and nonpainful inputs so that pain results without nociceptor stimulation. Several cellular and molecular mechanisms operating over different periods of time are thought to be involved in the abnormal peripheral and central nervous system activity associated with NeP.

Many of these mechanisms can be modulated by the anticonvulsants. First, nerve injury is reported to evoke spontaneous discharges from the cell bodies of myelinated fibers at the levels of the dorsal root ganglia. The mechanism of spontaneous activity is thought to be secondary to an increase in concentration of sodium channels in neuromas, dorsal root ganglion cells, and areas of demyelination. Second, spinal inhibitory interneurons modulate the peripheral-to-central transmission of pain signals, thus “gating” ascending sensory information. GABA and glycine and their receptors are abundant in the superficial
dorsal horn,\textsuperscript{10,11} but their levels are regulated by primary afferent input and change significantly after nerve injury.\textsuperscript{111}

Finally, increased glutamatergic neurotransmission may also contribute to hyperexcitability and NeP through activation of both NMDA receptors and non-NMDA amino-3-hydroxy-5-methylisoxazole propionic acid (AMPA)/kainate-type glutamate receptors.\textsuperscript{112} Therefore, drugs that modulate sodium channels, increase GABA, reduce glutamate release, or block glutamate effects can potentially reduce neuropathic pain.

The four main mechanisms of action of the anticonvulsants that result in pain reduction include (1) calcium channel modulation, (2) sodium channel blockade, (3) NMDA antagonism, and (4) GABA\textsubscript{A} agonism. The net result of the mechanisms of action consists of reduced spontaneous pain and hypersensitivity through membrane stabilization, reduced neurotransmitter release, and reduced postsynaptic cellular activation in the dorsal horn of the spinal cord. Table 66-10 summarizes the mechanisms of action of the various anticonvulsants.

With the exception of pregabalin and gabapentin, studies on the efficacy of the various anticonvulsants to treat pain have been inconsistent. Pregabalin is FDA approved to treat postherpetic neuralgia (PHN), painful diabetic peripheral neuropathy (DPN), and fibromyalgia; gabapentin is FDA approved to treat PHN; carbamazepine is FDA approved to treat trigeminal neuralgia; and valproic acid and topiramate are FDA approved to treat migraine headaches. The discussion in this chapter will focus on pregabalin and gabapentin, which have the most evidence in support of treating chronic neuropathic pain (Table 66-11).

**Mechanism of Action**

Pregabalin and gabapentin are synthetic molecules that are structurally related to GABA. Both structures are derived

| Table 66-7 Serotonin-Norepinephrine Reuptake Inhibitors: Pharmacodynamics |
|-------------------|-------------------|-------------------|
| **Venlafaxine**   | **Duloxetine**    | **Milnacipran**   |
| Half-life         | 5-7 hr            | 8-17 hr           | 6-8 hr            |
| Metabolism        | Hepatic           | Hepatic           | Hepatic           |
| Excretion         | Urine             | Urine 70%, feces 20% | Urine             |
| Protein binding   | 35%               | ≥90%              | 13%               |
| Therapeutic dose/day | 37.5-225 mg/day   | 20-60 g/day       | 12.5-200 mg/day   |
| Administration    | 2-3 times         | Daily             | Twice daily       |
| Titration         | ≥4-7 days/dose change | ≥7 days/dose change | 1-2 days/dose change |

**Table 66-8 Potential Tricyclic Antidepressant Drug Interactions**

<table>
<thead>
<tr>
<th>Monoamine oxidase inhibitors</th>
<th>Drugs that inhibit cytochrome P450 isoenzymes</th>
<th>Drugs that prolong the QT interval</th>
<th>Central nervous system–active drugs</th>
<th>Antidopaminergic drugs</th>
<th>Alcohol</th>
<th>Lithium</th>
<th>St. John’s wort</th>
<th>Tryptophan</th>
</tr>
</thead>
</table>

**Table 66-9 Potential Serotonin-Norepinephrine Reuptake Inhibitor Drug Interactions**

<table>
<thead>
<tr>
<th>Milnacipran</th>
<th>Duloxetine</th>
<th>Venlafaxine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monoamine oxidase inhibitors</td>
<td>Drugs that inhibit cytochrome P450 isoenzymes</td>
<td>Drugs that inhibit cytochrome P450 isoenzymes</td>
</tr>
<tr>
<td>Serotonergic drugs</td>
<td>CNS-active drugs</td>
<td>CNS-active drugs</td>
</tr>
<tr>
<td>Triptans</td>
<td>Serotonergic drugs</td>
<td>Serotonergic drugs</td>
</tr>
<tr>
<td>Diuretics</td>
<td>Alcohol</td>
<td>Alcohol</td>
</tr>
<tr>
<td>Drugs that interfere with hemostasis</td>
<td>Drugs that interfere with hemostasis</td>
<td>Drugs that interfere with hemostasis</td>
</tr>
<tr>
<td>Antidopaminergic drugs</td>
<td>Antidopaminergic drugs</td>
<td>Antidopaminergic drugs</td>
</tr>
<tr>
<td>St. John’s wort</td>
<td>St. John’s wort</td>
<td>St. John’s wort</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Tryptophan</td>
<td>Tryptophan</td>
</tr>
</tbody>
</table>

CNS, central nervous system.
by the addition of a cyclohexyl or branched-chain aliphatic carbohydrate moiety to the GABA backbone. However, the three-dimensional shapes of pregabalin and gabapentin are significantly different from GABA. The amine (NH₂) and carboxyl groups (CO₂H) are closer to each other than in the native GABA structure. It is proposed that this difference in the three-dimensional structure of pregabalin and gabapentin versus GABA accounts for their different pharmacologic activities.113,114

Although the exact mechanism of action of gabapentin and pregabalin is unknown, results from animal models suggest that pregabalin modulates neuronal hyperexcitability, resulting in analgesic and anticonvulsant effects. Reduction of neurotransmitter release occurs because pregabalin selectively binds to the α₂δ subunit of the N and P/Q subtypes of calcium channels of neurons in the brain and spinal cord, thereby modulating calcium influx into presynaptic cells.115 Therefore, there is less postsynaptic activation

### Table 66-10  Mechanism of Action of Anticonvulsants

<table>
<thead>
<tr>
<th>Sodium Channel Blockade</th>
<th>NMDA Antagonism</th>
<th>GABA Agonism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbamazepine</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Clonazepam</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Gabapentin</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lamotrigine</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Levetiracetam</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Oxcarbazepine</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Phenytoin</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Pregabalin</td>
<td>X</td>
<td>(through AMPA/kainate antagonism)</td>
</tr>
<tr>
<td>Topiramate</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Valproic acid</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*AMP, amino-3-hydroxy-5-methylisoxazole propionic acid; GABA, γ-aminobutyric acid; NMDA, N-Methyl-D-aspartate.*

### Table 66-11  Comparison of Gabapentin and Pregabalin Pharmacology

<table>
<thead>
<tr>
<th></th>
<th>Gabapentin</th>
<th>Pregabalin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FDA-approved pain indication</strong></td>
<td>Postherpetic neuralgia</td>
<td>Postherpetic neuralgia</td>
</tr>
<tr>
<td><strong>Mechanism of action</strong></td>
<td>Modulates calcium channel opening by binding to the α₂δ subunit</td>
<td>Modulates calcium channel opening by binding to the α₂δ subunit</td>
</tr>
<tr>
<td><strong>Pharmacokinetic profile</strong></td>
<td>Nonlinear: plasma concentration is not dose proportionate</td>
<td>Linear: plasma concentration is dose proportionate</td>
</tr>
<tr>
<td><strong>Oral bioavailability</strong></td>
<td>60% 900 mg</td>
<td>90% at all doses</td>
</tr>
<tr>
<td><strong>Effective dose</strong></td>
<td>1800 mg/day</td>
<td>150 mg/day</td>
</tr>
<tr>
<td><strong>Schedule</strong></td>
<td>Unscheduled</td>
<td>Schedule V</td>
</tr>
</tbody>
</table>

*FDA, U.S. Food and Drug Administration.*

Gabapentin has a nonlinear bioavailability with a reduction in absorption as the dose increases. This is the result of an active and saturable transport mechanism.116 Gabapentin 900 mg has 60% bioavailability, and a 3600-mg dose has only 33% bioavailability. More frequent delivery of smaller doses may improve bioavailability through the saturable transport mechanism. It is less than 3% protein bound, and because of its high lipid solubility, it readily penetrates into the CNS. Gabapentin is not metabolized and is excreted unchanged in the urine. Therefore, renal impairment will significantly increase the drug half-life. Studies have shown efficacy at doses up to 1800 mg and 2400 mg; however, doses above 1800 mg/day do not appear to be more efficacious, and the current maximum FDA-approved dose to treat pain is 1800 mg/day. Because of its nonlinear kinetics, titration of gabapentin to an effective dose can be prolonged with an average onset of 10 to 14 days after initiation.

### Pharmacology of Gabapentin and Pregabalin

#### Gabapentin

Gabapentin is not dependent on active transport for absorption and therefore has linear kinetics with bioavailability of about 90%. Unlike gabapentin, pregabalin absorption is independent of dose. It has negligible protein binding and readily penetrates into the central nervous system. Pregabalin is not metabolized and is excreted unchanged in the urine; therefore, renal impairment will significantly increase the drug half-life, and dose adjustments are required.113 The effective dose of pregabalin is between 150 and 300 mg per day. Owing to the linear kinetics and high bioavailability, the effective dose can be achieved in 2 to 3 days; therefore, onset is faster than with gabapentin.

### Toxicity of Gabapentin and Pregabalin

Neither gabapentin nor pregabalin has any significant drug interactions, and protein binding is minimal. Because gabapentin and pregabalin have similar mechanisms of action, their side effect profiles are very similar. The most common side effects with at least more than twice the incidence of placebo in controlled trials were dizziness, somnolence, dry mouth, peripheral edema, blurred vision, weight gain, and thinking abnormalities. Dizziness and somnolence usually began shortly after initiation of therapy, and most cases resolved with continued dosing. Dizziness and somnolence are dose related and can be reduced with slow titration. This side effect can also be an advantage when administered at night in that it will improve sleep. Controlled studies with gabapentin and pregabalin have demonstrated improved sleep quality when compared with placebo.1,2 Xerostomia (dry mouth) is dose dependent; however, it is usually mild,
Patients frequently describe muscular pain as “spasms,” and clinicians clearly know that loss of range of motion is detrimental to their patients. It makes sense that a medicine that acts as a “muscle relaxant” would be of great value in treating patients with these complaints. Unfortunately, skeletal muscle relaxants do not have a primary role in the treatment of chronic pain because of limited true effects on muscles, significant side effect profiles, drug interactions, or addiction potential.

The list of medications classified as skeletal muscle relaxants is highly diverse (Figure 66-4). Because each of the drugs in this class has a unique mechanism of action and side effect profile, each medication from this class must be examined separately. Adding to the difficulty of understanding these medications is the lack of consensus as to how or why they provide benefit. The two approved clinical indications are for treatment in upper motor neuron diseases that result in spasticity and in peripheral musculoskeletal disorders that cause pain and spasms. Each of these medications can thus be classified as an antispastic or antispasmodic agent. These medications should not be used as first-line medications and usually should be used in conjunction with other pain medications. In addition, most antispasmodic medications come with restriction of use to 2 to 3 weeks. Despite this, many skeletal muscle relaxants are prescribed on a long-term basis for chronic conditions.118

Understanding the reflex arc in muscle is important in understanding where these agents act. The simplest reflex is monosynaptic. A muscle spindle transmits an efferent signal through the Ia afferent neuron; this enters the dorsal horn of the spinal cord and synapses on an alpha motoneuron. This neuron then excites the spinal cord through the ventral root and innervates the extr fusal fibers of the same muscle, causing a contraction in the muscle. The Ia afferent signal is also transmitted polysynaptically through interneurons that inhibit alpha motoneurons of antagonist muscles, causing them to relax. Gamma motoneurons located in the ventral horn adjust the sensitivity of the muscle spindles to stretch. They travel with the alpha motoneurons and receive input from cutaneous receptors and many supraspinal pathways, including the corticospinal and reticulospinal tracts. Polysynaptic spinal reflexes with supraspinal connections are much more abundant than monosynaptic reflexes. These connections are both excitatory and inhibitory, providing improved control and feedback to set and fine-tune motor tone. The main tract that inhibits spinal reflexes is the dorsal reticulospinal tract, and the main excitatory pathway is the bulbopontine tegmentum.119

Spasticity occurs in upper motoneuron disorders such as multiple sclerosis, spinal cord injuries, traumatic brain injuries, strokes, and cerebral palsy. These diseases cause loss of descending inhibition from the brain to the spinal cord, leading to muscular hypertonicity and increased resistance to stretch. This presentation, which is classically described as a “clasp-knife phenomenon,” tends to affect the flexors of the upper extremities and the extensors of the lower extremities more than other muscle groups. Medications approved for spasticity include baclofen (Lioresal), dantrolene (Dantrium), tizanidine (Zanaflex), and diazepam (Valium).

The antispasmodic agents are used to treat muscle pain and spasms when hypertonicity, hyporeflexia, or other signs of upper motoneuron disorders are not present. This presentation is much more common than spasticity and muscular pain, and patients rarely ask for these treatments.
would include syndromes such as fibromyalgia, tension headaches, myofascial pain disorders, and nonspecific back pain. Medications include cyclobenzaprine (Flexeril), carisoprodol (Soma), metaxalone (Skelaxin), chlorzoxazone (Parafon Forte), methocarbamol (Robaxin), and orphenadrine (Norflex). Tizanidine and benzodiazepines are approved for both spasticity and musculoskeletal disorders. How each of these drugs improves these conditions is not completely clear, but the nonspecific sedative effects may have greater importance than the effects on muscles or spinal reflexes.

Antispasmodic Medications

Baclofen

Baclofen blocks presynaptic and postsynaptic GABA$_B$ receptors$^{120,121}$ and disrupts polysynaptic and monosynaptic reflexes at the spinal cord level.$^{122}$ Oral doses usually start at 5 mg three times per day with a maximum dose of 80 mg per day. This medication frequently can be used intrathecally for the same indication with a reduction in clinical side effects. Sedation, weakness, hypotension, nausea, depression, and constipation have been reported. Taper slowly and adjust for renal impairment. Withdrawal symptoms include hallucinations, seizures, and pruritus.

Dantrolene

Dantrolene acts directly on the muscle by decreasing the release of calcium from skeletal muscle sarcoplasmic reticulum.$^{123}$ It has a black box warning about hepatotoxicity.$^{124}$ Its risk is much greater with patients taking 800 mg per day or more compared with doses of 400 mg per day. It is recommended to stop this medication after 45 days if no benefit is derived.

**Tizanidine**

Tizanidine is a centrally acting agonist of the $\alpha_2$ receptor,$^{125,126}$ which reduces the release of excitatory amino acids from the presynaptic terminal of spinal interneurons.$^{127}$ This reduces tonic stretch reflexes and polysynaptic reflex activity.$^{128}$ Doses start at 2 to 4 mg at night and increase by 2 to 4 mg three to four times per day with a maximum dose of 36 mg per day. Side effects include hypotension, sedation, and dry mouth. Its major route of elimination is renal, and renal impairment can significantly decrease clearance. It is metabolized via CYP1A2, and concomitant use with ciprofloxacin or fluvoxamine is contraindicated.$^{129}$ Elevated liver functions—three times the upper level of normal—have occurred in 5% of patients$^{129}$ Monitoring of liver function is recommended. Acute withdrawal can cause hypertension, tachycardia, and hypertonia. Use with alcohol should be warned against, and oral contraceptives will decrease clearance with doses as low as 4 mg per day.

**Diazepam**

Diazepam works by central blockade of GABA$_A$,$^{123,130}$ which is an inhibitory neurotransmitter in the central nervous
system. It was the first drug used to treat spasticity and is often used to compare new agents in clinical trials. Efficacy for spasticity and muscle pain is believed to be similar to that of benzodiazepines as a class. Clinical trials showing efficacy with benzodiazepines are based on tetrazenapam, which currently is not available in the United States. It is not considered a first-line agent for spasticity or muscle pain because of its sedative effects, drug interactions, and significant abuse potential.

**Antispasmodic Medications**

**Cyclobenzaprine**

Cyclobenzaprine is structurally related to tricyclic antidepressants and thus causes sedation through its anticholinergic effects. It acts primarily at the brain stem to reduce tonic somatic motor activity influencing both gamma and alpha motoneurons. Dosing at 5 mg three times per day has shown similar efficacy when compared with 10 mg three times per day with reduced side effects. The maximum daily dose is 30 mg. This medication also comes in 15- and 30-mg extended-release formulations. Concomitant use with MAO inhibitors is contraindicated. This drug should not be used during the acute recovery phase of myocardial infarction, or in patients with arrhythmias, heart block, congestive heart failure, or hyperthyroidism. The risk of seizures with tramadol is increased if this medication is added. The risk of serotonin syndrome can be increased in patients taking selective serotonin reuptake inhibitors.

**Carisoprodal**

Carisoprodal in animal studies has shown muscle relaxation through altered interneuronal activity in the spinal cord and in the descending reticular formation of the brain. It is metabolized in the liver to meprobamate, a Schedule IV medication with abuse potential. Meprobamate binds to GABA receptors, which results in further sedation. Doses start at 250 mg up to four times per day with a maximum dose of 1400 mg per day. It is contraindicated in patients with a history of acute intermittent porphyria. CYP1C19 inhibitors such as omeprazole or fluvoxamine may increase carisoprodal levels and decrease meprobamate. CYP1C19 inducers such as rifampin or St. John’s wort increase exposure to meprobamate.

**Methocarbamol**

Methocarbamol, a carbamate derivative of guaifenesin, is a central nervous system depressant with no direct action on striated muscle, the motor end plate, or the nerve fiber. It comes in 500- and 750-mg tablets and can be given up to four times per day with maximum dose of 8 g per day. The parenteral form has been associated with skin sloughing and thrombophlebitis. It should be used with caution in patients with myasthenia gravis receiving anticholinesterase agents.

**Metaxalone**

Metaxalone also has no direct actions on muscles. Its effects occur mainly through generalized central nervous system depression. It tends to cause much less sedation than other drugs in this class and has limited drug interactions. It is contraindicated in drug-induced, hemolytic anemias and in significantly impaired renal or hepatic function. The dose most commonly used is 800 mg three or four times per day.

**Chlorzoxazone**

Chlorzoxazone is a centrally acting agent that acts primarily at the level of the spinal cord and subcortical areas of the brain, where it inhibits multisynaptic reflex arcs. Doses are 250 to 750 mg three or four times per day. Hepatocellular toxicity has been reported with this drug.

**Orphenadrine**

Orphenadrine is derived from diphenhydramine. It does not directly relax muscles and possesses greater anticholinergic properties than diphenhydramine. Its dosage recommendation is 100 mg twice a day. It is contraindicated in patients with glaucoma, pyloric or duodenal obstruction, stenosing peptic ulcers, prostatic hyperplasia or obstruction of the bladder neck, cardiospasm (megaesophagus), and myasthenia gravis.

**Efficacy**

Clinical studies using these drugs are very limited, poorly controlled, and of short duration. One meta-analysis found fair evidence that tizanidine and baclofen are roughly equivalent and are similar to diazepam for the treatment of spasticity. Investigators go on to state that data on efficacy for muscle relaxants with musculoskeletal conditions are limited, but for treatment compared with placebo, cyclobenzaprine, tizanidine, carisoprodol, and orphenadrine showed a consistent trend favoring active treatment. Data demonstrating effectiveness for chlorzoxazone, methocarbamol, baclofen, and dantrolene are limited. Last, data for metaxalone are mixed.

**EMERGING TARGETS**

**Nerve Growth Factor Inhibitors**

Nerve growth factor (NGF) is a ligand to the tyrosine kinase (TrKA) receptor located on sensory neurons. Inflammation results in increased levels of NGF, which stimulate the TrKA receptor leading to an increase in inflammatory mediators and increased sensitivity to pain. Overexpression of NGF also leads to a proliferation of sympathetic fibers, which may be important in the production of inflammatory and neuropathic pain. Because NGF is essential for the survival and development of sensory neurons, NGF inhibitors may lead to peripheral nerve dysfunction and neuropathies. Tanezumab, a recombinant humanized monoclonal antibody targeting NGF, is in phase II trials and is being developed to treat pain associated with osteoarthritis, low back pain, and metastatic bone cancer.

**Cannabinoid Agonists**

Cannabinoid receptors type 1 (CB1) are located at multiple locations in the peripheral and central nervous system,
whereas CB2 receptors are located on inflammatory cells (monocytes, B/T cells, mast cells). CB2 activation results in a reduction in inflammatory mediator release, plasma extravasation, and sensory terminal sensitization. Activation of peripheral CB1 receptors results in a reduction in the release of proinflammatory terminal peptides and a reduction in terminal sensitivity. Activation of central CB1 receptors leads to reduced dorsal horn excitability and activates descending inhibitory pathways in the brain. The net result is a reduction in both pain and hyperalgesia. Inhaled cannabis has been extensively studied in various pain syndromes with mixed results. More recent well-controlled trials have shown promise. Stativex is a sublingual spray containing a mixture of tetrahydrocannabinol and cannabidiol, which is currently in phase II trials for the treatment of cancer-related pain. It has been shown to reduce the pain of rheumatoid arthritis and chronic pain of various other causes.

**AMPA/Kainate Antagonists**

Glutamate, an excitatory amino acid neurotransmitter, has been implicated in pain perception. Two types of glutamate receptors are found in the central nervous system: ionotropic and metabotropic. Ionotropic receptors are further functionally categorized into NMDA and non-NMDA receptor subtypes (AMPA and kainate receptors). AMPA and kainate receptors are prevalent within the dorsal horn of the spinal cord, and activation is thought to mediate rapid excitatory neurotransmission, resulting in central sensitization mechanisms. Tezampanel, an intravenously administered antagonist of AMPA and kainate receptors, has been shown to be effective in reducing human experimental pain. NGX426, an oral prodrug of tezampanel, is in clinical development.

**Selected References**

1033

CHAPTER 66 | ANALGESIC AGENTS IN RHEUMATIC DISEASE


120. Baclofen table product labeling, Minneapolis, 2002, Upsher-Smith Laboratories Inc.


122. Dantrium (dantrolene sodium) package insert, Manson, Ohio, September 2002, Proctor & Gamble Pharmaceuticals.


Full references for this chapter can be found on www.expertconsult.com.
References


122. Baclofen table product labeling, Minneapolis, 2002, Upsher-Smith Laboratories Inc.


The average course of rheumatoid arthritis (RA) is about 27 years. The usual psychological defenses wear out. The illness becomes integrated, an unwelcome but integrated part of the individual.

“I used to believe that ‘my arthritis’ was separate from myself. I have come to understand, after 30 years of living with JRA [juvenile RA], that this is no less a part of me than my fingers, my voice. My appreciation for pain, especially that which goes unseen, for recognizing the reality of potential (not everyone can be whatever they want), and for the ability to work at accepting one's limitations has guided and still guides my perspective. Everything about who I am is wrapped in and around this illness.”

—Woman with juvenile RA since age 6

Additional Effects of Chronicity

Equal to pain for many patients is the loss of control and self-confidence that comes with chronicity.

“But when my symptoms are present, my self-image changes—I am aging, tentative, dependent, tinged with self-pity and regret. I am more willing to accept, rather than strive. In short, my confidence in my life and myself is reduced by how I feel physically. Despite this, a part of me rebels against this lowered self-esteem. It effectively refuses to accept the disease, to admit to not being able to do the things I once did with ease or even skill. These two conflicting personae—the hesitant reflective sick person and the unaccepting person who does what he wants despite how he feels—arise from each other in an uninterrupted, shifting pattern. Self-pity, which is the last stage of my sick person, awakens my healthy person. When he is eventually confronted with a task, however mundane, that symptoms prevent him from doing, the sick person reemerges, and the cycle begins anew.”

—Man with ankylosing spondylitis of 20 years’ duration

Initial events of chronic pain and loss of control and confidence give rise to reactions and adaptations analogous to stages of reactions to acute losses in general—anger, denial, anxiety, and resignation. Management of these reactions and adaptations in the chronic setting may represent new problems of their own from the patient’s point of view, depending particularly on the responses of the patient’s social support system (family, friends, co-workers). Above all, during this series of emotional events, there is the possibility of the existence of treatable depression.

“Any patient with RA who tells you they haven’t ever been depressed is lying. When I see those questionnaires about depression, I just mark something in the middle of the scale. It’s embarrassing to admit how bad you feel.”

—50-year-old registered nurse with RA of 10 years’ duration
FOUR LANDMARKS OF PSYCHOSOCIAL ADAPTATION IN RHEUMATIC DISEASES

Virtually every rheumatology patient faces four landmark events or processes that routinely require some active management: (1) the initial ambiguity of not getting a diagnosis, (2) getting a diagnosis and the denial that comes with it, (3) repeatedly struggling to understand risks and benefits, and (4) dealing with the impact of the disease on relationships with others. Everyone either receives a diagnosis or does not receive one. If there is no immediate diagnosis, it may help to realize that such ambiguity is a good prognostic factor. If the diagnosis is clinically and serologically obvious, the prognosis may be poor. When a clear diagnosis is made, it entails the patient overcoming denial and facing the fact of having a chronic illness. Subsequently, the risk-benefit decisions about therapies are ongoing. The disease changes; decisions have to be made again.

Family Relationships

Of these four landmark situations that almost every rheumatology patient encounters—facing the diagnosis of a chronic disease or the uncertainty of no diagnosis, understanding repetitive cost-benefit decisions and threats to family relationships—the last uses the most psychological capital. Helpful data in this fourth landmark area are available, but no template approach is known, especially when the patient is a child. In the latter case, the re-establishment of a peer group, through school, camp, or any other setting, is a primary goal (see Chapter 107). Despite the loss of self-confidence and the effects of pain, fatigue, and deformity, individuals with RA have the same divorce rate as the general population. RA patients have fivefold less remarriage if divorced, however. The rate of progression of disability in married patients is substantially less than for patients who are unmarried. Healthy husbands of wives with RA are affected by the realization of the vulnerability of their own health. Marriage increases social support and is a favorable psychosocial variable. RA is experienced in an interpersonal context that can be hard to understand. In couples, if the woman has RA that she perceives as severe, she is helped by congruent beliefs in the spouse. However, if the wife’s emotional response to the RA is high, she is helped by fewer congruent emotions on the part of the husband. Childhood and adolescence are particularly unfortunate times to encounter rheumatic disease.

The physician and the health care system can enable social support to some degree. From studies in systemic lupus erythematosus (SLE), it seems that such enabling effort has the greatest effect among patients who already have strong families, higher incomes, and low disease activity.

FATIGUE

Fatigue in rheumatic diseases seems to be a complex admixture of effects of circulating cytokines presumably mediated by binding to receptors within the periaqueductal gray matter (reticular activating system) of the brain stem, deconditioning, and depression. These factors appear in differing proportions in different diseases. Fatigue is the most common chief complaint of patients with RA. For patients with RA, fatigue is prevalent, correlated with depression when that occurs, and is the focus of elaborate coping strategies. The pathophysiology of chronic widespread pain, despite the lack of obvious inflammation clinically, includes a contribution from cytokine-related fatigue. Patients with chronic widespread pain have reduced levels of anti-inflammatory cytokines interleukin-4 and interleukin-10. The rapid time course of the dramatic response of RA fatigue to tumor necrosis factor blockade—less than 1 week in some patients—shows that in many RA patients, fatigue must be due to circulating cytokines directly, rather than to deconditioning and psychological factors, which would not be expected to improve so rapidly.

In these patients, an anti–tumor necrosis factor agent essentially turns RA fatigue into SLE fatigue, which apparently is predominately due to deconditioning and psychological factors. Maximum oxygen consumption for patients with RA is comparable with that of patients with SLE (slightly >50%). In contrast to RA, in SLE there is no association between fatigue and disease activity, or markers of inflammation. Unless cytokines are acting locally within the central nervous system with no concentration in serum, psychosocial factors may be a predominant cause of lupus fatigue. This mechanistic explanation begs the question of whether or not depression in SLE might be due to tonically elevated (or decreased) cytokines that do not vary much with disease activities. Transforming growth factor-β is an example of the latter type of cytokine elevation in SLE.

Another key reason for fatigue is treatable deconditioning. When 93 patients with SLE were compared with 44 sedentary controls, the patients were about 20% worse off with regard to muscular strength, forced expiratory volume, and maximum oxygen consumption. A smaller interventional study of exercise physiology preceded this one and found the same baseline. Patients were selected or pretreated to suppress SLE disease activity and underwent 8 weeks of aerobic conditioning, which was successful in reducing their fatigue by about half. A stepwise regression model explaining 37% of variation in SLE fatigue featured abnormal illness-related behaviors, greater age, greater degree of helplessness, and lack of health insurance as potentially responsible psychosocial variables.

DEPRESSION

Diagnosable psychiatric disorders occur no more frequently in patients with rheumatic diseases than in patients with any other chronic disease. Two exceptions stand out statistically. First, patients with SLE have cognitive deficits, particularly as related to learning new things. Mild cognitive deficits, such as word finding difficulty, are prominent. Second, patients with widespread, nonarticular, axial aching have more diagnosable psychiatric illness—usually depression or anxiety—and are more likely to somatize. Patients across the spectrum of those usually seen in office practice rheumatology were compared in an attempt to map out the relationship between pain, disability, and abnormal psychology across diagnostic groups. The highest pain scores and the greatest psychological distress were reported by patients with axial disorders such as fibromyalgia, back pain, and neck pain. Patients with serious peripheral joint disease from RA had the most disability but the least pain and the
most normal psychological test scores. As a rule, patients with RA psychologically adapt over time and usually are as well adjusted and productive as their anatomy allows.\\textsuperscript{19}

**Rheumatoid Arthritis**\\

Most patients with moderately severe RA have been depressed. In a preliminary review of veterans’ diagnoses (n = 8706 patients with RA), depression was more prevalent in RA (odds ratio, 1.48), more so in women and African-Americans, and less in older patients. Patients are resilient, however, and they adapt. Even the distress accompanying chronic joint pain in RA, described earlier as the “illness experience,” fails to lead to sustained, treatable clinical depression most of the time. In RA, depression is not more common than among other rheumatology office patients, and pain causes psychological distress rather than the other way around.\\textsuperscript{19} To an extent, the mental health aspect of RA may run an unexpectedly independent course. In one recent trial of a high-value (biologic) intervention, the mental health scale on the Short Form (SF)-36 was the outcome least correlated with the Disease Activity Score (DAS)\\textsuperscript{28} and the American College of Rheumatology 20% response criteria (ACR20), and the pain score was the best correlated.\\textsuperscript{29} However, the mind-body dichotomy is not a rigid Cartesian one in RA. Circulating cytokines can exacerbate depression, and interleukin (IL)-6 might be more important than tumor necrosis factor (TNF) or IL-1 in this regard.\\textsuperscript{30} It remains possible, although unproven, that psychological stress could exacerbate rheumatoid arthritis via hormonal effects upon neuronal transmission.\\textsuperscript{22} Conversely, a systematic review of 27 different trials of psychological interventions for RA concluded that these interventions do increase physical activity levels, which at least has the potential to favorably alter structure.\\textsuperscript{23}

**Knee Osteoarthritis**\\

Self-reported knee pain in the United States has the highest prevalence among elderly individuals and non-Hispanic black women.\\textsuperscript{24} Only younger and less educated patients presenting with knee osteoarthritis are more likely to be depressed, however.\\textsuperscript{25} Multiple regression analysis of a study of a mixed Asian population in Singapore revealed numerous independent variables associated with better outcome—some biologic, some fixed psychosocial, and some potentially modifiable. Among the best functioning patients with knee osteoarthritis were not only patients with less pain and less radiographic severity, but also patients with better education and less feeling of helplessness (“learned helplessness”).

**Chronic Widespread Pain, Fibromyalgia, and Low Back Pain**\\

Overall, population-based studies show patients with chronic, nonarticular, axial aching rheumatic syndromes to be the group most likely to have an underlying psychiatric diagnosis. It is usually a mood disorder, such as depression or anxiety. Somatization is an important causal factor, at least for chronic widespread pain ascertained at the community level. Arguably, fibromyalgia and chronic widespread pain are physiologically distinct. These conditions nonetheless form part of the same differential diagnostic group of nonarticular, axial aching and may overlap in the survey or population-based studies most generalizable to office practice.

Chronic widespread pain among unselected general populations is associated with a 17% incidence of a psychiatric diagnosis in Australia;\\textsuperscript{26} one-quarter of a community sample of similar patients in Manchester, United Kingdom, had a psychiatric diagnosis.\\textsuperscript{27} In the Manchester group, the odds ratio for having a psychiatric illness if accompanying self-reported chronic widespread pain was 4.9 (95% confidence interval, 2.6 to 9.5). In the Australian phone survey (n = 17,545), 18% of the population had widespread chronic pain, which interfered with the daily activity of two-thirds. Psychosocial markers statistically associated with widespread chronic pain included less education, less income, lack of health insurance, and unemployment and disability payments. The same strong ties to psychosocial factors distinguish low back pain manifesting as part of chronic widespread pain from low back pain by itself. Finally, in prospective studies, patients initially without generalized axial aching, but who somatize, have increased risk of developing chronic widespread pain later.

**DISABILITY, EMPLOYMENT, AND ECONOMICS**\\

Health-related quality of life is significantly associated with paid work even when disease severity and demographics are controlled for, and even when societal safety nets are robust.\\textsuperscript{30} Individual patients with RA lose 30% of their work capacity. Societal costs of RA in economic and employment terms have now been estimated accurately at $19.3 billion in 2005 dollars, with the burden divided as follows between four main stakeholder groups: patients 28%; employers 33%; and families and government nearly equal at about 20% each. Less firm estimates of quality of life deterioration and mortality double the total cost.\\textsuperscript{30} A bit less than half of the cost of a biologic can be paid for by the indirect cost gain from increased labor force participation.\\textsuperscript{30} Psychosocial factors affect the course of rheumatic diseases, especially with regard to employment and disability, against the background of this societal, economic context.

The effects of psychological and social factors on the presentation of many important diseases have been studied. These illnesses include the most statistically and economically important of the rheumatic diseases—osteoarthritis, RA, and low back pain—and SLE and fibromyalgia, as partially described earlier. Likewise, the effects of these diseases upon employment have been described. For Germans nationwide, ages 20 to 59, who had more than a decade of disease duration, the rank order of risk of unemployment from various rheumatic diseases was, from highest to lowest, granulomatosis with polyangiitis (formerly Wegener’s granulomatosis), rheumatoid arthritis, scleroderma, lupus, psoriatic arthritis, and ankylosing spondylitis. The largest differences were noted between better-off psoriatic and spondyloarthropathy groups and the others.\\textsuperscript{31} In Canada, the long-term work disability rate in RA after 10 years of disease duration is somewhere between 32% and 50%.\\textsuperscript{27} Whether early interventions to prevent unemployment are worth the
expense depends on accounting assumptions about how closely interests of the paying agency are aligned with those of patients receiving the intervention. In RA, failure to respond within the first 6 months at least to the level of an ACR20 predicts job loss even when a large response (ACR50) comes later. It is interesting to speculate on how much of this disproportionate impact of the first 6 months upon employment is due to downstream effects of biologic events unique to the first 6 months of clinical disease versus how much is due to early psychological adaptations.

After severity of RA, the top two determinants of RA work disability are the physical difficulty of the job and the educational level of the patient. Beyond these first three powerful factors, some psychological factors are important. These include most prominently a feeling of helplessness and employer support. One large difference has been noted between patients’ and doctors’ perceptions about continuation of employment with RA. This difference involves employer support, which providers tend to underweigh compared with the importance that patients ascribe to it. Low employer support could be viewed as a modifiable risk factor and therefore a target of psychosocial intervention.

### Disability Insurance

In understanding disability and communicating that understanding to patients, four different funds of knowledge are of special value: (1) detailed knowledge of the mechanics and criteria of Social Security disability insurance (SSDI) in the United States, (2) cross-cultural comparisons, (3) an appreciation of the major criticism raised in critique of the logical assumptions underlying the current system, and (4) knowledge of research results showing what determines disability in specific diseases. Achieving reasonable expectations on the part of the patient and less frustration for the provider and the patient are two attainable goals of disability management.

The mechanics of the U.S. SSDI system vary from state to state but are simple and easily summarized. Reforms have reduced the number of steps between the initial application and the administrative law judge level. At this level, as an example, the court is empowered to ask and answer the question of whether a skin lesion, joint symptoms, and cognitive deficits in a patient with SLE are equivalent in severity to a stroke or renal failure, the actual default criteria. The waiting period before final adjudication of the disability claim is now shorter but still varies from state to state in the United States. Opportunities for the applicant to introduce new medical facts after the initial application is submitted are diminishing, however. Therefore, completeness of data at the initial application has greater importance. The appeal level in the streamlined version of the SSDI process reviews the correctness of the decision but allows few new medical data to enter the picture during the review. The patient has less opportunity to buttress medical weak points in his or her case in response to feedback from the system. The algorithm for advocacy therefore encompasses three elements, in order: (1) accrue the greatest quantity of objective data as possible, whether or not they are clinically necessary, being especially careful to document radiographic changes in patients with RA or other inflammatory arthritis; (2) write details about limits in daily and work activities in the

notes before beginning correspondence about disability; and (3) begin the application using an attorney with steps 1 and 2 firmly in place.

An algorithm for employer-provided private disability insurance in the United States might be entirely different. This difference arises from answers to the questions of who pays what when the patient is on light duty, part-time accommodation, short-term disability, or long-term disability. Some employers prefer to push ambiguous cases onto a long-term disability account, which is insured, rather than carrying a less than 100% effective employee in a current budget. Such an assignment to long-term disability insurance (which might eventually run out) may impede the patient’s ultimate return to full-time work at the same firm, however.

Cross-cultural comparisons offer little help to individual patients experiencing economic losses, which constitute the indirect costs of rheumatic diseases. Between 1983 (Netherlands) and 1996 (Australia), many countries passed a milestone, however, at which the absolute numbers of individuals drawing disability payments surpassed the official number of unemployed for the first time. The United States and the United Kingdom passed this marker around 1993. Over the three years 2000 to 2003, 1.8 million people claimed work disability benefits in the United States. Economists speculate that these numbers were then large enough to affect macroeconomic statistics on the workforce participation rate. Economists also argue that political incentives to record low unemployment numbers and the personal incentives of more payments for disability than for unemployment favor continuation of this trend for as long as it is fiscally sustainable.

In the long term, this may mean that disability insurance will become more like unemployment insurance. Such a change would have to involve acknowledgment that disability and unemployment are related aspects of a single work problem, significantly greater incentives for employers to hire and promote disabled individuals, and new incentives for patients to avoid disability rolls in favor of remaining in the workforce. Regardless of the future, and of an individual patient’s struggle to salvage income levels in the face of illness, insight into the relationship between unemployment and disability can help maintain a perspective without which the disability issue has great potential to split the therapeutic alliance between physician and patient.

The major philosophical underpinning of the current SSDI system is the assumption that the amount of disease determines the amount of work disability. A necessary corollary is that the amount of disease is objectively quantifiable. From the rheumatologic standpoint, two things are wrong. First, pain is self-reported and, by definition, not objectively quantifiable. Second, and more important, psychosocial factors are stronger determinants of disability than disease factors in many cases. Examples of the latter situation have been studied in a range of illnesses as outlined subsequently.

In one multivariate analysis of knee osteoarthritis only, age, female sex, obesity, and severity of knee pain maintained independent effects ($P < .0009$). All of these factors, individually and together, had greater influence on health assessment questionnaire–determined disability than did radiographic severity. For 180 English patients with low
back pain, the best predictor of some pain persisting at 1 year of follow-up was concomitant widespread pain (odds ratio, 6.4). Other factors, from the most clinical, such as radicular pain, to the most psychological, such as job dissatisfaction, all showed equal odds ratios of about 3. Lack of influence over the work environment and poor social relations with co-workers are specific job-related factors influencing work disability from back pain.

Whiplash injury was definitively studied in an experiment of nature when Saskatchewan switched from tort to no-fault auto insurance in 1995, abruptly disallowing claims for pain and suffering. Analysis of 7462 claims showed that the number and prognosis of neck injuries improved by about 30%. Depression influenced the duration of neck pain. The relative parity of psychosocial factors with clinical disease factors seen in the spine seems to be maintained in analyses of more peripheral musculoskeletal problems, such as forearm pain. Patients with RA and patients with fibromyalgia report equal disability on a self-report scale, but patients with RA report less pain.

Labeling a patient with widespread pain or nonarticular axial aching as having fibromyalgia probably does not increase work impairment. Disability awards for RA are heavily influenced by objective radiographic data—erosions and narrowing. Age, pain score, and depressed mood secondary to physical disability determine overall disability in RA, however, from the patient’s point of view. Fatigue plays a role in RA disability that fatigue, more disability, more anxiety, and less social support all are correlated with one another in RA. The odds are 2:1 to 3:1 against return to work for a patient with RA who stops working. SSDI, older age, and less education decrease the odds. In SLE multivariate models for patients at university centers, predictors of early disability in the first 3 years included low educational level (P = .0004), a physical job (P = .0028), and disease activity at diagnosis (P = .0078). Cumulative organ damage and disease duration, race, and gender did not predict stopping work.

One randomized, unblinded trial of 13,077 Madrid workers with temporary work disability in 1998 and 1999 showed a significant effect of an inexpensive, organized program. This program cut the incidence of long-term disability by half and saved about $5 million using more frequent visits and patient education early on.

**DETERMINANTS OF HEALTH CARE UTILIZATION AND DISPARITIES**

**Large Joint Osteoarthritis, Early Course**

Depression seemed to influence younger patients and less well educated patients to seek attention for osteoarthritis of the hip or knee among a community interview sample of 108 patients older than 50 years. In this study, about half of the variance in the incidence of depression was explained by the effects or interactions of age, education, and self-assessment of osteoarthritis impact. Biologic and sociologic variables were analyzed together in a second community-based, much larger (n = 1272) survey study. Pain and other factors, none of which were modifiable psychosocial features, were independent influences on disability. Of several psychosocial variables affecting disability and usage, only the psychological variable—depression in a younger patient—is amenable to direct intervention.

**Large Joint Osteoarthritis, Late Course, Total Joint Arthroplasty**

Later in the course of osteoarthritis, patient education, rather than detection and treatment of depression, may play the strongest role. In countering the effects of some sociologic factors important in selecting patients for total joint arthroplasty (TJA), specific defects in a patient’s fund of knowledge can be targeted. African-American veterans have 10% higher Western Ontario and McMaster Universities Arthritis Index (WOMAC) scores in a multivariate regression model adjusting for severity. Nevertheless, the subsets of African-Americans and older elderly patients tend to accept fewer arthroplasties. In a study of 596 older Cleveland veterans with symptomatic hip or knee osteoarthritis, African-American men in particular were only about half as likely (odds ratio, 0.50; 95% confidence interval, 0.30 to 0.84) to accept a recommendation for TJA. Underlying reasons for this racial disparity consist predominantly of expectations on the part of African-Americans of a more prolonged and troublesome postoperative period. African-Americans also were less familiar overall with TJA than were nonblacks. Socioeconomic status per se has not been found to determine acceptance of TJA. In a Canadian survey of 48,218 people older than 55, 3307 had hip and knee problems verified as osteoarthritis by clinical examination and radiography. Individuals with the lowest socioeconomic status were equally willing to undergo TJA and needed it more than others, as might be expected from their lower educational attainment, leading perhaps to more physical occupations.

In-depth interviews of elderly patients reveal three common misconceptions influencing decisions about TJA. First, older patients expect that if their physician fails to make an explicit recommendation for TJA, this means they are not candidates. Patients believed that their provider was actively screening for the procedure, and that they had already been screened out. Second, elderly patients tend to believe that more pain and daily impairment are needed to make a favorable risk-benefit tradeoff than do orthopedic surgeons. It has been suggested that this view is due partially to the fact that elderly patients are much more sensitive than their physicians to risk of mild postanesthesia cognitive deficits. Third is a philosophical difference: Elderly patients viewed osteoarthritis as normal aging to be tolerated, rather than as a disease to be operated on.

**INFLUENCE OF READING AND EDUCATIONAL LEVELS**

Generally, people with the lowest literacy have the worst health. Whether reading level per se carries an independent effect or is associated with other sociodemographic features such as poverty is uncertain. A mechanism by which impaired literacy might have direct effects is clear enough, however. Patient information pamphlets, medication sheets, and other materials are written at reading levels too difficult for most American adults, putting a premium on
Table 67-1  Evidence-Based Psychosocial Interventions

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Target</th>
<th>Method</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthritis self-help/courses,* self-management programs</td>
<td>All arthritis diagnoses, fund of knowledge</td>
<td>4-12 hr of didactic sessions, trained leader, flip charts, overheads, handouts, discussion aimed at problem solving; pain, stress and depression management, medications, and communicating with providers</td>
<td>43% decrease in office visits over 4-yr follow-up; decreased pain satisfying criteria for clinical significance (15%-20% on visual analogue scale); increased exercise, increased fund of knowledge and self-efficacy; 9% decrease in disability; most annual cost savings by patients with RA; $162; sustained effect 4 yr after the intervention</td>
</tr>
<tr>
<td>Aquatics exercise, YMCA*</td>
<td>All arthritis diagnoses; range of motion, strength</td>
<td>12-30 hr over 6-10 wk; heated pool, lifeguard, trained instructor, physical therapist</td>
<td>13%-17% improvement in range of motion, strength in lower extremities, and self-efficacy; exercise compliance declines over time</td>
</tr>
<tr>
<td>Land-based exercise programs such as Arthritis Foundation PACE or walking programs Phone follow-up of osteoarthritis</td>
<td>All arthritis diagnoses; range of motion, strength</td>
<td>Varies</td>
<td>Decreased depression and increased self-efficacy with respect to activities of daily living; decreasing adherence to exercise over time57</td>
</tr>
<tr>
<td></td>
<td>Symptomatic, radiographic knee osteoarthritis</td>
<td>Symptom monitoring phone call on a scheduled basis</td>
<td>Pain improvement similar in magnitude to that attained with an NSAID56</td>
</tr>
</tbody>
</table>

*Organized, current effort of Arthritis Foundation local chapters with certified instructor. NSAID, nonsteroidal anti-inflammatory drug; PACE, People with Arthritis Can Exercise program; RA, rheumatoid arthritis.

non–literacy-dependent patient education. This persistent fact may be among the several reasons why face-to-face programs with groups of patients are particularly effective (Table 67-1).

The Internet promises to multiply the effects of reading and education and research,58 but it may also contribute to a digital divide effect in the care of rheumatic diseases. On a theoretical basis, sociologic interventions touching on fundamental factors such as reading and education levels of the population as a whole might be expected to be the most effective psychosocial interventions for arthritis and rheumatic diseases and other diseases. It is logical that most proven psychosocial interventions consist of individual or small group provider-patient communications or patient education interventions that have the potential to overcome educational deficits and compensate for uncoordinated care (see Table 67-1).

Reading and education are complex psychosocial factors that may be associated with outcome through interaction with other variables as explanatory variables that help understanding, but offer poor targets for intervention; as epiphenomena rather than causes; and as carrier variables that are markers for other factors making the real contribution to outcome. The many abilities needed to gather and act on information used to participate actively in the care of a complex rheumatic disease such as SLE or RA might be seen in a study only through the carrier variable of educational attainment.

Another example is the moderately strong dose-response relationship between physical workload and the incidence of hip osteoarthritis. Farming and heavy lifting mediate this association, with an odds ratio of 3:1.59 In this context, educational attainment is inversely related to exposure to heavy labor. Likewise, if high-salaried individuals with gratifying jobs and bonuses in their salary structures avoid disability applications for RA, whereas hourly workers do not, this tendency might be an understandable indirect effect of education on continued employment. These complex sorts of relationships of education to clinical outcome should not obscure more direct effects, however. Education and reading level also can operate directly on visit and medication compliance and have particularly noticeable effects on outcomes of complex diseases such as RA and SLE.

PSYCHOSOCIAL INTERVENTIONS

To be a target for intervention, a psychosocial factor should be shown to be a potential cause of a clinical outcome and should be modifiable. Psychosocial factors, such as education and mood, fit these criteria and bear directly on psychosocial outcomes, such as patient satisfaction and job preservation. In addition, socioeconomic and psychosocial factors help determine biologic outcome. The effects of psychosocial factors on biologic outcome are mediated primarily by visit compliance, medication adherence and availability, and funds of knowledge about disease and the health care system. Psychosocial interventions effective in randomized clinical trials tend to revolve around the most modifiable factors (e.g., funds of knowledge). These interventions involve transfer of information or attitudes via patient education or peer support groups (see Table 67-1).

A gap between educational efforts and long-term results persists. A systematic review of 11 randomized psychoeducational interventions for patients with RA found that better designed studies showed smaller effect sizes, and long-term changes in health status (as opposed to improved knowledge per se) were not demonstrated.60 However, at least one trial of cognitive-behavioral therapy in fibromyalgia produced physical functioning effects lasting 6 months.61
Fixed versus Modifiable Determinants

There is an obvious logic to studying psychological and social factors together. Distinct features separate psychological factors from socioeconomic factors as they apply to practice, however. First, psychological characteristics of physicians and patients are made up of fairly fixed personality traits combined with states such as mood and other modifiable factors, such as fund of knowledge about a disease or medication. In contrast to psychological states, sociocultural features are almost certainly fixed. As Virchow wrote of the turbulent 1840s in Germany, “Politics is medicine written large.” His observation remains applicable. In addition, because sociologic variables cannot be easily changed in a hypothesis-testing experiment, inference of causality is problematic, although analysis of variance can help.

Even when the biology of the disease is complex, the genetics heterogeneous, and access barriers likely, as in U.S. urban or rural populations with SLE, detailed analysis suggests that education and counseling, coordinated with medical care, improve outcome. These studies found SLE disease activity and overall health status to be most strongly associated with potentially modifiable psychosocial factors, such as self-efficacy for disease management. SLE organ damage, in contrast, was best predicted by conventional clinical factors, such as duration of disease. Notably, outcomes were not associated with race. Another large study indicates that noncompliance is the largest psychosocial factor influencing outcome and confounded with race. African-American race per se had no independent effect in this large cohort in Baltimore. In patients with SLE, even when the sociologic factors seem overwhelming, and the biology of the disease and its therapies complex, targeted psychosocial interventions concentrating on patient education seem to have the potential to change outcome. In lupus, for example, self-efficacy and couples interactions are identified as such modifiable factors.

Psychosocial interventions fall into three main categories as follows: (1) efforts to improve physician-patient communication in usual care, (2) organized programs aimed at teaching or influencing groups of patients, and (3) psychopharmacology or psychotherapy. A more robust literature describes psychosocial interventions as treatment adjuncts to maximum medical treatment in ameliorating RA symptoms, improving coping and reducing disability. A recent review has reported on worksite interventions for people with RA or osteoarthritis (OA). Effect sizes are modest but represent improvement that is available cheaply after even after medical therapy is maximized. The first trial, which attempted head-to-head comparison of the three basic elements of these interventions, concluded that education effect size was about equal to that attained with more expensive-to-administer stress management and cognitive-behavioral therapies emphasizing coping. Thus the focus returns to patient education per se, and therefore to health literacy and physician-patient communication.

Physician-Patient Communication

Differing concerns, agendas, or fears; sociocultural differences; and assignment of different meanings to formulations such as “I’m fine” can all constitute barriers. Response shift is the term for psychological adaptation to a worsening disease course affecting the reproducibility of responses to questions as simple as “How are you doing?” or affecting responses to more complex questionnaire instruments. Response shift could affect observer-independent measures of outcome and quality of life in studies, as well as providers’ conclusions regarding individual patient encounters.

Canadian patients (n = 197) with established diagnoses representative of office practice except for the omission of widespread pain and fibromyalgia answered the following question: What do patients want? Factor analysis condensed responses to questionnaires covering 32 possible concerns down to eight points. Patients asked for help in five areas and extra information in two. The five areas mentioned for help were psychological, coping, getting medication, social, and financial. The information deficits were in disease-specific areas of natural history and therapy, including conventional and unproven remedies. Physician sources and written materials were preferred. Patients with scleroderma were particularly interested in support groups. The most common worry for patients across diagnostic groups was probably the same as the number one worry of their physicians: progression of the disease. From contrasting studies, it seems that patients with SLE and their physicians have particularly divergent worries.

Patients and physicians rate disease activity in SLE differently. Physicians place greater emphasis on laboratory features, whereas patients place more emphasis on function, particularly the detrimental effects of fatigue.

Data-based recommendations for improving physician-patient communications have been derived partially from taped usual care visits and post hoc interviews with participants (Table 67-2). An ethnographic approach to each individual patient, a method taken from medical anthropology, has generated a protocol for bridging sociocultural gaps between patient and provider. These seven questions, known colloquially as “the Kleinman questions,” focus on determining what the patient feels is at stake in a certain disease situation (Table 67-3). These questions might also be viewed as facilitators for the first four Daltroy techniques presented in Table 67-2, which also aim to determine the patient’s agenda.

Table 67-2  Techniques to Improve Communication in Usual Care Office Visits

| Encourage patients to write down their concerns before each visit. |
| Address each concern specifically, however briefly. |
| Ask patients what they think has caused their problems. |
| Tailor treatments to patients’ goals and preferences as possible. |
| Explain the purpose, dosage, common side effects, and inconveniences, and how to judge the efficacy of each treatment, including length of trial. |
| Assess patients’ understanding. |
| Anticipate problems in compliance with treatment plans, and discuss methods to cope with common problems. |
| Write down diagnosis and treatment plan to help patients remember. |
| Distribute written materials that are now widely available. |
| Reinforce patients’ confidence in their ability to manage their regimen. |
| Use ancillary personnel in patient education. |
| Refer patients to organized programs in the community. |

Table 67-3  Explanatory Models Approach  
(Kleinman’s Ethnographic Approach to Bridge Sociocultural Gaps)

<table>
<thead>
<tr>
<th>Question</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do you call this problem?</td>
<td>How do you know it’s a problem? Other problems you know of?</td>
</tr>
<tr>
<td>What do you believe is the cause?</td>
<td>Other people consider it to be a disease. You consider it...</td>
</tr>
<tr>
<td>What course do you expect it to take? How serious is it?</td>
<td>How serious is the problem? Other people consider it to be a disease.</td>
</tr>
<tr>
<td>What do you think this problem does inside your body?</td>
<td>Other people consider it to be a disease. You consider it...</td>
</tr>
<tr>
<td>How does it affect your body or your mind?</td>
<td>Other people consider it to be a disease. You consider it...</td>
</tr>
<tr>
<td>What do you most fear about this condition?</td>
<td>Other people consider it to be a disease. You consider it...</td>
</tr>
<tr>
<td>What do you fear most about the treatment?</td>
<td>Other people consider it to be a disease. You consider it...</td>
</tr>
</tbody>
</table>


Time Constraints
Time constraints in the context of usual care office visits argue for maximal use of organized programs and for the use of self-completed questionnaires and forms, such as the American College of Rheumatology new patient form. Patients sometimes see forms as an imposition if they are not referred to in the subsequent interview. If the form is part of the conversation and is even marked with a few highlights or marginal notes, it is interpreted as an element of thoroughness. Scheduled telephone contact is poorly reimbursed but is effective enough to reduce knee pain from osteoarthritis in a randomized study. Legal and institutional restrictions notwithstanding, a reasonable rule of thumb for clinical e-mail is to use it for nothing more sensitive than what would be said over the telephone. Overall, the main psychosocial impact of time pressure is to make the interview more physician-directed than patient-directed, putting up an obstacle to physician-patient communication.

Organized Programs
A variety of interventional programs improve outcome measured in their own terms when other psychosocial variables or a health status scale are the primary outcomes. These include support groups, which improve coping more than half the time, and stress management, group psychotherapy, and prayer. A smaller number (see Table 67-1) also save money, improve pain, protect anatomy, and have a durable effect. Exercise programs are effective, but patients have difficulty sustaining the effort these programs require over time.

Psychotherapy and Psychopharmacologic Interventions
Depression often goes unrecognized. With a psychiatric telephone interview in a survey study taken as the “gold standard,” only about 19% of depressed individuals were diagnosed and treated in the course of usual care over 1 year. Treatment was less likely for men, African-Americans, and individuals with lower educational levels. Individuals 30 to 60 years old were more likely to get treatment. Brief psychotherapy and antidepressants are effective. Depression in rheumatic disease occurs more frequently in particular settings, including patients reporting widespread chronic pain, patients coming to grips with work disability, and patients presenting at a young age with knee osteoarthritis.

More recent findings in relevant psychopharmacology include selective serotonin reuptake inhibitor equivalence for efficacy across the class. Recognition of large relative increases in the risk of sudden cardiac death on moderate doses of antipsychotics has also increased.

SUMMARY
The results of patient education and psychosocial intervention seem modest compared with very effective, biologically based interventions such as tumor necrosis factor blockade. The latter costs about $18,000 to $25,000 per year and makes inroads into the lifetime loss of 16 quality-adjusted life-years per the average 27-year course of RA, including 6.6 years of lost survival that were incurred in the absence of biologics and methotrexate. In contrast, the best organized and delivered psychosocial intervention administered on a group basis, the Arthritis Foundation Self-Help Course, has minimal costs and cuts outpatient visits by 40%, saving each osteoarthritis patient and RA patient less than $300 per year.

Meta-analyses and reviews indicate that organized programs aimed at specific subsets of rheumatic disease patients have effects that are favorable, but the effect sizes are small, making them hard to measure, and the effects on outcomes not purely within psychosocial domains are hard to find. Finally, it is difficult to prove effects that are sustained over time. Specific psychosocial management techniques stand as modest adjuncts to powerful specific therapy but must be understood as the middle part of a range of promising opportunities.

Increasing health literacy and improving individual physician-patient communications are effective psychosocial interventions, along with recognition and treatment of concomitant depression. Large-scale social interventions, such as improving the reading level of the patient population where health literacy is lacking, would change disease outcome more than psychosocial interventions aimed at individual diseases. Interpreted broadly over their whole range, including health literacy, psychosocial management techniques are potentially as powerful as biologic therapies. Organized programs applied to defined patient subgroups are the most easily measured and discussed interventions in a scientific sense, however. This thought illustrates French anthropologist-philosopher Rene Dubois’ dictum that “the measurable drives out the important,” which still leaves individual patients and their physicians a fascinating set of possibilities to study and apply, and from which to benefit.

References


The references for this chapter can also be found on www.expertconsult.com.
Nutrition plays a role in the management of most chronic diseases. Physicians provide dietary advice to patients with diabetes, heart disease, and obesity as part of standard clinical care. Although the role of diet and nutrition is well established in the etiology and management of gout, the role of nutrition in other rheumatic diseases such as rheumatoid arthritis (RA) is less well accepted. In general, dietary advice is not part of standard clinical practice for patients with inflammatory rheumatic diseases. Despite a widespread lack of conviction among physicians about the role of nutrition, many people with arthritis believe food plays an important role in their symptom severity and approximately 50% will have tried dietary manipulation in an attempt to improve their symptoms.\(^1\)

It is well recognized that inflammatory disease processes can interfere with nutritional status. More recently, the impact of nutritional factors on the inflammatory response has been recognized. Furthermore, diets and lifestyles have changed considerably through industrialization with significant effects on rates of obesity, food choices, and consumption of different dietary components. In light of this, the role nutrition in the etiology and management of rheumatic diseases is becoming increasingly important.

**Role of Omega-3 Fatty Acids and the Inflammatory Process**

**Fatty Acid Biochemistry**

Fatty acids are divided into three groups on the basis of the number of double bonds they contain (saturated fatty acids (no double bond), monounsaturated fatty acids (one double bond), and polyunsaturated fatty acids (PUFAs) (≥2 double bonds)). C18 fatty acids are prominent in the diet and provide the index fatty acids for the above classification. PUFAs are further grouped according to the site of the double bond proximal to the methyl (omega) terminus as n-6 or n-3. Vertebrates do not have the enzymes required to introduce double bonds in the n-3 and n-6 positions; therefore these fatty acids must be obtained from the diet and are thus known as essential fatty acids.

In general, the Western diet contains more n-6 fats than n-3 fats due to the dominance in processed foods and visible fats of soybean, safflower, sunflower, and corn oils, which contain the n-6 fat linoleic acid (LA; 18:2n-6). The n-3 homologue of LA, α-linolenic acid (ALA; 18:3n-3), is present in flaxseed oil, which is generally a minor dietary component. LA and ALA may be used in energy metabolism or be converted to the C20 fatty acids arachidonic acid (AA) and eicosapentaenoic acid (EPA), respectively. AA and EPA are then incorporated into cell membranes and tissues. EPA can be further metabolized to the n-3 fatty acid docosahexaenoic (DHA).

The critical process linking fatty acids and inflammation is the metabolism of AA and EPA to eicosanoids, which act as inflammatory mediators. AA is metabolized via cyclooxygenase (COX) to n-6 eicosanoids (prostaglandin [PG] E\(_2\), thromboxane [TX] A\(_2\) or via 5-lipoxygenase [5-LOX] to n-6 leukotrienes [LTs]). In comparison, EPA is metabolized via COX and 5-LOX to n-3 PGs and LTs, respectively. In comparison to the n-6 eicosanoids produced from AA, EPA is...
n-3 fatty acids are known as E-resolvins, whereas those derived from n-3 fatty acids via COX-2 with increased proinflammatory actions of cytokine inhibition. For example, fatty acids on eicosanoids, there also appears to be eicosanoid-dependent cytokine inhibition. n-3 fatty acids contribute to inhibition of inflammatory cytokine production.

**Effect of n-3 Fatty Acids on Adhesion Molecule Expression**

Adhesion molecules expressed on endothelial cells and leukocytes mediate the transit of cells from the circulation into tissues. Intercellular adhesion molecule-1 (ICAM-1) and its cognate receptor, leukocyte function-associated antigen (LFA)-1, have been shown to be important in migration of leukocytes into inflamed synovium in animal models. ICAM-1 blockade has also been reported to reduce disease activity in RA. In vitro n-3 fatty acids decrease human monocyte expression of HLA-DR and HLA-DP molecules and reduces the ability of monocytes to present antigen to autologous lymphocytes. Thus n-3 fatty acids may have an anti-inflammatory effect via suppression of pathogenic T cell activation through inhibition of APC function.

**Figure 68-1** Metabolism of linoleic acid (LA) and α-linolenic acid (ALA) to n-6 and n-3 prostaglandins and leukotrienes. AA, arachidonic acid; COX, cyclooxygenase; EPA, eicosapentaenoic acid; IL-1β, interleukin-1β; LOX, lipoxygenase; LTB, leukotriene B; PGE, prostaglandin E; TNF, tumor necrosis factor; TXA, thromboxane A.

Proinflammatory Actions of Eicosanoids

In general the n-6 eicosanoids (PGE$_2$ and TXA$_3$) are proinflammatory, whereas the n-3 eicosanoids are either less potent in their effects (TXA$_3$) or less abundant (PGE$_3$). TXA$_3$ promotes interleukin-1β (IL-1β) and tumor necrosis factor (TNF) production by mononuclear cells, whereas PGE$_3$ results in vasodilatation, increased vascular permeability, and hyperalgesia (see Figure 68-1). PGE$_3$ is edemogenic, although little is produced. LTB$_5$ is 10 to 30 times less potent than LTB$_4$ as a neutrophil chemotaxin.

Effect of n-3 Fatty Acids on Proinflammatory Cytokine Production

IL-1β and TNF production may be reduced as a consequence of dietary n-3 fatty acid supplementation. Although some of this cytokine inhibition is mediated through effects on eicosanoids, there also appears to be eicosanoid-independent cytokine inhibition. For example, fatty acids may have direct effects on intracellular signaling mechanisms including nuclear factor κB (NFκB) and PPAR-γ, thereby affecting cytokine production.

The resolvins (resolution phase interaction products) are derived from n-3 fatty acids via COX-2 with increased production in the presence of aspirin. Resolvins derived from EPA are known as E-resolvins, whereas those derived from DHA are known as D-resolvins. The resolvins have a variety of anti-inflammatory actions including inhibition of TNF-induced transcription of IL-1β and inhibition of human polymorphonuclear leukocyte transendothelial migration (for review, see Kohli and Levy's study). The identification of resolvins provides another mechanism through which n-3 fatty acids contribute to inhibition of inflammatory cytokine production.

Effects of n-3 Fatty Acids on MHC Expression

The number of MHC molecules expressed on antigen-presenting cells (APCs) is an important determinant of T cell response to antigen. Patients with RA have high levels of MHC class II expression on T cells and synovial lining cells. In vitro studies show that EPA and/or DHA reduces monocyte expression of HLA-DR and HLA-DP molecules and reduces the ability of monocytes to present antigen to autologous lymphocytes. Thus n-3 fatty acids may have an anti-inflammatory effect via suppression of pathogenic T cell activation through inhibition of APC function.
Apoptosis, thereby inhibiting DC-dependent T cell activation. In addition, 1,25(OH)₂D₃ can promote DC expression of tolerizing functions, which instruct T regulatory (Treg) cells, which in turn may inhibit the development of autoimmunity. Vitamin D inhibits monocyte/macrophage proinflammatory cytokine production including TNF, IL-6, and IL-1α. Vitamin D has direct effects on T cells, in particular inhibition of proliferation and cytokine production by Th1 cells, and may enhance Th2 cytokine production. 1,25(OH)₂D₃ has also been shown to reduce Th17 cell differentiation through its effects on DCs, as well as direct effects on Th17 cells leading to reduced IL-17A production. 1,25(OH)₂D₃ inhibits the proliferation of activated B cells, induces activated B cell apoptosis, and inhibits plasma cell differentiation and immunoglobulin secretion. Thus vitamin D deficiency may have a role in the etiology of B cell–mediated autoimmune disorders, whereas vitamin D supplementation may have beneficial effects in B cell–mediated autoimmune diseases such as SLE and RA.

The array of relevant effects of vitamin D on the immune system suggests that vitamin D insufficiency may have an important role in the etiology and management of rheumatic diseases more generally.

**Effect of n-3 Fatty Acids on Degradative Enzymes**

Proteinases have a pivotal role in cartilage degradation and bone erosion. n-3 Fatty acids added in vitro can suppress proteinases ADAMTS-4, ADAMTS-5, and MMP-3 in IL-1α-stimulated bovine chondrocytes. This inhibition of chondrocyte proteases is a mechanism through which n-3 fatty acids may inhibit cartilage degradation and bone erosion.

The RANK/RANKL/OPG pathway is also important in bone pathophysiology in RA. Increased RANK/RANKL and decreased OPG contribute to bone erosion. Three months of dietary fish oil supplementation has been reported to decrease the RANK/OPG ratio, which may help prevent bone resorption that leads to erosions.

**Importance of the Balance of n-3 and n-6 Fatty Acids in the Inflammatory Process**

The balance of AA and EPA can be altered through dietary fatty acid intake. In humans, the conversion of dietary ALA to tissue EPA is inefficient and fish/fish oils are a more effective way to increase EPA and DHA in tissues. Changes in AA/EPA ratios in tissues have downstream effects on eicosanoid production and the resulting proinflammatory/anti-inflammatory environment. Dietary supplementation with fish oil in humans results in decreased production of PGE₂, TXA₂, and LTβ with increased production of PGI₂ and LTβ. These data provide a mechanistic basis for beneficial effects of dietary n-3 fatty acid supplementation in the control of inflammatory diseases. Dietary fish oil supplements have been shown to increase vascular production of prostacyclin (PGI₂). Although the role of PGI₂ in inflammation is not well defined, it is a potent vasodilator and inhibits platelet aggregation, as well as disaggregating platelets. These effects likely contribute to the protective effects of dietary fish and fish oils against thrombotic vascular events. Importantly, patients with several of the major rheumatic diseases (e.g., RA, systemic lupus erythematosus [SLE] and gout) are at high risks for serious cardiovascular events and mortality, to which nonsteroidal anti-inflammatory drug (NSAID)–associated COX-2 inhibition may also contribute.

**Vitamin D and the Inflammatory Process**

Vitamin D has multiple immunosuppressive effects in addition to its effects on bone and calcium metabolism. The biologically active form of vitamin D (1,25-dihydroxyvitamin [OH]₂D₃) interacts with vitamin D receptors, which are expressed on a variety of cells including osteoblasts, T cells, dendritic cells (DCs), macrophages, and B cells. These cells also have the capacity to convert more abundant 25(OH)D₃ to 1,25(OH)₂D₃ and to degrade 1,25(OH)₂D₃. Accordingly, the molecular machinery is in place for important autocrine and paracrine effects of vitamin D at sites of inflammation.

DCs have a central role in activation of the immune system and in response to self. 1,25(OH)₂D₃ inhibits the differentiation of monocyte precursors into mature DCs, downregulates expression of MHC class II molecules on DCs, inhibits IL-12 production, and promotes DC apoptosis, thereby inhibiting DC-dependent T cell activation. In addition, 1,25(OH)₂D₃ can promote DC expression of tolerizing functions, which instruct T regulatory (Treg) cells, which in turn may inhibit the development of autoimmunity. Vitamin D inhibits monocyte/macrophage proinflammatory cytokine production including TNF, IL-6, and IL-1α. Vitamin D has direct effects on T cells, in particular inhibition of proliferation and cytokine production by Th1 cells, and may enhance Th2 cytokine production. 1,25(OH)₂D₃ has also been shown to reduce Th17 cell differentiation through its effects on DCs, as well as direct effects on Th17 cells leading to reduced IL-17A production. 1,25(OH)₂D₃ inhibits the proliferation of activated B cells, induces activated B cell apoptosis, and inhibits plasma cell differentiation and immunoglobulin secretion. Thus vitamin D deficiency may have a role in the etiology of B cell–mediated autoimmune disorders, whereas vitamin D supplementation may have beneficial effects in B cell–mediated autoimmune diseases such as SLE and RA.

The array of relevant effects of vitamin D on the immune system suggests that vitamin D insufficiency may have an important role in the etiology and management of rheumatic diseases more generally.

**Reactive Oxygen Species/Antioxidants and the Inflammatory Process**

Production of reactive oxygen species (ROS) such as superoxide and hydrogen peroxide are part of the normal immune response. ROS acting through transcription factors such as NFκB increase production of proinflammatory eicosanoids and cytokines including PGE₂, TNF, and IL-1β. Thus unchecked production of ROS may cause inflammation and tissue damage. Antioxidant enzymes such as superoxide dismutase and glutathione peroxidase remove superoxide, thus providing protection from oxidative damage. Vitamin C (ascorbic acid), vitamin E (α-tocopherol), and β-carotene are acquired through the diet and can act as ROS scavengers.

**Obesity and the Inflammatory Process**

With excess energy intake and reduced energy expenditure, body weight and adiposity increase. Obesity (body mass index [BMI] > 30 kg/m²) is a significant health problem globally with the World Health Organization estimating that by 2015 approximately 2.3 billion adults will be overweight and more than 700 million will be obese.

Adipose tissue was originally thought to be simply a fat store. However, it is now recognized that adipose tissue and adipocytes are metabolically active and contribute to systemic inflammatory responses (Figure 68-2). Adipocytes release the proinflammatory cytokines TNF, IL-1β, and IL-6. IL-6 enters the systemic circulation and increases C-reactive protein and serum amyloid A production by the liver.

Adipocytes produce adipokines: leptin, resistin, and visfatin (proinflammatory); and adiponectin (anti-inflammatory) (see Figure 68-2). Although the primary function of leptin is appetite control, it also has a number of proinflammatory actions. Leptin increases the expression of adhesion molecules such as ICAM-1 and monocyte...
lymphocytes and adipocytes, has similar proinflammatory effects including induction of IL-8, IL-6, IL-1β, and TNF; increased endothelial expression of ICAM-1, VCAM-1, and MMPs; and enhancement of B cell differentiation. The net result of obesity is thus an inflammatory state associated with an increase in circulating C-reactive protein.

**Figure 68-2** Inflammatory mechanisms of obesity. BMI, body mass index; ICAM, intercellular adhesion molecule; IL, interleukin; MCP-1, monocyte chemotactic protein; MMPs, matrix metalloproteinases; NFκB, nuclear factor κB; Th, T helper; TNF, tumor necrosis factor; VCAM-1, vascular cell adhesion molecule-1.

Probiotics are defined as live microorganisms, which confer a health benefit to the host. The most common probiotics are lactobacilli and bifidobacteria. These organisms are present in foods, commonly dairy fermented products such as yogurt, as well as in supplements such as probiotic drinks in lyophilized form. Probiotics have anti-inflammatory properties, with the mechanism of anti-inflammatory action varying between different probiotics species and different
species strains. Thus although certain probiotic strains may confer health benefits, these benefits may not be seen with other probiotic species or strains.

Probiotics may exert their putative anti-inflammatory effects by interacting with intestinal epithelial cells. Resultant modulation of intestinal microflora and fortification of intestinal barrier function may lead to altered exposure of the immune system to microbes and direct effects on immune cells within the intestine. Associated effects relevant to anti-inflammatory responses include accumulation of CD4+ Treg cells in inflamed areas; inhibition of DC activation; antagonism of NFkB, which leads to decreased IL-1β and TNF production; and induction of the inflammation-modulating cytokine TGF-β.

Summary

Dietary components can have a variety of effects on the inflammatory process and on bone destructive pathways (Table 68-1). Thus alterations in diet may have effects on both the risk and management of rheumatic diseases, as discussed later.

### Table 68-1  Effects of Dietary Components on Inflammatory and Bone Destructive Pathways

<table>
<thead>
<tr>
<th>n-3 Fatty Acids</th>
<th>Decreased production of n-6 derived eicosanoids (PGE₂, TXA₂, LTB₄), which have proinflammatory effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Increased production of n-3-derived eicosanoids (PGE₃, TXA₃, LTB₃), which in general are less proinflammatory</td>
</tr>
<tr>
<td></td>
<td>Decreased IL-1β and TNF production</td>
</tr>
<tr>
<td></td>
<td>Increased production of resolvins</td>
</tr>
<tr>
<td></td>
<td>Decreased major histocompatibility complex II expression by antigen-presenting cells</td>
</tr>
<tr>
<td></td>
<td>Decreased adhesion molecule expression: ICAM, VCAM, LFA</td>
</tr>
<tr>
<td></td>
<td>Decreased expression of matrix metalloproteinases</td>
</tr>
<tr>
<td></td>
<td>Alteration in RANK/OPG ratio</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vitamin D</th>
<th>Inhibition of monocyte differentiation into DCs and promotion of DC apoptosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Induction of tolerogenic Tregs with enhanced suppressive activity</td>
</tr>
<tr>
<td></td>
<td>Inhibition of monocyte/macrophage IL-1β and TNF production</td>
</tr>
<tr>
<td></td>
<td>Inhibition of Th1 cell proliferation and cytokine production</td>
</tr>
<tr>
<td></td>
<td>Enhancement of Th2 cytokine production</td>
</tr>
<tr>
<td></td>
<td>Reduction in Th1 cell differentiation and IL-17A production</td>
</tr>
<tr>
<td></td>
<td>Inhibition of proliferation of activated B cells, differentiation of activated B cells and immunoglobulin secretion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Antioxidants</th>
<th>Scavenge reactive oxygen species</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reduce proinflammatory eicosanoids, TNF and IL-1β</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probiotics</th>
<th>Generation of CD4+ Tregs in inflamed areas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inhibition of DC activation</td>
</tr>
<tr>
<td></td>
<td>Decreased IL-1β and TNF production via inhibition of nuclear factor κB</td>
</tr>
<tr>
<td></td>
<td>Increased production of transforming growth factor-β</td>
</tr>
</tbody>
</table>

A number of epidemiologic studies have examined the role of nutrition in the etiology of RA. Assessing dietary intake in epidemiologic studies is difficult, and identifying the effects of a single dietary variable and distinguishing it from other nutritional and lifestyle factors may not be possible. Notwithstanding, a number of nutritional factors are candidates for an association with the development of RA.

### Omega-3 Fatty Acid Consumption

The long chain n-3 fats, EPA and DHA, are most abundant in fish and fish oils. A protective effect against RA of fish intake has been reported. For example, the Seattle Women’s Health Study showed reduced risk for developing RA in subjects consuming two or more fish meals per week with an adjusted odds ratio (OR) of 0.57 (95% confidence interval [CI], 0.35 to 0.93) compared with subjects consuming less than one fish meal per week. A more recent population-based, case-controlled study reported a modest decrease in the risk of RA in subjects who consume oily fish one to seven times per week compared with those who seldom or never consumed fish (OR, 0.8; 95% CI, 0.6 to 1.0), which did not change even allowing for RF and anticyclic citrullinated protein (CCP) antibody status.

### Red Meat and Protein Consumption

High consumption of red meat has been associated with an increased risk of inflammatory polyarthritis (OR, 1.9; 95% CI, 0.9 to 4.0). Although one study reported that meat and offal were associated with an increased risk of developing RA, this was not confirmed in other studies. Whether the association between red meat consumption and inflammatory arthritis is causative remains unclear, although the presence of significant amounts of AA in red meat may provide some explanation for the association.

### Tea and Coffee Consumption

Tea and coffee have been identified as potential risk factors for the development of RA. In the Finnish National Health Study, consumption of four or more cups of coffee per day was associated with an increased risk of RF-positive, but not RF-negative, RA after adjustment for potential confounders such as age, smoking, and gender (relative risk [RR], 2.2; 95% CI 1.13 to 4.27). By contrast, the Iowa Women’s
Health Study reported no association between daily caffeine intake and risk for RA. However, women who consumed four or more cups of decaffeinated coffee per day were at increased risk for RA compared to non–coffee drinkers (RR, 2.58; 95% CI, 1.63 to 4.06). Furthermore, women who consumed three or more cups of tea per day had a reduced risk of RA (RR, 0.39; 95% CI, 0.16 to 0.97). More recent studies have not shown an association between tea/coffee and RA.31,35

Although the epidemiologic evidence for an association between coffee/tea and RA is equivocal, interest has been sustained by the presence of metabolically active agents in these beverages. For example, the major catechin in tea inhibits induction of inducible nitric oxide synthase (iNOS) by stimulated macrophages.36 iNOS generates highly reactive free radical products while di-imidating substrate arginine moieties to citrulline. Citrullinated peptides/proteins are recognized immunogens that provide a focus for autoimmunity in RA.

Alcohol Consumption

Alcohol consumption may reduce the risk of developing RA. In a case-control study of 515 patients with RA, alcohol consumption was associated with a reduced risk of anti-CCP positive RA.33 A dose-dependent inverse relationship between alcohol consumption and risk of RA has also been demonstrated from two independent case-control studies (Swedish EIRA and Danish CACORA).38 The reduction in risk of RA was more pronounced in patients with the shared epitope compared those without the shared epitope and most pronounced in smokers with the shared epitope.38

With regard to candidate mechanisms for this putative reduction in risk for RA, alcohol has been shown to down-regulate the production of proinflammatory cytokines and upregulate production of the anti-inflammatory cytokine IL-10.39,40 Furthermore, in a murine model of arthritis, ethanol almost totally prevented the development of collagen-induced arthritis and in those mice that did develop arthritis, disease was less severe. These anti-inflammatory effects of ethanol were associated with reduced leukocyte migration, downregulation of NFkB, and reduced production of the proinflammatory cytokines IL-6 and TNF but not the anti-inflammatory cytokine IL-10.41

Vitamin D

As noted earlier, vitamin D has anti-inflammatory actions. Although vitamin D has been implicated in reducing the risk of the autoimmune diseases, diabetes,42 and multiple sclerosis, the association with risk of RA is less clear.43

The Iowa Women’s Health Study reported that a higher intake of vitamin D was associated with a reduced risk of RA (RR, 0.67; 95% CI, 0.44 to -1.00, P = 0.05) in women aged 55 to 69 years.44 However, in a more recent large study of 186,389 women followed for 22 years, there was no association between dietary intake of vitamin D and the risk of developing RA.45 However, apart from supplements, the main source of vitamin D is de novo synthesis in the skin and estimated dietary intake may be a poor predictor of serum vitamin D concentrations. In a study of 79 patients with RA, no association was found between prior serum vitamin D concentrations and subsequent development of RA.46 However, it is notable that the geometric means for both cases and controls in this study were only half the lower reference level of 60 nmol/L, which has been set subsequently to reflect a level that suppresses secondary hyperparathyroidism due to vitamin D insufficiency.

Antioxidants and Risk of Rheumatoid Arthritis

Oxygen free radicals (e.g., nitric oxide, superoxide, hydroxyl radical) are implicated in the tissue damage observed in RA.47 Antioxidants including vitamin E (α-tocopherol), vitamin C (ascorbic acid), β-carotene, and selenium may have a protective role against tissue damage caused by these oxygen free radicals. This combined with evidence that markers of antioxidant nutritional status are lower in patients with established RA compared with normal controls48 led to the hypothesis that antioxidants may protect against the development of RA. Despite this biologic plausibility, the available data do not provide clear evidence for a protective effect of antioxidants as dietary supplements in relation to the development of RA.

Low serum concentrations of vitamin E, β-carotene, retinol, and selenium have been reported to be weakly associated with an increased risk for developing RA in some but not all studies.49-51 The strongest association between risk of RA and antioxidants was found for a combined antioxidant index rather than any one specific antioxidant.49

A higher dietary intake of β-cryptoxanthin (a carotenoid found in fruit and vegetables) and zinc may protect against the development of RA.52,53 Low vitamin C intake has also been associated with an increased risk of inflammatory polyarthritis with an adjusted OR of 3.3 (95% CI, 1.4 to 7.9) for the lowest tertile of vitamin C intake (<55.7 mg/day) compared with the highest intake tertile (>94.9 mg/day).54 However, another study found no association between intake of vitamin C, vitamin E, zinc, or selenium and the development of RA.31

The Women’s Health Study, a randomized, double-blind, placebo-controlled trial designed to evaluate the effects of low-dose aspirin and vitamin E in the primary prevention of cardiovascular disease and cancer, has been used to address the influence of vitamin E supplementation (600 IU on alternate days) on the development of RA. During the 10-year follow-up period, 106 cases of RA occurred with no significant association between vitamin E supplementation and incidence of RA. A statistically insignificant trend toward inverse association between vitamin E and development of seropositive RA was seen on subgroup analysis.55

Obesity and Rheumatoid Arthritis

Two studies have reported that obesity increases the risk of RA,56,57 whereas two other studies report no association.58-60 Increased plasma concentrations of the adipokines leptin, adiponectin, and visfatin have been observed in patients with RA compared with healthy controls.61 Furthermore, visfatin and leptin were associated with increased and reduced radiographic joint damage, respectively.61 These data combined with the proinflammatory state observed in obesity suggest that BMI may affect disease activity and outcomes in RA, whereas, perhaps paradoxically, increased
BMI has been associated with less radiographic damage. An alternative explanation is that patients with more active inflammatory disease, who tend to develop more radiographic erosions, may have lower BMI and rheumatoid cachexia.

**Gout**

### KEY POINTS

- The link between diet and gout has been recognized for centuries.
- High intake of meat, seafood, and alcohol is associated with increased risk of gout.
- Fructose increases serum urate.
- Higher intake of low-fat dairy products has been associated with a reduced risk of gout.
- Increased BMI predisposes to gout.

For centuries gout has been associated with overindulgence in rich food and wine. Gout occurs when serum urate (SU) concentrations reach super-saturation concentrations (=6.8 mg/dL at 37° C), resulting in the formation of monosodium urate crystals, which deposit in joints and soft tissues. Urate is the end product in the breakdown of purines, which are a product of cellular turnover or are ingested in the diet. The purines adenosine and guanine are present in nucleic acids and the intracellular energy transporters adenosine triphosphate and guanosine triphosphate. Accordingly, foods derived from metabolically active animal tissues may increase dietary purine load.

**Dietary Factors and Gout**

Several large clinical studies have identified an association between high intake of meat and seafood (but not total protein intake) and both SU concentrations and gout. In comparison, the risk of gout is reduced with higher intake of low-fat dairy products and long-term coffee consumption. More recently, fructose, which is found in corn syrup, sugar-sweetened soft drinks, and fruit juices, has been associated with hyperuricemia and gout. Fructose increases SU through increased purine degradation. Furthermore, urate and fructose share a common transporter within the kidney (SLC2A9). Alcohol consumption other than moderate wine intake is associated with an increased risk of gout, with beer conferring a higher risk than liquor.

Dietary vitamin C supplementation has been shown to reduce the risk of gout (relative risk of gout with no supplementation vs. 1000 to 1499 mg vitamin C/day, 0.66 [95% CI, 0.49 to 0.88]). Vitamin C (ascorbic acid) is an important vitamin that can only be obtained through dietary intake. Because ascorbic acid is water soluble, it is not stored within the body and thus must be regularly supplemented through the diet to maintain the ascorbic acid pool. Dietary sources of ascorbic acid include fresh fruits and vegetables, in particular citrus fruits and green leafy vegetables such as broccoli.

**Fasting and Gout**

Prolonged periods (2 weeks to 8 months) of fasting in obese subjects result in a significant increase in SU and in some cases the development of gout. In patients with previously documented gout, a 1-day fast led to an increase in SU of 0.5 to 2.1 mg/dL with a mean rise of 1.1 mg/dL, with refeeding associated with a return in SU to baseline levels after 24 hours. Explanations for the increase in SU during fasting include increased urate production, decreased excretion due to decreased glomerular filtration rate, altered renal tubular transport of uric acid, and competition with ketones for renal tubular excretion.

**Obesity and Gout**

High BMI predisposes to gout. In a study comparing obese men (mean BMI 34 ± 4 kg/m²) with healthy controls (BMI, 21 ± 1 kg/m²), SU concentrations were raised to a similar degree in the obese subjects (=8.0 ± 1.6 mg/dL vs. controls, 5.2 ± 0.81 mg/dL), irrespective of body fat distribution (predominantly visceral or predominantly subcutaneous). By contrast, 80% of subjects with hyperuricemia and accumulation of fat subcutaneously had low 24-hour urinary urate excretion compared with 10% of their counterparts with fat accumulation predominantly in a visceral distribution. These data suggest that the mechanism of hyperuricemia may vary depending on body fat distribution. In a separate study using computed tomography to determine transverse area of abdominal fat components, visceral fat was shown to correlate strongly with SU, whereas no relationship was seen between SU and BMI or subcutaneous fat area.

**Osteoarthritis**

### KEY POINTS

- Obesity is associated with knee osteoarthritis (OA).
- Direct biomechanical effects of obesity contribute to OA.
- Increased leptin provides another link between obesity and OA.

Associations have been established between obesity and onset and progression of knee OA. The evidence for an association between obesity and OA of the hip or hand is less well defined. In addition to direct biomechanical effects of obesity on the joint, recent data suggest a role for the adipocytokine leptin. Expression of leptin was increased in advanced OA cartilage compared with minimally damaged OA cartilage, with the leptin mRNA expression in advanced OA cartilage correlating with BMI. Furthermore, leptin was shown to reduce chondrocyte proliferation and increase IL-1β, MMP-9, and MMP-13 expression.

Dietary antioxidants (vitamin E, vitamin C, and β-carotene) have been reported to reduce progression of knee OA but have no effect on onset of OA. The association between serum vitamin D concentrations and OA remains unclear, with one study reporting no association between vitamin D and risk of joint space narrowing or cartilage loss in knee OA and another study reporting a
positive association between knee cartilage volume and serum vitamin D concentrations.\textsuperscript{32}

**NUTRITION IN THE MANAGEMENT OF RHEUMATIC DISEASES**

The role of nutrition in the management of gout is well accepted. However, its role in other rheumatic diseases such as RA is less routine. Despite a relative lack of interest or emphasis from physicians, many patients consider food may contribute to their arthritis and seek information about or try putative dietary remedies. A convincing, informed approach to nutritional advice is thus an important aspect of management, which can help patients avoid worthless interventions that are expensive, time consuming, in some cases harmful, and diversions from access to more effective measures (summarized in Figure 68-3). Furthermore, positive advice regarding dietary choices may empower the patient at a time when there is often a sense of loss of control. Well-informed, authoritative dietary advice can also protect patients from poorly grounded advice from relatives, friends, and nonauthoritative Internet sites about dietary measures.

### Rheumatoid Arthritis

**KEY POINTS**

- Omega-3 fatty acids reduce disease activity and NSAID requirement.
- There is no evidence for benefit of antioxidants in the management of RA.
- Fasting, vegetarian/vegan, and elimination diets are difficult to sustain, and it is difficult to predict which patients may respond.

![Figure 68-3](summary_of_nutrition_in_rheumatic_diseases_bmi_body_mass_index_acpa_anticitrullinated_protein_antibody_cvd_cardiovascular_disease_mtx_methotrexate NSAID_nonsteroidal_anthipeflammatory_drug_RA_rheumatoid_arthritis_SU_serumurate.png)

**Figure 68-3** Summary of nutrition in rheumatic diseases. BMI, body mass index; ACPA, anticitrullinated protein antibody; CVD, cardiovascular disease; MTX, methotrexate; NSAID, nonsteroidal anti-inflammatory drug; RA, rheumatoid arthritis; SU, serum urate.
The vascular benefits of omega-3 fatty acids are important in RA patients because they have an increased risk of cardiovascular disease.

Some dietary variables may interact with methotrexate.

### Dietary n-3 Fatty Acids in the Management of Rheumatoid Arthritis

Fish oil is a rich source of the anti-inflammatory long-chain n-3 fatty acids EPA and DHA. The threshold dose of EPA and DHA usually required to obtain an anti-inflammatory effect is 2.7 g/day of EPA plus DHA, which is the equivalent of 9 or more standard 1-g fish oil capsules or 10 mL of bottled fish oil per day. This dose is generally more than that with which patients will self-prescribe, although community awareness of required doses varies. Tissue levels of EPA and DHA from fish and fish oils are increased when dietary n-6 fatty acid intake is reduced concomitantly through substitution of n-6 rich visible fats (e.g., with a base of corn oil, soy oil, sunflower oil) with unsaturated products with less n-6 fat (with a base of olive oil, rapeseed/canola oil, or flaxseed oil).83

In general, patients with recent-onset or more established RA can be expected to achieve better disease control when anti-inflammatory doses of fish oil are co-administered with appropriately intensive combination disease-modifying antirheumatic drug (DMARD) therapy. In a longitudinal cohort study of patients with recent-onset RA (<12 months’ duration), response-driven, intensive combination DMARD therapy was combined with either supplemental fish oil or placebo. At 3 years, those patients compliant with fish oil therapy had improved self-reported function in activities of daily living, lower tender joint counts, lower erythrocyte sedimentation rate (ESR), higher remission rates (72% vs. 31%), and less NSAID use than those who did not consume fish oil.84 The better outcome for patients taking fish oil is consistent with randomized controlled trial data showing reduced symptoms with fish oil compared with placebo treatment. This study also demonstrates the feasibility of long-term fish oil use, in this case more than 3 years. Although fish oil usage may identify more compliant patients, overall DMARD usage was similar in the two groups.

Anti-inflammatory doses of fish oil have been shown to reduce NSAID requirements in patients with RA in other studies.85,86 NSAIDs alter the balance of PGE₂/TXA₂ in favor of TXA₂, thereby increasing the production of IL-1β and TNF by monocytes.87 By contrast, fish oil has been shown to reduce the production of these cytokines.12 Fish oil also lacks many of the adverse effects associated with NSAIDs (Table 68-2). The direct effects of fish oil and their influence on NSAID use may potentially reduce long-term tissue damage associated with release of these cytokines, although this remains to be established.

Important for both patients and physicians to recognize is that, as with most standard DMARDs, there is a latent period of up to 15 weeks before the symptomatic benefits of anti-inflammatory doses of fish oil are experienced. One small pilot study has shown that this latent period can be reduced with intravenous administration of n-3 fats.88 Although the inconvenience and cost of intravenous therapy is a barrier to this approach, the benefits of short-term intravenous therapy can be prolonged by a switch to oral n-3 supplementation.89

### Adverse Effects of Fish Oils

The most common adverse effects associated with fish oil in anti-inflammatory doses are a fishy aftertaste, gastrointestinal upset, and nausea. These side effects are clearly neither organ nor life-threatening, but they can be dose limiting. Patient preferences vary, but in general bottled fish oil layered on juice is the most efficient way of taking an anti-inflammatory dose (one quick swallow compared with 10 to 15 standard fish oil capsules). Fish oil is best tolerated with food and not on an empty stomach. The term fish oil defines oil prepared from fish bodies in order to provide a distinction from cod liver oil, which is prepared from fish livers and is rich in the fat-soluble vitamins A and D. Standard fish oil contains more EPA plus DHA (30% w/w) than cod liver oil (EPA plus DHA =20%) and is the preferred material. Vitamin D can be given separately as needed, and vitamin A supplementation is best avoided because negative effects on bone mineral density and fracture risk have been reported.90

Although anti-inflammatory doses of fish oil have not been associated with serious toxicity, concerns may be raised. One consideration is a putative bleeding tendency associated with therapeutic fish oil ingestion. This notion has its origins in studies of Greenland Eskimos consuming their aboriginal diet, in whom myocardial infarction is rare and bleeding times were prolonged relative to those observed in Danes.91 However, these concerns need to be placed within the context of the high amounts of EPA and DHA in the aboriginal Eskimo diet (more than double the threshold anti-inflammatory dose) and the lower amounts of antagonistic n-6 fatty acids compared with Western diets. Although bleeding episodes have not been a feature of long-term fish oil therapy for RA54 and platelet EPA is fourfold

### Table 68-2 Comparison between NSAIDs and Anti-inflammatory Doses of Fish Oil

<table>
<thead>
<tr>
<th></th>
<th>NSAIDs</th>
<th>Fish Oil</th>
</tr>
</thead>
<tbody>
<tr>
<td>COX inhibition</td>
<td>COX-1/COX-2 selectivity varies depending on agents</td>
<td>Nonselective</td>
</tr>
<tr>
<td>NSAID sparing</td>
<td>No</td>
<td>Yes (↓ prostaglandin E₂)</td>
</tr>
<tr>
<td>Serious cardiovascular events</td>
<td>Increased (except naproxen)</td>
<td>Reduced</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Increased</td>
<td>Reduced</td>
</tr>
<tr>
<td>TNF and IL-1β</td>
<td>Increased</td>
<td>Reduced</td>
</tr>
<tr>
<td>Upper gastrointestinal bleeding</td>
<td>Increased</td>
<td>Not reported</td>
</tr>
<tr>
<td>Mortality</td>
<td>Increased</td>
<td>Reduced (especially sudden cardiac death)</td>
</tr>
<tr>
<td>Time to effect</td>
<td>Prompt</td>
<td>Delayed (≤3 mo)</td>
</tr>
</tbody>
</table>

COX, cyclooxygenase; IL-1β, interleukin-1β; NSAID, nonsteroidal anti-inflammatory drug; TNF, tumor necrosis factor.
lower than found in the Eskimos, surgeons may ask that fish oil be discontinued before elective surgery. Also, pharmacists may advise patients to discontinue fish oil while taking warfarin, even though no increase in bleeding tendency has been seen in patients undergoing cardiac surgery when taking fish oil concomitantly with either warfarin or aspirin therapy.\(^9\)

Another concern is the possible presence of the environmental contaminants methylmercury, polychlorinated biphenyls (PCBs), and dioxins, which are concentrated in large carnivorous fish. These toxins are excluded from properly prepared fish oil for therapeutic use. The FDA has accorded “generally regarded as safe” status to intakes of up to 3 g/day of long-chain n-3 fatty acids from marine sources.

**Role of Omega-3 Supplementation in Rheumatic Diseases with Increased Cardiovascular Disease Risk.** A number of rheumatic diseases including RA, SLE, gout, and psoriatic arthritis are associated with increased cardiovascular mortality. n-3 Fatty acid supplementation has been shown to be beneficial in the primary and secondary prevention of ischemic heart disease, as well as in heart failure.\(^7\) The potential mechanisms by which n-3 fatty acids reduce cardiovascular risk include stabilization of the myocardium leading to reduction in cardiac arrhythmias, reduced blood pressure, stabilization of atheromatous plaques, reduced triglycerides and increased high-density lipoprotein (HDL), decreased platelet thromboxane release, increased vascular prostacyclin release, and anti-inflammatory effects (Figure 68-4).

To date there have been no specific studies of the cardiovascular benefits of n-3 fatty acid supplementation in patients with RA or other rheumatic diseases. However, patients with early RA receiving fish oil have been shown to have lower triglycerides, increased “good” HDL cholesterol, less NSAID use, greater disease suppression, and reduced platelet synthesis of TXA\(_2\), all of which can be expected to reduce cardiovascular risk.\(^8\)

**Antioxidants in the Management of Rheumatoid Arthritis**

Despite the association between ROS and RA, a recent meta-analysis shows no convincing evidence that antioxidant supplementation improves disease control in RA.\(^9\)

**Vitamin E.** Serum concentrations of vitamin E are similar in patients with RA and healthy controls.\(^9\) A 12-week placebo-controlled study of vitamin E supplementation reported a reduction in pain scores but no effect on joint tenderness score, duration of morning stiffness, swollen joint count, or laboratory parameters.\(^10\)

**Vitamin C.** Although animal models have shown benefits with vitamin C supplementation, human studies have not shown any clinical benefit in patients with RA.\(^11\)

**Selenium.** Although not an antioxidant per se, selenium is found at the active site of the enzyme glutathione peroxidase, an important antioxidant enzyme. Plasma concentrations of selenium are reduced in patients with RA compared with healthy controls,\(^12\) and an inverse association between serum selenium concentrations and the number of active joints has been reported.\(^13\) Notwithstanding, studies of selenium supplementation have not shown clinical benefit in RA, despite increases in serum and red blood cell selenium concentrations.\(^14,15\) However, polymorphonuclear cell selenium concentrations do not increase.

---

**Figure 68-4** Effects of omega-3 fatty acids on the cardiovascular system. HDL, high-density lipoprotein; IL-1, interleukin-1; LDL, low-density lipoprotein; TNF, tumor necrosis factor; TXA\(_2\), thromboxane A\(_2\); VLDL, very low density lipoprotein.
with dietary supplementation, which may explain the lack of clinical effect.\textsuperscript{102}

**Vitamin D and the Management of Rheumatoid Arthritis**

Animal models have shown that 1,25(OH)\(_2\)D can prevent and have therapeutic benefits in experimental arthritis.\textsuperscript{103} Studies examining the relationship between serum vitamin D concentrations and disease activity in RA have been conflicting with both an inverse relationship\textsuperscript{104} and no relationship\textsuperscript{105} observed.

In humans the vitamin D receptor has been shown to be expressed at sites of cartilage erosion and in chondrocytes and synoviocytes in patients with RA but not normal controls, suggesting that vitamin D may have local effects within the inflamed joint.\textsuperscript{106} Supplemental vitamin D is usually prescribed to patients with RA for treatment or prophylaxis of osteoporosis. However, a 12-month trial of vitamin D (calciferol) 100,000 IU/day showed improvement in RA disease activity and reduction in the requirement for analgesics and NSAIDs.\textsuperscript{107} In another small study of 19 patients with RA, oral \(\alpha\)-calcitriol 2 \(\mu\)g/day was reported to have a beneficial effect on disease activity compared with placebo.\textsuperscript{108} There were neither significant side effects nor increases in serum calcium in either of these studies.

It is important to recognize that cod liver oil, a rich source of dietary n-3 fatty acids, is also rich in vitamin D. Cod liver oil has been used in the management of RA in sufficient doses to deliver anti-inflammatory doses of long chain n-3 fatty acids.\textsuperscript{109} Although serum vitamin D concentrations rose significantly, the study did not allow for the respective contributions of vitamin D and co-ingested n-3 fatty acids to be evaluated. The relatively high content of vitamin A in cod liver oil militates against its use for delivering anti-inflammatory doses of n-3 fatty acids.

**Dietary Restriction in Rheumatoid Arthritis—Fasting, Vegetarian, and Elimination Diets**

A number of studies have examined elimination of different nutrients from the diet in the management of RA. At the extreme, total fasting and subtotal fasting reduced clinical and laboratory parameters of disease activity in some patients within a few days, but deterioration occurred on reintroduction of food.\textsuperscript{110} Thus fasting is an impractical management strategy because it should last no longer than 7 days and the positive effects are short-lived, whereas RA is a chronic disease. A number of potential mechanisms for this improvement have been postulated including a psychologic or placebo effect, weight loss, and immunosuppression as a result of reduced caloric intake, alteration in fatty acids, and IL-6 concentrations and alterations in gut flora.

Elemental diets aim to provide major food groups as simpler components (e.g., protein as free amino acids, fat as medium chain triglycerides, carbohydrates as small sugars). Minimal general benefit was seen in only a minority of patients with lack of benefit observed in objective parameters such as inflammatory markers, swollen, and tender joint counts.\textsuperscript{111,112}

Vegetarian and vegan diets have also been trialed in RA. Response to these diets has been variable with poor compliance and high dropout rates due to both lack of efficacy and adverse effects (mainly nausea and vomiting).\textsuperscript{113,114} More recently, a meta-analysis of four clinical trials of fasting, followed by vegetarian diets which lasted at least 3 months, showed clinically significant long-term benefits in RA can occur.\textsuperscript{115} However, the diet required is strict and there are no means currently to predict which patients will respond.

Elimination diets involve avoidance of foods that are putatively allergenic. Foods are deemed to be “allergenic” if their elimination results in decreased disease activity and subsequent ingestion results in increased disease activity. Although such diets may result in improvement in a proportion of patients, variable responsiveness and compliance can be a significant limitation.\textsuperscript{116,117} Notwithstanding, from a practical perspective, if a patient believes their arthritis is due to hypersensitivity to a particular food item or group, it may be worth assessing the effects of dietary avoidance in an n = 1 study of sequential withdrawal and challenge. Objective clinical and laboratory signs of disease activity should be documented after one or more cycles of withdrawal and reintroduction of the suspected food item. The findings can then be used as a basis for deciding whether the food item should be avoided by the individual in the longer term.

**Interactions between Diet and Disease-Modifying Antirheumatic Drugs**

It is well recognized that certain dietary components may interact with DMARDs. Methotrexate remains the anchor drug in the management of RA and is also used commonly in many other rheumatic diseases. Alcohol may increase the risk of hepatotoxicity in patients receiving methotrexate, and alcohol intake should be moderate at most. Methotrexate has multiple potential actions, many of which can be traced to its action as a dihydrofolate reductase antagonist. Routine supplementation with folic acid is recommended to reduce potential adverse effects associated with methotrexate.\textsuperscript{118} However, higher red blood cell folate concentrations have been associated with higher disease activity in RA patients receiving methotrexate.\textsuperscript{119} In some countries fortification of flour, pasta, rice, and bread with folic acid is a governmental requirement. Not surprisingly, such dietary fortification may lead to increased methotrexate doses in some patients.\textsuperscript{120}

Methotrexate treatment has been shown to increase extracellular accumulation of adenosine, which has multiple anti-inflammatory effects on neutrophils and macrophages including inhibition of IL-1\(\beta\) and TNF synthesis. Caffeine is a methylxanthine and acts as an adenosine receptor antagonist. Caffeine may therefore interfere with the effects of methotrexate. In rat adjuvant arthritis, caffeine reversed the anti-inflammatory effects of methotrexate.\textsuperscript{121} In patients with RA, less intense caffeine intakes have been shown to have no effect on methotrexate efficacy\textsuperscript{122} or, at lower mean dose of methotrexate, to impair efficacy in caffeine doses greater than 180 mg/day compared with doses less than 120 mg/day.\textsuperscript{123}
Gout

**KEY POINTS**

Dietary interventions in gout need to address the association between gout and the metabolic syndrome and increased cardiovascular disease.

Dehydration is a trigger for acute gout.

Weight loss may reduce serum urate.

Sustained reduction of SU to less than 6 mg/dL is critical for the effective management of gout. This is generally achieved through a combination of urate-lowering therapy and dietary/lifestyle modifications. When considering dietary interventions in the management of gout, it is important to recognize that gout is associated with the metabolic syndrome and with an increased risk of cardiovascular disease and mortality.

The diet usually advocated for gout entails restricted intake of foods and drinks that contain higher amounts of purines or that are thought to precipitate acute attacks of gout. Restricted items include meat, seafood, beer/wine, and legumes. Such diets are frequently high in saturated fats and carbohyrdrates, which may add to the risk of metabolic syndrome.

Dehydration is a potential trigger for acute gout attacks, and increased water intake has been associated with a reduced risk of gouty attacks. In a small short-term study in healthy volunteers, ingestion of cherries reduced serum urate (before urate 3.6 ± 0.2 mg/dL vs. after urate 3.1 ± 0.25 mg/dL; P < 0.05) and increased urinary urate excretion. Whether the same effects are observed in patients with gout remains to be determined.

Vitamin C supplementation reduces SU concentrations. In a study of 184 patients randomized to placebo or vitamin C 500 mg/day for 2 months, SU was reduced significantly in the vitamin C group with a mean reduction of −0.5 mg/dL (95% CI, −0.5 to 0.02 mg/dL) compared with the placebo group (P < 0.0001). In the subgroup of 21 patients with baseline SU greater than 7 mg/dL, the mean reduction in serum urate was 1.3 mg/dL. This reduction in SU suggests vitamin C is a possible additional therapy for patients with gout and persistent hyperuricemia.

In obese patients weight loss may reduce SU, as well as other health benefits. In one study, 13 males with gout lost weight over 16 weeks (pre BMI, 30.5 ± 8.1 kg/m²; post BMI, 27.8 ± 7.9 kg/m²). SU decreased in all subjects; pre SU, 9.6 ± 1.7 mg/dL versus post SU, 7.9 ± 1.5 mg/dL; P = 0.001. A decrease in frequency of gouty attacks was seen from 2.1 ± 0.8 attacks/month to 0.6 ± 0.7 attacks/month (P = 0.002). However, not all studies have shown a reduction in SU with weight loss.

Osteoarthritis

**KEY POINTS**

Weight loss is important in the management of OA.

From a mechanistic viewpoint n-3 fatty acids may be beneficial and can be an alternative or adjunct to NSAIDs for long-term analgesia.

Weight loss is generally considered an important aspect in the management of OA of weight-bearing joints. A recent meta-analysis of weight loss studies in patients with knee OA reported that moderate weight loss (≥5%) is associated with statistically significant improvement in self-reported disability scores. Although knee pain was also reduced, this did not reach statistical significance.

n-3 Fatty acids have recently been shown to reduce IL-1α–induced expression of the degradative enzymes such as MMP-3, MMP-13, and ADAMTs in bovine chondrocyte cultures, suggesting that n-3 supplementation may be beneficial in OA. Results of suitably powered studies of fish oil in OA are awaited.

Probiotics in the Management of Rheumatic Diseases

**KEY POINT**

There are insufficient data to support routine use of probiotics in inflammatory rheumatic diseases.

Although most interest has focused on the role of probiotics in inflammatory bowel diseases, there are a number of studies in animal models of arthritis, as well as human studies in RA and spondyloarthritis.

Probiotics have been reported to enhance the antiarthritic effects of methotrexate in adjuvant-induced arthritis in rats, whereas in collagen-induced arthritis probiotics downregulated Th1 effector cells, leading to suppression of joint inflammation and reduction in cartilage destruction.

In studies of patients with rheumatic diseases, results have varied depending on the probiotic administered. In a study of 21 patients with RA, 12 months’ supplementation with Lactobacillus rhamnosus had no effect on disease activity. In 63 patients with active spondyloarthritis, there was no difference in outcomes after 12 weeks of probiotic supplement containing Streptococcus salivarius, Bifidobacterium lactis, and Lactobacillus acidophilus compared with placebo. By contrast, in a 60-day study of 45 patients with RA, improvement in pain and patient global scores were seen with Bacillus coagulans GBI-30 supplements compared with placebo.

Alterations in gut microbial flora may also influence absorption and metabolism of drugs. Of particular importance for patients with rheumatic diseases are potential effects on DMARDs, especially salazopyrin, which is metabolized to the active metabolites sulfapyridine and 5-aminosalicylic acid by intestinal microbes. Probiotics that alter intestinal microflora may potentially alter efficacy and toxicity of salazopyrin. Although alterations in salazopyrin metabolism with probiotic supplements have been demonstrated in an animal model, similar effects were not observed in a human study.

In summary, although theoretic considerations and empiric data provide some support for use of probiotics in RA, the evidence is not sufficient to recommend routine use.
CONCLUSIONS

Rheumatic diseases are typically chronic, and, as with any chronic disease, nutritional issues are intrinsic to optimization of long-term health. Although patients as individuals can make their own nutritional choices, it behooves rheumatologists to be adequately informed regarding the plausibility and evidence for frequently considered nutritional supplements, toward which this chapter has been directed.

In terms of positive advice for patients and referring physicians, the strongest case can be made for dietary supplementation with fish oil in adequate doses. For long-term analgesia in rheumatic diseases, fish oil can be recommended in favor of NSAIDs. To the extent that fish oil is taken instead of NSAIDs, the risks for serious upper gastrointestinal complications and increased risk for serious thrombotic cardiovascular events associated with NSAIDs will be avoided because fish oil has not been associated with these risks. Furthermore, fish oil has been shown to reduce cardiovascular risks through mechanisms independent of and additional to its NSAID-sparing effects.

Selected References


References

Rheumatoid arthritis (RA) is a complex disease involving numerous cell types, including macrophages, T cells, B cells, fibroblasts, chondrocytes, neutrophils, mast cells, and dendritic cells.

Several genes are implicated in susceptibility to RA and severity of disease, including class II major histocompatibility complex genes, PTPN22, and peptidylarginine deiminases.

Evidence of autoimmunity, including high serum levels of autoantibodies such as rheumatoid factors and anticitrullinated protein antibodies, can be present for many years before the onset of clinical arthritis.

Adaptive and innate immune responses in the synovium have been implicated in the pathogenesis of RA.

Cytokine networks involving tumor necrosis factor, interleukin-6, and many other factors participate in disease perpetuation and can be targeted by therapeutic agents.

Bone and cartilage destruction are primarily mediated by osteoclasts and fibroblast-like synoviocytes, respectively.

Rheumatoid arthritis (RA) is the most common inflammatory arthritis, affecting from 0.5% to 1% of the general population worldwide. Although the prevalence is surprisingly constant across the globe, regardless of geographic location and race, there are some exceptions. For instance, in China the occurrence of RA is somewhat lower (∼0.3%), whereas it is substantially higher in other groups such as the Pima Indians in North America (∼5%). Because of its prevalence and the ready accessibility of joint samples for laboratory investigation, RA has served as a useful model for the study of all inflammatory and immune-mediated diseases. As such, the information gleaned from these studies provides new and unique insights into the mechanisms of normal immunity.

Although RA is primarily considered a disease of the joints, abnormal systemic immune responses are evident and can cause a variety of extra-articular manifestations. These manifestations clearly show that RA has features of a systemic disease that can involve many organs. In some cases, autoantibody production with the formation of immune complexes that fix complement contribute to these extra-articular findings. One of the mysteries of RA is why the synovium is the primary target, although the unique structure of its vascular bed could provide an environment that is ideal for innate and adaptive immune responses.

Although the precise causes of RA remain uncertain, environmental and genetic influences clearly participate. Clues have been provided by detailed immunogenetic studies and the observation that underlying autoimmunity antedates onset of arthritis by up to a decade. Progress in understanding the pathogenesis has been even more robust: The roles of small-molecule mediators of inflammation (e.g., arachidonic acid metabolites), autoantibodies, cytokines, growth factors, chemokines, adhesion molecules, and matrix metalloproteinases (MMPs) have been carefully defined. Synovial cells can exhibit behavior resembling a localized tumor that invades and destroys articular cartilage, subchondral bone, tendons, and ligaments. Irreversible loss of articular cartilage and bone begins soon after the onset of RA, and early interventions can probably improve long-term outcomes. Increased appreciation of how comorbidities, especially cardiovascular disease and accelerated atherosclerosis, can affect mortality has also led to attempts to suppress synovial and systemic inflammation.

ETIOLOGY AND PATHOGENESIS OF RHEUMATOID ARTHRITIS: ROLES OF INNATE AND ADAPTIVE IMMUNITY

The etiology and pathogenesis of RA are complex and multifaceted. A variety of predetermined (genes) and stochastic (random events and environment) factors contribute to susceptibility and pathogenesis. As an introduction, a summary of how these mechanisms interact to create and perpetuate RA is shown in Figure 69-1. Individual mechanisms are discussed in greater detail throughout the chapter.

The initiation of RA probably begins years before the onset of clinical symptoms. This process involves certain specific genes that can help break tolerance and lead to autoreactivity. It is likely that the earliest phases are marked by repeated activation of innate immunity (see Figure 69-1).
Cigarette smoke, bacterial products, viral components, and other environmental stimuli can contribute to these responses. This process probably occurs often in normal individuals but is self-limited. In individuals, a predetermined propensity for immune hyper-reactivity or auto-reactivity might lead to a different outcome. The genome of these individuals might encode for a variety of genes implicated in RA including class II major histocompatibility complex (MHC) genes, protein tyrosine phosphatase-22 (PTPN22), cytokine promoter polymorphisms, signal transduction gene polymorphisms, population-specific genes (e.g., PADI4 in Japanese or Koreans), and other undefined genes. Abnormal T cell selection could also contribute by allowing autoreactive T cells to escape deletion. The environmental stresses can lead to post-transcriptional modification of proteins, especially citrullination of arginine residues, in mucosal surfaces or the synovium. Although this commonly occurs without sequelae in normal individuals, people with a propensity for RA can develop antibodies against these modified proteins with production of rheumatoid factors (RFs) and anticitrullinated protein antibodies (ACPAs).

Activation of synovial innate immunity can also increase vascular leakage in the synovium, production of chemotactants that recruit immune cells to the joint, and processing of antigens by dendritic cells. Antigen presentation can potentially occur in the synovial germinal centers or, more commonly, in central lymphoid organs after the loaded dendritic cells migrate via the lymphatics. Naive T cells can then be activated through interactions with the T cell receptor and co-stimulatory signals. T cells can help B cells produce pathogenic antibodies and/or migrate to the joint, where they can influence other cells through the production of cytokines such as interleukin (IL)-17 or through cell contact mechanisms that do not require a specific antigen. Although it is uncertain what transforms subclinical inflammation to symptomatic arthritis, this process can take up to a decade before it reaches fruition.

Ultimately, a destructive phase proceeds, which can have antigen-dependent and -independent mechanisms and is mediated by mesenchymal elements such as fibroblasts and synoviocytes. Bone erosions are subsequently caused by osteoclasts, whereas cartilage dissolution results from proteolytic enzymes produced by synoviocytes in the pannus or synovial fluid neutrophils. Anti-inflammatory mechanisms such as soluble TNF receptors, suppressive cytokines, cytokine binding proteins, protease inhibitors, lipoxins, antioxidants, antiangiogenic factors, and natural cytokine antagonists are not present in sufficient concentrations to truncate the inflammatory and destructive process. The only way to suppress this response is through therapeutic interventions that either modulate pathogenic cells or neutralize the effector molecules produced by the rheumatoid process, or restore tolerance.
The heterogeneity of mechanisms provides an explanation for the unpredictable response to therapeutic agents and also allows clinicians to consider new therapeutic targets to either prevent RA or interfere with the immunologic, inflammatory, or destructive components as separate but interrelated entities. Each of these mechanisms is discussed in detail later. Brief summaries are also provided intermittently to help guide the reader through this complex maze.

** ETIOLOGY OF RHEUMATOID ARTHRITIS **

**KEY POINTS**

Genes play a key role in susceptibility to RA, as well as disease severity.

Class II major histocompatibility genes, especially those containing a specific 5 amino acid sequence in the hypervariable region of HLA-DR4, are the most prominent genetic association.

Newly defined genetic associations including polymorphisms in **PTPN22** and **PADI4** suggest that the associations in RA are complex and involve many genes.

Although the etiology of RA remains unknown, a variety of studies suggest that the interaction of environmental and genetic factors is responsible; either one is necessary but not sufficient for full expression of the disease. The most compelling example for a genetic component is in monozygotic twins, in whom the concordance rate is perhaps 12% to 15% when one twin is affected, compared with 1% for the general population. The fact that concordance is not higher provides key evidence that other influences such as the environment, epigenetics, or even microchimerism from maternal-fetal transfer might be as important as or even more important than the genetic component. The risk for a fraternal twin of a patient with RA is also high (≈2% to 5%) but similar to the rate for other first-degree relatives.

Although the immunogenetics is, at best, incompletely understood, one of the best-studied and perhaps most influential genetic risk factor is the class II MHC haplotype of an individual. **PTPN22** and **PADI4** increase risk in some racial and ethnic groups, but not all. Genome-wide screens have implicated at least 35 genes, many of which are involved with immune function. However, most have a relatively modest contribution and the susceptibility polymorphism confers only a 1.1- or 1.2-fold increase. Combinations of genes can clearly interact with one another, and a 45-fold increase in risk is conferred by a combination of HLA-DR, PTPN22, and the TRAF1-C5.\(^1\) This combination, however, is found in less than 1% of individuals with RA. The RA-associated alleles identified to date contribute approximately 40% of total genetic susceptibility. Additional progress in understanding the role of genes in RA including rare variants that might be more important than some common polymorphisms will require sophisticated bioinformatics to clarify how individual alleles contribute to susceptibility, severity, and response to targeted therapies.

** Role of HLA-DR in the Susceptibility to and Severity of Rheumatoid Arthritis **

The structure of class II MHC molecules on antigen-presenting cells is associated with increased susceptibility and severity of RA and accounts for about 40% of the genetic influence. A genetic link between HLA-DR and RA was initially described in the 1970s with the observation that HLA-DR4 occurred in 70% of RA patients, compared with about 30% of controls, giving a relative risk of having RA to those with HLA-DR4 of approximately 4 to 5.

The susceptibility to RA is associated with the third hypervariable region of DRB\(^\beta\)-chains, from amino acids 70 through 74. The epitope is glutamine-leucine-arginine-alanine-alanine (QKRAA), a sequence found in DR4 and DR14, in addition to some DR\(\beta\)-chains. The “susceptibility epitope” (SE) on DR\(\beta\)-chains with the greatest association with RA are DR\(\beta\)*0401, DR\(\beta\)*0404, DR\(\beta\)*0101, and DR\(\beta\)*1402 (Table 69-1). Up to 96% of patients with RA exhibit the appropriate HLA-DR locus in some populations.\(^3\) In certain ethnic and racial groups, however, the association with QKRAA is not as prominent or is not associated. The QKRAA epitope might also predict the severity of established RA, with a greater prevalence of extra-articular disease and erosions in patients with two copies. Other HLA genes such as DR\(\beta\)*1301 contain the DERAA sequence and are associated with decreased susceptibility to RA.\(^3\)

One intriguing possibility that could account for some patients who do not fit within this paradigm is microchimerism.\(^4\) Maternal cells expressing the SE can survive and persist in the circulation throughout adulthood. These non-inherited maternal antigens (NIMAs) could then confer increased risk of disease in the children of SE-expressing women.

The region associated with RA (QKRAA) primarily faces away from the antigen-binding cleft of the DR molecule that determines the specificity of peptides presented to

**Table 69-1 Nomenclature for HLA-DR Alleles and Associations with Rheumatoid Arthritis**

<table>
<thead>
<tr>
<th>Old Nomenclature (HLA-DRB(\beta) Alleles)</th>
<th>Current Nomenclature</th>
<th>Association with Rheumatoid Arthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DR1</td>
<td>0101</td>
<td>+</td>
</tr>
<tr>
<td>HLA-DR4 Dw4</td>
<td>0401</td>
<td>+</td>
</tr>
<tr>
<td>HLA-DR4 Dw14</td>
<td>0404/0408</td>
<td>+</td>
</tr>
<tr>
<td>HLA-DRw14 Dw16</td>
<td>1402</td>
<td>+</td>
</tr>
<tr>
<td>HLA-DR4 Dw10</td>
<td>0402</td>
<td>–</td>
</tr>
<tr>
<td>HLA-DR2</td>
<td>1501, 1502, 1601, 1602</td>
<td>–</td>
</tr>
<tr>
<td>HLA-DR3</td>
<td>0301, 0302</td>
<td>–</td>
</tr>
<tr>
<td>HLA-DR5</td>
<td>1101-1104, 1201, 1202</td>
<td>–</td>
</tr>
<tr>
<td>HLA-DR7</td>
<td>0701, 0702</td>
<td>–</td>
</tr>
<tr>
<td>HLA-Dw8</td>
<td>0801, 0803</td>
<td>–</td>
</tr>
<tr>
<td>HLA-Dw9</td>
<td>0901</td>
<td>–</td>
</tr>
<tr>
<td>HLA-Dw10</td>
<td>1001</td>
<td>–</td>
</tr>
<tr>
<td>HLA-Dw13</td>
<td>1301-1304</td>
<td>1301 associated with protection</td>
</tr>
<tr>
<td>HLA-Dw14 Dw9</td>
<td>1401</td>
<td>–</td>
</tr>
</tbody>
</table>

The shared epitope might not be an independent risk factor for RA but instead a marker for immunoreactivity and anticitrullinated protein antibodies (ACPs). In a large series of patients with early undifferentiated inflammatory arthritis, one-third of patients met criteria for RA within 1 year. Progress to RA occurred regardless of HLA-DR genotype if patients were anticitrullinated protein (anti-CP) positive. When patients were stratified according to ACPA, the shared epitope did not make an additional contribution to progression from undifferentiated arthritis to RA. The shared epitope probably contributes to immune hyperreactivity, but ACPAs are more closely associated with RA. In other studies, however, the presence of the shared epitope and ACPAs together is associated with even greater disease severity.

**Additional Polymorphisms: Cytokines, Citrullinating Enzymes, PTPN22, and Others**

The genetic influence on RA has also led to studies evaluating non-MHC genes. Single nucleotide polymorphisms (SNPs) in promoter regions, coding regions, or areas with no known function have been extensively investigated in RA with a variety of methods including genome-wide association studies. Table 69-2 shows some of the SNPs and microsatellites that have been associated with RA. The relative contribution for most is modest, and variations in technique, stage of disease, and patient populations result in some disagreement among various reports.

Given the importance of cytokines in RA (see following), it is not surprising that many studies have focused on these genes. The most intriguing evidence relates to tumor necrosis factor (TNF). This proinflammatory factor is a major cytokine in the pathogenesis of RA, and the TNF genes are located in the MHC locus on chromosome 6 in humans. Several polymorphisms of the TNF promoter including two at positions −238 and −308 can alter gene transcription. Associations among the TNF polymorphisms and RA susceptibility and radiographic progression have been reported, although there is not uniform agreement. In addition, certain polymorphisms in cytokines, especially TNF or Fc receptors, have been associated with differential response to therapy. For instance, substitution of a T for a C at position −857 in the TNF promoter might confer greater responsiveness to TNF inhibitors.

Among the many noncytokine and non-MHC genetic linkages described for RA, the ones associated with peptidyl arginine deiminase (PADI) and PTPN22 have the strongest effect on susceptibility. The PADI genes are responsible for the post-translational modification of arginine to citrulline. Four isoforms have been identified, known as PADI1 through PADI4. In light of the striking associations of RA with ACPAs, several groups have investigated potential associations with these genes. The most promising is an extended haplotype in the PADI4 gene that can lead to increased levels of PADI4 protein due to enhanced messenger RNA (mRNA) stability. In a Japanese cohort, a twofold increase in risk of RA was observed with PADI4 SNPs. Confirmatory reports have been mixed because the association has been confirmed in other Asian populations but not in Western Europe. These studies suggest that the contribution of PADI4 to RA might be restricted, depending on the overall genetic background of the patient population.

Protein tyrosine phosphatase-22 (PTPN22) associations have been discovered in large-scale screening efforts to identify SNP associations in RA. Using 12,000 SNPs in the initial screens, a novel association was discovered at position 1858 in the PTPN22 gene that, like PADI4, conferred a twofold increase in risk. The allele containing thymidine leading to an amino acid substitution (R620W) was present in 8.5% of controls but was found in nearly 15% in patients with seropositive RA. Subsequent studies have demonstrated a similar association with systemic lupus erythematosus (SLE), type 1 diabetes, and several other autoimmune diseases. PTPN22 is a phosphatase that regulates the phosphorylation status of several kinases important to T cell activation including Lck and ZAP70. The R620W allele surprisingly results in a gain of function that alters the threshold for T cell receptor (TCR) signaling. Because the PTPN22 allele is rare in Japan, it is another gene (e.g., PADI4) where susceptibility is specific for particular ethnic or racial populations.

The list of genes associated with RA consistently involves immune regulation. Cytokine polymorphisms such as for TNF and the IL-1 inhibitor, IL-1Ra are not surprising. Genes that regulate adaptive immune responses in T cells such as PTPN22 and the co-stimulation receptor CTLA have also been associated with RA. Other genes associated with B cell function and/or antigen presentation such as BTLA (B- and T-lymphocyte attenuator), Fc receptors, and CD40 are also implicated. Polymorphisms have also been identified in signal transduction pathways that regulate immune function such as TRAF1-C5 and STAT4. The consistent thread in this analysis is that most gene associations for RA cluster to innate immunity, adaptive immunity, and inflammation. Aside from providing insight into the mechanisms of disease, they could also potentially contribute to responses to targeted therapies.

### Table 69-2 Key Genetic Associations in Rheumatoid Arthritis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Odds Ratio for Risk Alleles</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DR</td>
<td>4-5 fold</td>
<td>Not in Asian populations</td>
</tr>
<tr>
<td>PTPN22</td>
<td>2 fold</td>
<td>Primarily in Asian populations</td>
</tr>
<tr>
<td>PADI4</td>
<td>2 fold</td>
<td></td>
</tr>
<tr>
<td>TRAF1-C5</td>
<td>&gt;1.2 fold; &lt;2 fold</td>
<td></td>
</tr>
<tr>
<td>STAT4</td>
<td>&gt;1.2 fold; &lt;2 fold</td>
<td></td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>&gt;1.2 fold; &lt;2 fold</td>
<td></td>
</tr>
<tr>
<td>IL2/2</td>
<td>&gt;1.2 fold; &lt;2 fold</td>
<td></td>
</tr>
</tbody>
</table>

Other genes with odds ratio >1.0 and <1.2: CTLL4, CD40, CCL21, CD244, IL2RB, TNFRSF14, PRKCR, PIK3CC, IL2RA, AFF3, REL, BLK, TAGAP, CD28, TRAF6, PP1RIC, FCGR2A, PRDM1, CD2-CD8, IRF5, CCR6, CCL21, IL6ST, RBPJ.
Interactions between Genes and Environment

A number of environmental factors clearly contribute to RA susceptibility, although no specific exposure has been identified as the pivotal agent. Smoking is the best defined environmental risk factor for seropositive RA. The reason for its influence on the development of synovitis is not fully defined but could involve the activation of innate immunity and PADJ in the airway. Citrullinated proteins have been detected in bronchoalveolar lavage samples of smokers, and this could provide a stimulus for generation of ACPAs in susceptible individuals. Repeated activation of innate immunity, especially in an individual with underlying genetically determined autoreactivity, could potentially contribute to autoreactivity and the initiation of synovitis. Other environmental factors such as oral contraceptives appear to modestly protect from RA, perhaps due to changes in the hormone milieu.

The interaction between HLA-DR and tobacco exposure is perhaps the best example of how genes and the environment conspire to enhance risk. Although smoking and the SE alone modestly increase the likelihood of developing RA, the combination is synergistic. An individual with a history of cigarette smoking and two copies of the SE increases the odds of developing RA by up to 40-fold. The mechanism of the interaction is not known, but it could potentially relate to the increase in protein citrullination in smokers and increased ability of SE-containing HLA-DR molecules to bind some citrullinated proteins. The extent of smoking is also predictive, with the greatest risk seen with at least 20 pack-years. The risk declines slowly with cessation of smoking, taking more than a decade to begin approaching nonsmokers. Alcohol consumption can decrease this risk, and exposure to other inhaled irritants like silica dust increases risk, demonstrating the complexity of environment and human behavior on understanding disease susceptibility.

Gender

RA is one of many chronic autoimmune diseases that predominates in women. The ratio of female-to-male patients is 2:1 to 3:1, which is not as high as Hashimoto’s thyroiditis (25:1 to 50:1) or SLE (9:1). The gender effect is often observed in some animal models of autoimmunity such as the NZB/NZW model of SLE, in which female mice have more severe disease. Estrogens are one obvious explanation, and some data support the concept that these hormones modulate immune function. For example, autoantibody-producing B cells exposed to estradiol are more resistant to apoptosis, suggesting that autoreactive B cell clones might escape tolerance. The effect on T lymphocytes is harder to reconcile with the female preponderance in RA because estrogens tend to bias T cell differentiation toward the Th2 phenotype. The cytokines produced by this subset such as IL-4 and IL-13 are usually considered anti-inflammatory in animal models of arthritis and are present in only limited amounts in the RA synovium. Estrogen receptors are expressed on fibroblast-like synoviocytes (FLS) and increase production of metalloproteinases. In macrophage cell lines, estrogen can enhance production of TNF. Nulliparity has also been suggested as a risk factor in early studies, but more recent reports do not support this notion. Thus the effects of estrogens are complex, and the specific mechanisms responsible for the female preponderance of RA are not fully understood.

Pregnancy is often associated with remission of the disease in the last trimester. More than three quarters of pregnant patients with RA improve in the first or second trimester, but 90% of these experience a flare of disease associated with a rise in RF titers in the weeks or months after delivery. The mechanism of protection is not defined but might be due to the expression of suppressive cytokines such as IL-10 during pregnancy, production of α-fetoprotein, or alterations in cell-mediated immunity. One intriguing finding is that fetal DNA levels in the maternal peripheral blood correlate with the propensity for improved symptoms in pregnant RA patients. It is not certain whether the DNA itself contributes or whether it is a marker for increased leakage of fetal cells into the maternal circulation. Immune responses directed against paternal HLA antigens can occur and lead to the production of alloantibodies in the maternal circulation. Maternal-fetal disparity in human leukocyte antigen (HLA) class II phenotypes can correlate with pregnancy-induced remission. More than three-fourths of pregnant women with maternal-fetal disparity of HLA-DRB1, DQA, and DQB haplotypes have significant improvement, whereas disparity is only observed in one-fourth of women whose pregnancy is characterized by continuous active arthritis. Therefore suppression of maternal immune responses to paternal HLA haplotypes might be protective. This question remains unsettled because another study failed to find a correlation between the HLA disparity and clinical improvement during pregnancy.

Epigenetics

Epigenetics describes phenotypic or gene expression properties caused by mechanisms other than changes in the underlying DNA sequence. Modification of CpG DNA sequences by methylation, for example, can suppress gene expression, and it plays a role in cell differentiation. Histone acetylation also alters accessibility of DNA for transcription factors and RNA polymerases. microRNAs can bind to DNA and suppress expression of key genes involved with the inflammatory process. Some epigenetic information such as DNA methylation can be transferred from one generation to the next, thereby providing an alternative mechanism for rapidly altering disease susceptibility in a population due to the environment.

Most information on epigenetics in RA comes from studies of RA synovium or cultured synoviocytes. Some evidence of imprinting is available in the latter, with evidence of global DNA hypomethylation. Only low levels of a key DNA methylase, Dnmt1 are expressed in RA synovium, suggesting a molecular basis for this observation. Because this enzyme is carried by gametes, the methylation pattern can be transgenerationally maintained. One particular CpG site in the IL-6 promoter of peripheral blood mononuclear cells has decreased methylation, and this is associated with higher levels of IL-6 production. Dietary influences such as ingestion of methyl donors such as folate, or even exposure to methyl donors in utero, can profoundly alter DNA methylation and adaptive
immune functions and affect susceptibility to autoimmune disease.

The histone deacetylase HDAC1 is overexpressed in RA FLS. When this gene is suppressed, synoviocyte proliferation decreases and expression of tumor suppressor proteins such as p53 increases. HDAC inhibitors are also effective in collagen-induced arthritis, markedly delaying the onset of disease and decreasing bone erosions. Finally, analysis of synoviocytes shows that some individual microRNAs such as microRNA-124a are decreased in RA compared with osteoarthritis (OA) cells. This particular microRNA can suppress cell cycling and chemokine genes. Increasing microRNA-124a levels in RA synoviocytes decreased the production of the chemokine MCP-1. Forced expression of microRNA-203 increases metalloproteinase and IL-6 expression by synoviocytes as well.

The role of epigenetics in RA is not understood. It is not clear whether these observations occur before the onset of disease, are involved with the transition from asymptomatic autoimmunity to clinical disease, or participate in the destructive phase in established RA. Environmental stress plays a major role in disease susceptibility, perhaps greater than DNA polymorphisms. The mechanisms probably involve epigenetic deregulation of gene expression leading to decreased thresholds for autoreactivity in the adaptive immune system.

Changing Epidemiology of Rheumatoid Arthritis

The history of RA reveals the surprising observation that it is a relatively new disease in Europe and Northern Africa. Examination of ancient skeletal remains in Europe and Northern Africa fails to reveal convincing evidence of RA, even though other rheumatic diseases such as OA, ankylosing spondylitis, and gout are readily discernable. In contrast, typical marginal erosions and rheumatoid lesions are present in the skeletons of Native Americans found in Tennessee, Alabama, and Central America from thousands of years ago. The first clear descriptions of RA in Europe appeared in the seventeenth century, and the disease was distinguished from gout and rheumatic fever by Garrod in the mid-nineteenth century. Although still controversial, the disease might have migrated from the New World to the Old World coincident with opening the trade and exploration routes. Because genetic admixture was relatively limited, an undefined environmental exposure potentially caused RA in susceptible Europeans. The most obvious explanation would, of course, be that an infectious agent is responsible. However, other environmental influences like tobacco smoking were also introduced to the Old World at the same time and could play a role.

Equally intriguing, the severity and incidence of RA appeared to decrease in the late twentieth century (Figure 69-2). In certain well-defined populations including Native Americans, the incidence of RA has gradually declined by as much as 50% over the past half of the twentieth century. Changes in hygiene and other lifestyle modifications related to industrialization might contribute, and an infectious agent might be less prevalent secondary to these societal changes, as with many other infectious diseases. Recent data from 1995 to 2007 suggest that the incidence might be rising again in women, but not in men.

Dissecting the environmental exposures will be key to understanding how the susceptibility to disease varies over time.

**PATHOGENIC MECHANISMS IN RHEUMATOID ARTHRITIS**

**KEY POINTS**

Although an etiologic link has not been established, pathogens such as viruses, retroviruses, bacteria, and mycoplasma have been associated with RA.

A single specific “RA pathogen” is unlikely.

Repeated inflammatory stress, especially through specialized receptors that recognize common molecules produced by pathogens, in a genetically susceptible individual might contribute to breakdown of tolerance and subsequent autoimmunity.

Considerable effort has been expended to assess the role of infectious agents in RA (Table 69-3). A potential pathogen could initiate disease through a variety of mechanisms including direct infection of the synovium, activation of innate immunity by pattern-recognition receptors that bind to components of the agent, or through molecular mimicry that induces an autoreactive adaptive immune response.

**Infectious Agents: Direct Infection and Innate Immune Responses**

**Toll-like Receptors and the Inflammasome in the Joint**

Infectious agents could contribute to the initiation or perpetuation of RA through a variety of mechanisms. Some arthrotropic microorganisms could potentially infect the synovium and cause a local inflammatory response. There is increasing awareness that the innate immune system could also directly affect the onset and course of synovitis.
Pathogen-associated molecular pattern receptors, especially the Toll-like receptors (TLRs), are expressed by sentinel cells in the host that provide a first line of defense. These receptors recognize preserved structures in bacteria and other infectious agents and permit rapid release of inflammatory mediators, activation of antigen-presenting cells, and enhancement of adaptive immune responses.

At least 11 TLRs exist in humans such as TLR2 (binds peptidoglycans), TLR3 (binds double-stranded RNA [dsRNA]), TLR4 (binds lipopolysaccharide), and TLR9 (binds bacterial DNA containing CpG motifs). Many of these pattern recognition receptors are expressed by rheumatoid synovial tissue and cultured FLS including TLR2, TLR3, TLR4, and TLR9. Exogenous TLR ligands such as bacterial peptidoglycan and DNA, as well as endogenous ligands (e.g., heat shock proteins, fibrinogen, and hyaluronan), are present in arthritic joints (see later). Engagement of these receptors participates in certain animal models of arthritis and can exacerbate synovial inflammation. TLR3, which recognizes viral dsRNA and activates the antiviral response, is also expressed by synovial cells in the intimal lining. Necrotic debris containing mRNA from RA synovial fluid cells activates TLR3 signaling and proinflammatory gene expression in synovitis.

The role of innate immunity in RA led to the notion that repeated engagement of TLRs in the synovium could help initiate disease. This hypothesis could explain why specific pathogens have been difficult to identify in the joint. In contrast, a genetically susceptible individual could potentially break tolerance if the TLRs are repeatedly engaged and permit autoimmune responses against articular antigens. Several animal models of disease require TLR ligands for initiations such as TLR9 in adjuvant arthritis. TLR2 is required for streptococcal cell wall arthritis, and the chronic T cell–dependent phase of that model requires TLR4. Mice lacking TLR4−/− have significantly less joint damage induced by IL-1 overexpression, even though synovial inflammation is still robust.24 These data suggest that endogenous TLR ligands play a key role in matrix regulation independent of inflammatory responses.

A second mechanism that regulates innate immunity involves a novel structure called the inflammasome. This complex includes several proteins involved in recognition of “danger signals” and pathogens such as muramyl dipeptides and uric acid. One central component is cryopyrin, also called NALP3, which is linked to caspase 1 (IL-1 convertase) by adapter proteins. When the inflammasome is engaged, caspase 1 is activated and IL-1 is produced. Mutations in this pathway, especially in cryopyrin, have been associated with autoinflammatory disorders such as Muckle-Wells syndrome and familial cold autoinflammatory disease. Inflammation induced by uric acid crystals or ATP uses this pathway and can be abrogated by IL-1 inhibitors. Cryopyrin is abundant in RA synovium and is constitutively expressed by FLS and macrophages. Expression in cultured FLS is markedly increased by TNF. Although the role of the inflammasome in RA has not been fully defined, its ability to induce cytokine production by exposure to bacterial products and other danger signals suggests that it participates in IL-1 and IL-18 regulation.

### Table 69-3 Etiology of Rheumatoid Arthritis: Possible Infectious Causes

<table>
<thead>
<tr>
<th>Infectious Agent</th>
<th>Potential Pathogenic Mechanisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mycoplasma</td>
<td>Direct synovial infection; superantigens</td>
</tr>
<tr>
<td>Parvovirus B19</td>
<td>Direct synovial infection</td>
</tr>
<tr>
<td>Retroviruses</td>
<td>Direct synovial infection</td>
</tr>
<tr>
<td>Enteric bacteria</td>
<td>Molecular mimicry (QKRAA, e.g., in bacterial heat shock proteins)</td>
</tr>
<tr>
<td>Mycobacterium</td>
<td>Molecular mimicry (proteoglycans, QKRAA), immunostimulatory DNA (Toll-like receptor 9 activation)</td>
</tr>
<tr>
<td>Epstein-Barr virus</td>
<td>Molecular mimicry (QKRAA in gp110)</td>
</tr>
<tr>
<td>Bacterial cell walls</td>
<td>Toll-like receptor 2 activation</td>
</tr>
</tbody>
</table>

### Bacteria, Mycobacteria, Mycoplasma, and Their Components

Active infection of synovial tissue by pyogenic bacteria is an unlikely cause of RA, and extensive searches for a unique or specific organism in synovial tissue or joint effusions have been negative. Antibodies to certain organisms such as Proteus are reportedly elevated in the blood of patients with RA, but this could represent an epiphenomenon or a non-specific B cell activation. Most RA and reactive arthritis patients contain bacterial DNA sequences in their synovium. The bacteria identified are not unique and generally represent a cross-section of skin and mucosal bacteria including Acinetobacter and Bacillus spp. It is possible that the synovium functions as an adjunct to the reticuloendothelial system in arthritis, allowing local macrophages to accumulate circulating bacterial products.

In addition to prokaryotic DNA, bacterial peptidoglycans have been detected in RA synovial tissue (Figure 69-3). Antigen-presenting cells containing these products express TLRs and produce proinflammatory cytokines such as TNF. It is not known whether the peptidoglycans activate cells in situ or whether phagocytic cells from other sites or the blood engage the molecules and then migrate to the joint. In either case, it is not difficult to imagine how they can contribute to synovial inflammation.

Several animal models of arthritis are dependent on TLR2, TLR3, TLR4, or TLR9. For instance, rodents injected with streptococcal cell walls (TLR2 ligand) develop severe polyarticular arthritis. The initial phase of disease resolves and is then followed by a chronic T cell–dependent phase that resembles RA. The arthritogenicity of complete Freund’s adjuvant in the rat adjuvant arthritis model is dependent on mycobacterial DNA that binds to TLR9 and activates an adaptive immunity. Endogenous TLR4 ligands such as heat shock proteins and fibrinogen also play a role in immune complex models such as passive K/BxN arthritis.

Mycoplasma-derived superantigens such as from Mycoplasma arthritidis can directly induce T cell–independent cytokine production by macrophages and can exacerbate or trigger arthritis in mice immunized with type II collagen. There is also a higher prevalence of antimycoplasma pneumonie IgG antibodies in RA patients than matched controls. Despite this and other circumstantial evidence, most efforts to identify Mycoplasma and Chlamydia organisms or DNA in joint samples have been negative, and there is no direct evidence to support these organisms as etiologic agents.
might cause an immune response directed at innocent bystander cells through “molecular mimicry.” This hypothesis could potentially account for disease perpetuation in the absence of active infection in patients with a specific MHC genotype. However, the data are circumstantial and gp110 is only one of many xenoproteins such as the *Escherichia coli* heat shock protein dnaJ that contain QKRAA. RA T cells, especially synovial fluid T cells, have increased proliferative responses to gp110, perhaps supporting the molecular-mimicry link between a variety of QKRAA-containing proteins and arthritis.

**Parvovirus**

Antecedent infection with parvovirus B19 has been implicated in some patients with RA based on serologic evidence including the nonstructural protein NS1. However, only about 5% of patients have evidence of recently acquired parvovirus B19 infection at the time of disease onset. Of interest, 75% of RA synovium samples contain B19 DNA compared with about 20% of non-RA controls. Immunohistochemical evidence of the B19 protein VP-1 was detected in patients with RA but not other forms of arthritis. However, evidence of the B19 genome in RA joint samples was not found in other studies.

The mechanisms of B19-induced synovitis, when it does occur, could be related to alterations in the function of FLS. In a cell-culture model of synoviocyte invasion into cartilage, infection with the parvovirus significantly increased the migration of cells into the matrix. Mice that are transgenic for the B19 protein NS1 were more susceptible to collagen-induced arthritis and developed high titers of anti-type II collagen antibodies. These data suggest that the B19 genome might not cause arthritis but can enhance an arthritogenic response to other environmental stimuli.

**Other Viruses**

Because rubella virus and the rubella vaccine can cause arthritis in humans, the virus has attracted some attention as a possible triggering agent. Live rubella virus can be isolated from synovial fluid of some patients with chronic inflammatory oligoarthritis or polyarthritis without clinical evidence of rubella. However, most rubella patients do not have the classic polyarticular involvement and display an oligoarthritis involving large joints. As with B19 infection, it is possible that a small subset of patients with chronic polyarthritides that are called RA actually have direct infection with wild-type or attenuated rubella virus.

Studies of synovial tissue in a variety of inflammatory and noninflammatory arthropathies have also demonstrated DNA of other viruses such as cytomegalovirus and herpes simplex, but not adenovirus or varicella-zoster. As with bacterial DNA, parvovirus, and EBV, the localization of viral DNA to the inflamed joint might be related to the migration of inflammatory cells containing the viral genome or other nonspecific mechanisms rather than an active infection. Although the hypothesis that one or more of these viral infections might serve as a triggering agent in the genetically susceptible host is both appealing and intellectually satisfying, the pathogenic role of these agents is unlikely.
Retroviral infections have been suggested as a cause of RA. Extensive searches for potential agents have not been fruitful. Endogenous retroviruses are abundant in inflamed and normal synovium, and certain transcripts are expressed in RA cells. In one study, higher levels of HERV-K10 gag protein from a common endogenous retrovirus were detected more often in RA compared with OA and normal peripheral blood mononuclear cells. Some indirect studies are suggestive of retroviral infection such as the demonstration of zinc-finger transcription factors in cultured synoviocytes that can increase signaling through enzymes like p38 mitogen-activated protein (MAP) kinase. In addition, the pX domain of one human retrovirus, human T lymphotropic virus-1 (HTLV-1), causes synovitis in transgenic mice, and synoviocytes from patients infected with HTLV-1 have increased cytokine production. Other studies failed to demonstrate increased expression of human retrovirus-5 proviral DNA in rheumatoid synovium. There is still no direct evidence that retroviruses cause RA, but some viral products could activate TLR3 or TLR7 to enhance cytokine and chemokine production.

Autoimmunity

KEY POINTS

Evidence of autoimmunity can be present in RA many years before the onset of clinical arthritis.

- Autoantibodies such as RFs and anticitrullinated protein antibodies are commonly associated with RA.
- Autoantibodies in RA can either recognize joint antigens such as type II collagen, or systemic antigens such as glucose phosphate isomerase.
- These autoantibodies can potentially contribute to synovial inflammation through several mechanisms including local activation of complement.

The idea that aberrant immune responses are directed toward self-antigens in RA was recognized with the discovery of RF in the blood of patients with the disease. Initially described by Waaler and later by Rose, it was not until the mid-1950s that Kunkel and colleagues firmly established that RF is an autoantibody. Although our understanding of autoantigens has changed over the years and the relative contributions of cellular and humoral immunity have been debated, emphasis on the role of autoantibodies in RA has enjoyed a resurgence over the past few years. Clinical improvement can be associated with decreases in levels of RFs or ACPAs, although the changes tend to be modest and are inconsistent.

Rheumatoid Factor

The identification and characterization of RF as an autoantibody that binds to the Fc portion of IgG was the first direct evidence that autoimmunity might play a role in RA. For many years, immune complexes comprising RF and other immunoglobulins were thought to play a primary role in the pathogenesis of synovitis (see Chapter 56). Even today, the presence of RF and its resultant pathogenic consequences are still considered cardinal features of RA. Longitudinal studies show that production of RF and other autoantibodies can precede the onset of RA by many years (Figure 69-4). Although some patients are initially “seronegative” for RA and subsequently convert to “seropositive,” this is rather unusual and seroconversion typically occurs during the first year of disease activity.

The role of RF and other autoantibodies in the pathogenesis of RA has been suggested by circumstantial evidence. For instance, patients with a positive test result for RF in blood have more severe clinical disease and complications than seronegative patients including increased cardiovascular complications. RF can also fix and activate complement by the classic pathway, and there is clear evidence of local complement production and consumption in the rheumatoid joint. Large quantities of IgG RF are produced by rheumatoid synovial tissue and form complexes through self-association. RF-containing immune complexes are readily detected in RA synovial tissue, as well as the surface layers of cartilage. The latter is especially relevant because immobilized complexes can facilitate complement fixation with resultant release of chemotactic peptides. In experiments performed in patients with RA, a marked inflammatory response was elicited when RF from one patient was injected into a joint, but not when normal IgG was given. B cell–targeted therapies such as rituximab can deplete peripheral B lymphocytes and modestly decrease titers of RF. This does not correlate precisely with clinical responses, and synovial RF production is not significantly changed by B cell depletion. Nevertheless, RF levels often decrease in responders and increase again coincident with clinical relapses.

Three-quarters of patients with RA are seropositive using standard tests for RF, although the percentage can be as high as 90% when assayed for IgM RF with enzyme-linked immunosorbent assays. First-degree relatives of seropositive patients with RA are frequently seropositive, suggesting a
genetic contribution. Although IgG and IgM RFs are the most abundant in RA, IgE RF has also been demonstrated, especially in patients with extra-articular manifestations. IgE RF can potentially complex with aggregated IgG in synovial tissue, and the subsequent complexes then could degranulate synovial mast cells through activation of Fc receptors in the synovium. IgA RFs are also produced in RA including patients who are seronegative as determined by standard clinical tests that primarily detect IgM RF.

The RFs produced in RA differ from those produced by healthy individuals or from patients with paraproteins. The avidity of RF for the Fc portion of IgG is much greater in RA than in Waldenström’s macroglobulinemia or in cryoglobulins. Many RFs expressed by abnormal B cells (such as Waldenström’s macroglobulinemia) and by normal B cells in human tonsils are derived from the germline. In contrast, RFs in RA are derived through rearrangements and somatic mutations of the germline genes. RF production can be quite high in rheumatoid synovium; IgM RF represents about 7% of the total IgM and 3% of the total IgG produced by synovial cell cultures.

RFs in RA primarily use the variable heavy 3 (VH3) gene and a variety of variable light (VL) genes, whereas natural antibodies with RF activity use VH1 or VH4 and the VK3 genes. The kappa light chain repertoire expressed in RF-producing cells isolated from one RA patient was enriched for two specific VK genes but contained many somatic mutations and non-germline-encoded nucleotides. Therefore the selection and production of these specific RFs were likely due to antigenic drive rather than derived directly from the germline. Additional RFs have been identified with characteristics similar to an antigen-driven response, although some examples of germline RFs have also been isolated from RA synovium. A crystal structure of one IgM RF bound to IgG showed a key contact residue of the RF with the Fc portion of IgG containing a somatic mutation, supporting the notion that the mutations are related to affinity maturation.

Anti-citrullinated Protein Antibodies (ACPAs). One of the most striking recent observations related to autoantibodies is the observation that immunoglobulins that bind to citrullinated proteins are produced by patients with RA and have significant prognostic implications. The discovery originated with reports in the 1970s that antibodies directed against keratin were detected in rheumatoid serum and that the primary target antigen was filament-aggregating protein, filaggrin. These antibodies actually bind to epitopes on filaggrin that contain citrulline, which is derived from post-translational modification of arginine by PADI. Humans have four isoforms of PADI. PADI2 and PADI4 are especially abundant in synovium, and certain SNPs are associated with RA in Asian populations. The function of PADIs in normal immune responses is not certain; citrullination of some chemokines can decrease activity, and modification of histones can regulate gene expression in stressed cells.

Induction of PADI expression and citrullination of peptides are not specific to RA and can occur in many inflammatory settings. Not only are other inflammatory arthropathies marked by citrullinated proteins, but other organs such as the lungs in smokers have significant PADI activity. The presence of CPs in the lungs of smokers could provide the systemic antigen exposure that can contribute to anti-CP antibody production and begin the long road to developing RA. CPs are present in most animal models of arthritis. Immunohistochemistry demonstrates citrullinated proteins in RA synovial tissue infiltrating cells (Figure 69-5), as well as in extracellular deposits that often colocalize with various isoforms of PADI, especially PADI2 and

![Figure 69-5](image-url)
The specific proteins that are modified vary widely but include many normal constituents like fibrinogen, vimentin, and fibronectin, as well as xenoproducts like EBV-derived peptides.\textsuperscript{14} The pattern of protein citrullination or the specificity of antibodies to particular proteins does not, thus far, have predictive value for developing RA or the severity of disease. One interesting subset involves antibodies that recognize mutant citrullinated vimentin (MCV). These antigens have two modifications that render them more immunogenic: First, there are amino acid changes in the primary sequence such as glycine to arginine. Second, either the new amino acid or another arginine on the protein is citrullinated.\textsuperscript{15} The mechanism of amino acid changes has not been defined but could be due to oxidative stress that causes mutations in the genes encoding the protein. These anti-MCV antibodies might be more specific for RA in established disease and also might be more predictive for radiographic progression than standard ACPA tests.\textsuperscript{36}

ACPAs are present in the serum of 80% to 90% of RA patients. In some studies, they are also more specific for RA than RF, with specificity approaching 90%. Perhaps most interesting, ACPA, like RF, can appear long before the onset of clinical arthritis and could be a marker for immune hyper-reactivity and subclinical inflammation leading to protein citrullination in a variety of tissues. The more peptides recognized by the repertoire of ACPAs, such as to vimentin, enolase, or various fragments of fibronectin, the greater the likelihood that an individual with arthralgias will develop RA.\textsuperscript{37} Of interest, elevated serum cytokines also predicts development of arthritis that meets criteria for RA.\textsuperscript{38} A genetic contribution was confirmed in the Native American population, where ACPAs are present in nearly 20% of unaffected first-degree relatives and more than 10% of more distant relatives.\textsuperscript{39} ACPAs are also produced by synovial tissue B cells and can be detected in synovial fluid. However, the fact that ACPAs, like RFs, can be detected years before the onset of arthritis and can be found in some normal relatives of RA patients suggests that the autoantibodies alone cannot account for the etiology of the disease.

ACPAs are predictors of more aggressive disease marked by bone and cartilage destruction. In fact, some data suggest that the HLA-DR associations in RA are actually due to an association between the susceptibility epitope and ACPA production. This association extends to accelerated atherosclerosis in RA, where anti-CP positivity is an independent risk factor for ischemic heart disease. In patients with early undifferentiated inflammatory arthritis, ACPAs are also predictive for individuals who will progress to RA.\textsuperscript{40}

Despite these caveats, ACPAs do have pathogenic potential. For instance, ACPAs can activate the classical and the alternative complement pathways.\textsuperscript{41} In addition, IgE ACPAs from patients with RA can sensitize basophils and mast cells to degranulate. This could enhance increased vascular permeability and influx of inflammatory cells into the synovium.\textsuperscript{42} The antibodies have minimal effect when directly injected into mice, and they enhance the arthritogenic potential of anti-type II collagen antibodies in the collagen-induced arthritis model.\textsuperscript{43} Mice immunized with citrullinated fibronectin, but not native fibronectin, can develop inflammatory arthritis.\textsuperscript{44} Hence the autoreactivity and autoantibodies are not simply a marker of disease but can participate in the disease process. Citrullination can also increase T cell responses to arthritogenic antigens. For instance, citrullination of albumin leads to the formation of antibodies that also cross-react with the unmodified protein. The citrullinated type II collagen is more immunogenic than the native protein, most likely due to increased affinity in the binding groove of HLA-DR proteins that contain the SE.\textsuperscript{45}

**Autoimmunity to Cartilage-Specific Antigens**

Because synovial tissue inflammation is a hallmark of RA, it is only natural to assume that certain joint-specific antigens might play an etiologic or pathogenic role. The number of potential antigens is extensive, and there is no convincing evidence to date that one specific “rheumatoid” antigen exists. In contrast, the emerging picture of autoimmunity in RA tends to implicate patterns of self-directed responses, rather than a single epitope that encompasses all patients at all times during the disease. It is quite possible that articular autoimmunity could vary with the stage of disease, the clinical manifestations, and treatment.

**Type II Collagen.** The discoveries that immunization with type II collagen can cause arthritis in rats and mice and that the disease can be passively transferred by IgG fractions containing anticollagen antibodies or by transfer of lymphocytes from affected animals have spawned extensive experiments that illustrate the antigenicity of collagen, the arthrotropic nature of the disease produced, and the dependence on class II MHC genes. T cells are required for initiation of collagen-induced arthritis, and a major immunogenic and arthritogenic epitope on type II collagen resides in a restricted area of the type II collagen chains.

However, type II collagen antibody generation is probably not the initiating event in RA but can amplify the inflammatory response (Table 69-4). Sera from patients with RA contain antibody titers to denatured bovine type II collagen that are significantly higher than those found in control sera; however, there is no difference in antibody titers to native collagen, indicating that the denatured form generated after the breakdown of connective tissue might serve as the immunogen. Anticollagen antibodies purified from the sera of patients with RA can activate complement,

<table>
<thead>
<tr>
<th>Table 69-4</th>
<th>Examples of Autoantigens in Rheumatoid Arthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cartilage antigens</td>
<td></td>
</tr>
<tr>
<td>Type II collagen gp39</td>
<td></td>
</tr>
<tr>
<td>Cartilage link protein gp39</td>
<td></td>
</tr>
<tr>
<td>Proteoglycans</td>
<td></td>
</tr>
<tr>
<td>Aggrecan</td>
<td></td>
</tr>
<tr>
<td>Citrullinated proteins</td>
<td></td>
</tr>
<tr>
<td>Glucose-6-phosphoisomerase</td>
<td></td>
</tr>
<tr>
<td>HLA-DR (QKRAA)</td>
<td></td>
</tr>
<tr>
<td>Heat shock proteins</td>
<td></td>
</tr>
<tr>
<td>Heavy-chain binding protein (BiP)</td>
<td></td>
</tr>
<tr>
<td>hnRNP-A2</td>
<td></td>
</tr>
<tr>
<td>Immunoglobulins (IgG)</td>
<td></td>
</tr>
</tbody>
</table>
generating C5a when they bind to cartilage. In addition, isolated synovial tissue B lymphocytes actively secrete antitype II collagen antibodies in almost all patients with seropositive RA, whereas articular cells from non-RA patients do not. Synovial fluid T cells also recognize and respond to type II collagen, and 3% to 5% of RA synovial fluid–derived T cell clones are autoreactive to the protein. Of interest, T cell responses to type II collagen, especially a dominant epitope at amino acid 263-270, are much greater if the epitope is glycosylated or if the protein is citrullinated.17

gp39 and Other Cartilage-Specific Antigens. Several other cartilage components besides type II collagen have been implicated as potential autoantigens in RA. Among the most provocative is cartilage glycoprotein gp39. Several gp39 peptides can bind to the HLA-DR*0401 molecule and stimulate proliferation of T cells from patients with RA. BALB/c mice, which are often resistant to experimental arthritis, develop polyarticular inflammatory arthritis after immunization with gp39 and complete Freund’s adjuvant. Although anti-gp39 antibodies are only detected in a small percentage of patients, it appears to be relatively specific for RA.48 Other examples of potential cartilage autoantigens include proteoglycans, aggrecan, cartilage-link protein, and other types of collagen. Proteomic analysis of RA serum using peptide arrays to detect multiple autoantibodies also identifies anti-gp39 antibodies in patients with early RA, which may be associated with less aggressive disease.

Autoimmunity to Nonarticular Antigens

Autoimmune responses in RA can also involve antigens that are broadly expressed beyond the joint.49 These antigen-antibody systems comprise a pattern of autoimmune responses that can potentially lead to synovial inflammation.

Glucose-6-Phosphoisomerase

Spontaneous inflammatory arthritis develops in K/BxN mice due to antigen-specific immunity against a seemingly irrelevant nonarticular antigen.50 Autoantibodies to the ubiquitous enzyme, glucose-6-phosphate isomerase (GPI) cause the disease, and transient synovitis also develops in normal mice injected with the serum of the affected K/BxN mice. This passive arthritis model serves as a unique tool to study antibody-dependent arthritis and requires the alternate complement pathway, Fc receptors (especially FcRγIII), and mast cells, but not T or B cells. IL-1 is more important than TNF in this model, and the IL-1 knockout mouse is almost completely protected from disease. Notably, this effect can be overcome by administration of a TLR ligand like lipopolysaccharide, which shares a downstream signaling pathway with IL-1. Other cytokines, such as IL-6, and signaling pathways, such as p38 MAP kinase and upstream kinases such as MKK3, are also required for full expression of the disease.

Immunohistochemical studies show that the target protein, GPI, adheres to the surface of cartilage, which permits local antibody binding and complement fixation. Initiation of synovial inflammation requires mast cells that increase vascular permeability and provide access to the synovium and cartilage.51 The initial phase involving mast cell and vascular permeability is augmented by antibody-mediated complement fixation when serum proteins have access to GPI-decorated cartilage.

Although the model appears on first blush to be due to a ubiquitous antigen, articular homing and display of the GPI suggest that it behaves like other arthritis models with “joint-specific” antigens. Although initial data suggested some specificity for RA, anti-GPI antibodies are detected in a relatively small percentage of RA patients and are not specific for the disease. Nevertheless, it might, along with several other antibody systems, contribute to local complement fixation and inflammation.

Heterogeneous Nuclear Ribonuceloprotein-A2 and Heavy-Chain Binding Protein. Several other autoantigens that are expressed in synovium have been characterized in RA, although they are also produced in many other locations. For instance, antibodies directed against the heterogeneous nuclear ribonucleoprotein-A2 (hnRNP-A2), sometimes called RA33, occur in about one-third of RA patients, as well as patients with other systemic autoimmune diseases. However, there may be some specificity for RA when compared with OA and seronegative spondyloarthopathies. Anti-RA33 antibodies are also produced in the TNF transgenic mouse model of rheumatoid arthritis, suggesting that proinflammatory cytokines can independently lead to a breakdown of tolerance for this particular protein.52 Of interest, RA33-positive patients with early RA tended to have less destructive disease.53 Although not especially sensitive or specific for RA when used in isolation, an algorithm involving anti-RNP-A2, RF, and anti-CP can be used to predict patients with early synovitis who will progress to erosive RA.54

Autoantibodies that bind to stress-protein immunoglobulin heavy-chain binding protein (BiP) have also been observed. About 60% of RA patients have anti-BiP antibodies, and the specificity is reportedly more than 90%. In addition to humoral responses, RA T cells can proliferate in response to this protein. Immunization of mice with BiP does not cause arthritis, but it can cross-tolerize mice and prevent collagen-induced arthritis if administered before immunization with type II collagen. BiP is normally expressed in many tissues but is markedly increased in RA synovium.

Heat Shock Proteins. The heat shock proteins (HSPs) are a family of mainly medium-sized (60 to 90 kD) proteins produced by cells of all species in response to stress. Another smaller HSP (HSP27) can serve as substrates for enzymes such as p38 MAP kinase that regulate stress responses. Immunity against HSPs can serve as substrates for enzymes such as p38 MAP kinase that regulate stress responses. Immunity against HSPs can contribute directly to synovitis and joint destruction in the adjuvant arthritis model in rats in which T lymphocytes recognize an epitope of mycobacterial HSP65 (amino acids 180 through 188). Some of these cells also recognize cartilage proteoglycan epitopes, perhaps explaining the targeting of joints.

Some patients with RA have elevated levels of antibodies to mycobacterial HSPs, especially in synovial fluid. The majority of T cell clones isolated from RA synovial fluid with specificity to mycobacterial components express the γδ–T cell receptor and do not display CD4 or CD8 surface antigens. Freshly isolated synovial fluid T cells from patients with RA briskly proliferate in response to recombinant 65-kD HSP. However, proliferation to other recall antigens
CHAPTER 69 | ETIOLOGY AND PATHOGENESIS OF RHEUMATOID ARTHRITIS

The synovium in RA is marked by intimal lining hyperplasia and sublining infiltration with mononuclear cells, especially CD4⁺ T cells, macrophages, and B cells.

Intimal lining FLS display unusually aggressive features.

Macrophages in the intimal lining are highly activated and produce many cytokines.

Lymphocytes can either diffusely infiltrate the sublining or form lymphoid aggregates with germinal centers.

Sublining CD4⁺ T cells mainly display the memory cell phenotype.

Synovial B cells and plasma cells in RA exhibit evidence of antigen-driven maturation and antibody production.

Dendritic cells can potentially present antigens to T cells in synovial germinal centers.

Mast cells produce small molecule mediators of inflammation.

Neutrophils are rarely present in RA synovium but can be abundant in synovial effusions.

The primary inflammatory site in RA is the synovium. Infiltration of synovial tissue with mononuclear cells, especially T cells and macrophages, and synovial intimal lining hyperplasia are hallmarks of the disease (Figure 69-6). In this section, the various cell lineages and histologic patterns of rheumatoid synovium are discussed.

**Synovial Intimal Lining Cells: Type A and Type B Synoviocytes**

The synovial intimal lining is a loosely organized collection of cells that form an interface between the synovium and the synovial fluid space. The intimal lining cells lack tight junctions and a definite basement membrane. The increase in cell number in RA can be quite substantial. In the normal joint, the lining is only one to two cell layers deep, whereas in RA it is often 4 to 10 cells deep. Two major cell types are found in the lining: a macrophage-like cell known as a type A synoviocyte and a fibroblast-like cell called a type B synoviocyte. The former are derived from the bone marrow and express macrophage surface markers such as CD68, Fc receptors, and CD14, as well as abundant HLA-DR, whereas the latter express little if any class II MHC antigens, are devoid of macrophage markers, and have a scant endoplasmic reticulum. The type B cells, also called FLS, express certain proteins that are unusual for mesenchymal cells including vascular cell adhesion molecule-1 (VCAM-1), CD55 (decay activating factor), cadherin-11, junctional adhesion molecule C (JAM-C), and the proteoglycan-synthesis enzyme, uridine diphosphoglucose dehydrogenase (UDPGD). The relative numbers of type A and B cells are usually similar in normal synovium. There is an absolute increase in both cell types in RA, although the percentage increase in macrophage-like cells is often greater. In addition, the type A synoviocytes tend to accumulate in the more superficial regions of the intimal lining.

Intimal lining synovial macrophages and sublining macrophages are terminally differentiated cells that presumably do not divide in the joint, and the accumulation of cells in RA is likely from the ingress of new bone marrow–derived precursors. Mesenchymally derived type B synoviocytes can divide locally in response to the proliferative factors generated by the activated immune response. Platelet-derived growth factor (PDGF), transforming growth factor-beta (TGF-β), TNF, and IL-1 produced by many different cells combine with products of arachidonic acid metabolism to induce proliferation of these cells. In addition, pluripotent mesenchymal stem cells that arise in the bone marrow and circulate through the blood can migrate into the synovium and differentiate into type B synoviocytes. Retention of macrophages in the intimal lining is probably due to expression of adhesion molecules like VCAM-1 and JAM-C on FLS.

Although local proliferation of cells in the intimal lining likely occurs, rheumatoid synovium rarely shows mitotic figures, and thymidine uptake occurs in only a small percentage of synovial cells. Using a monoclonal antibody that recognizes dividing cells, an even lower rate of cell division...
organize the intimal lining formation, but also, like T cells, B cells, and macrophages, play a critical role in the pathogenesis of inflammatory arthritis.

Two major populations of adherent cells can be readily identified when rheumatoid synovium is enzymatically dispersed and cultured in vitro. One type of cell is macrophage-like, which expresses HLA-DR antigens, Fc receptors, and monocyte lineage-differentiation antigens and is capable of phagocytosis. The macrophages, which comprise about 20% of the total cell number in the rheumatoid joint, can be derived either from the intimal lining or the sublining region. These cells are highly activated in the synovium and produce large amounts of inflammatory mediators including cytokines and arachidonic acid metabolites.

A second type is defined by the presence of antigens expressed primarily on fibroblasts and by the absence of phagocytic capability, DR antigens, or antigens of the monocytic lineage. When the enzymatically dispersed cells are cultured for several passages, this latter cell type survives and proliferates, resulting in a relatively homogeneous population of fibroblast-like cells.

Fibroblast-like cells grow slowly, with a doubling time of 5 to 7 days, and can be passaged for several months in vitro. Their doubling rate is rapid at first, perhaps due to the presence of cytokines produced by contaminating macrophages in the culture or a carryover effect from the synovial milieu. Over time, proliferation slows, and after 12 to 15 passages, (=0.05%) is apparent. A somewhat higher percentage of cells that express the cell cycle-specific antigen proliferating cell nuclear antigen (PCNA) is present in RA lining compared with OA. This correlates with the lining cell expression of theproto-oncogene c-myc, a gene that is intimately linked with fibroblast proliferation.

The architecture of the synovial intimal lining is distinct from other lining layers in the body. In contrast to serosal surfaces, the intimal lining does not include epithelial cells, it lacks a basement membrane, and has no tight junctions. Rather than serving as a discrete barrier, it is a loose association of cells that is discontinuous in some locations. Cadherin-11, from a class of adhesion proteins that are ubiquitous in various tissues, serves as the major mediator of homotypic aggregation by FLS. Immunohistochemistry shows abundant expression of this protein in the intimal lining. Its importance in the synovial architecture was confirmed in cadherin-11 knockout mice, in which the intimal lining was virtually nonexistent. Finally, cadherin-11 mediates self-aggregation of FLS in vitro. When the cells are cultured in “micromasses” made of laminin, they migrate to the surface of the particles (Figure 69-7). Macrophages cocultured in the micromasses leads to recruitment of these cells to the surface as well, thereby recapitulating a lining layer with macrophage-like and fibroblast-like cells. Blocking cadherin-11 with antibodies suppresses arthritis in the passive K/BxN model. These data suggest that FLS not only

Figure 69-7  Fibroblast-like synoviocytes are programmed to form the synovial intimal lining. Human and mouse fibroblast-like synoviocytes (FLS) spontaneously form lining structures when embedded in a three-dimensional matrix composed of the matrix protein laminin. The cells migrate to the surface of the structure over 3 weeks and resemble an intact synovial lining. The figure shows parallel experiments using human (A) and mouse (C) fibroblast-like synoviocytes. Of interest, dermal fibroblasts from human (B) or mouse (D) skin do not migrate to the surface, indicating that the synoviocytes are preprogrammed for this function and that it is not a property of all fibroblasts. (From Kiener HP, Watts GF, Cui Y, et al: Synovial fibroblasts self-direct multicellular lining architecture and synthetic function in three-dimensional organ culture, Arthritis Rheum 62:742, 2010.)
the cells gradually become senescent and ultimately cease to grow. Although it has not been proven that these cells originate solely from the synovial intimal lining, the fact that a significant percentage of cells express VCAM-1 and CD55 suggests that at least some are derived from this region. Synovial fibroblasts from RA have some characteristics reminiscent of tumors or transformed cells (see following).

Experiments examining synovial tissue and FLS gene expression profiles suggest that specific patterns correlate with synovial histopathology. FLS from highly inflamed synovium exhibited a TGF-β gene signature that has also been described in cells with features of both smooth muscle cells and fibroblasts that are found in many mucosal surfaces known as myofibroblasts. FLS appear to be involved in wound healing. A second pattern observed in cells derived from relatively noninflammatory RA tissue had increased expression of insulin-like growth factor–regulated genes. Whether the pathogenic processes in the synovium imprint the synoviocytes or vice versa is unknown. Attempts to distinguish RA and OA FLS expression profiles from each other have offered mixed results.

**Aggressive Features of RA Fibroblast-like Synoviocytes**

**Tumor-like Properties.** Rheumatoid FLS demonstrate some aggressive properties compared with cells obtained from other synovia or tissues. For instance, adherence to plastic or extracellular matrix is generally required for normal fibroblasts to proliferate and survive in culture. Although FLS typically grow and thrive under conditions that permit adherence, RA synoviocytes can also proliferate in an anchorage-independent manner. In addition, cultured RA synoviocytes can exhibit defective contact inhibition and express a variety of transcription factors, such as c-Myc, that are typically abundant in tumor cells. Poorly regulated cell growth likely occurs in vivo as well, and studies examining X-linked genes demonstrate oligoclonality in the synoviocyte population from RA, but not OA, synovium. This is especially true of cells derived from the invading pannus, which is the most aggressive region of the synovium. Increased telomerase activity, another feature of transformed tissue, is also present in RA synovium and can be observed in FGF-stimulated RA synoviocytes. Epigenetic analysis suggests overexpression of certain microRNAs associated with increased cytokine production.

**Matrix Invasion.** The most compelling evidence suggesting that RA synoviocytes are permanently altered is derived from studies in a severe combined immunodeficiency (SCID) mouse model where synoviocytes and cartilage are coimplanted. The rheumatoid cells adhere to and invade the cartilage matrix, whereas osteoarthritis and skin fibroblasts do not. Because these synoviocytes are devoid of T cells and macrophages, there is no contribution from an immune response to murine antigens. The invading cells express VCAM-1, which could potentially facilitate adherence to cartilage or chondrocytes, as well as proteases that digest the cartilage matrix. More recent studies in which explants are introduced into two locations in the SCID mice show that the rheumatoid FLS can migrate from one site to another, presumably via lymphatics and the bloodstream. This raises the intriguing possibility that imprinted, aggressive cells can “metastasize” from joint to joint and contribute to the polyarticular nature of RA.

Blocking IL-1 function in the SCID model with interleukin-1 receptor antagonist (IL-1Ra), a natural antagonist to IL-1, has no effect on synoviocyte invasion but decreases perichondrocyte matrix loss. In contrast, IL-10 or blocking type I metalloproteinase, Ras, or c-Myc expression decreases invasion. Surprisingly, overexpression of soluble TNF receptors has little effect in this model. These studies suggest that excessive production of IL-1 and underexpression of IL-10 contribute to the invasive properties of RA synoviocytes. In another study, transfecting normal synoviocytes with the human papillomavirus gene encoding E6 induced the rheumatoid phenotype. The E6 protein degrades endogenous p53 tumor suppressor protein in synoviocytes. Therefore deficient p53 function can mimic the aggressive phenotype of RA FLS. Overall, these studies paint an interesting picture of RA FLS serving as key effector cells that display unique aggressive properties in inflammatory synovitis.

**Synovial T Lymphocytes**

**Immunohistologic Patterns**

In chronic RA, the synovium contains a collection of T lymphocytes that can lead to an organizational structure that resembles a lymph node. The distribution of lymphocytes in the tissue varies from discrete lymphoid aggregates to diffuse sheets of mononuclear cells, with the most prominent location for T cells being the perivascular region. These collections consist of small, CD4+ memory T cells (CD45RO+) with scant cytoplasm. Scattered few CD8+ T cells accumulate in the aggregates, and formation of ectopic germinal centers in RA synovial tissue may depend on them. Peripheral to these foci is a transitional zone with a heterogeneous mixture of cells including lymphocytes, occasional undifferentiated blast cells, plasma cells, and macrophages (see Figure 69-6). Synovial lymphoid aggregates occur in about 15% to 20% of patients; although
originally considered specific for RA, it is now clear that they are equally prevalent in psoriatic arthritis and are even seen in some individuals with osteoarthritis. The presence of aggregates correlated with higher production of ACPAs and RF in synovium, but not with clinical disease activity.

Considerable heterogeneity exists in the histologic patterns from patient to patient and within a single joint. Synovial biopsy studies suggest that at least six sites must be evaluated to decrease the risk of sampling error to 10% to 20% or less. In situations where the synovial tissue of more than one joint from an individual patient is available, the same general histopathologic patterns are usually apparent in tissue from separate sites.

**Regulation of T Cell Aggregate Formation.** T cells often constitute 30% to 50% of cells in RA synovia, and most are CD4+. About 5% of cells are B lymphocytes or plasma cells, although in some tissues the percentage can be considerably higher. The B cells are located primarily within reactive lymphoid centers, whereas plasma cells and macrophages are often found outside these centers. This arrangement is consistent with T cell–dependent B lymphocyte activation. Plasma cells, the main immunoglobulin producers, migrate away from the germinal centers after differentiation. CD4 cells in RA synovium are in intimate contact with B lymphocytes, macrophages, and dendritic cells.

Aggregate formation is complex and involves numerous signals to orchestrate the organization of individual cell lineages. In some cases, follicular dendritic cells are present and are involved in forming true germinal centers (see later). The presence or absence of aggregates and germinal centers is a dynamic process. In one study evaluating synovial biopsies, the presence of lymphoid neogenesis was not restricted to patients with autoantibodies. Even though the cytokine and chemokine profile supports the formation of these structures, progression to fully differentiated follicles was uncommon.

Chemokines play a key role in the organization of tissues into lymphoid structures such as aggregates and germinal centers. CXCL13 and CCL21 appear to be especially important, and their expression in rheumatoid synovium correlates with the presence of this microarchitecture. The former, in particular, is produced by synovial follicular dendritic cells. Similarly, plasma cells expressing the chemokine receptor CXCR3 are present in the rheumatoid synovium. The CXCR3 ligand, Mig/CXCL9, is highly expressed by intimal lining synoviocytes and sublining cells and recruits these cells to the T cell aggregates.

The architecture of lymphoid structures in rheumatoid synovium is also regulated by members of the TNF superfamily. Lymphotoxin-α and lymphotoxin-β (LTα and LTβ, respectively), as well as lymphotoxin-related inducible ligand that competes for glycoprotein D binding to herpesvirus entry mediator on T cells (LIGHT) form trimeric molecules in various combinations that can bind to distinct cell surfaces. These three cytokines regulate the function and organization of lymphoid tissues. Deletion of either LTα or LTβ seriously impairs lymphoid development, whereas lymph nodes develop normally in LIGHT-deficient mice. In the SCID mouse model using RA tissue explants, depletion of CD8+ T cells led to loss of follicular dendritic cells, depletion of LTα1β2, and disintegration of the lymphoid follicles.

Although LTα is difficult to detect in RA, LTβ and LIGHT are present. In addition to regulating lymphoid aggregate and germinal center formation, LT can also directly stimulate FLS to produce chemokines such as CCL2 and CCL5 that attract T cells into the joint. LIGHT also enhances osteoclast differentiation and induces expression of MMP-9, TNF, IL-6, and IL-8 by macrophages. LTβ and LIGHT blockade using soluble receptors decreases the severity of collagen-induced arthritis. However, the inhibitor was not effective in the passive arthritis model in which the pathogenic antibodies were directly administered to mice. Hence the LIGHT axis has a more important role in the early phases of disease when adaptive immunity to joint antigens is developing.

**Synovial T Cell Phenotype**

**Co-stimulatory Molecules.** RA synovial T lymphocytes display an activated surface phenotype, with high expression of HLA-DR, CD69, and CD27. The co-stimulatory molecule CD28 is highly expressed by synovial T cells in RA. Its ligands, CD80 and CD86, are also displayed on antigen-presenting cells in the joint, thereby providing an excellent environment for T cell activation. The importance of the CD80/86-CD28 interaction is supported by the observation that abatacept blocks CD80/86 and is effective in RA. Oligoclonal expansion of an unusual T cell population that is CD4+CD28- is present in RA, especially patients with extra-articular disease. The cells can be cytotoxic and can respond to autoantigens. Other co-stimulatory molecule pairs such as CD40L on T cells and CD40 on antigen-presenting cells are also expressed in RA synovium.

**Adhesion Molecules.** Synovial lymphocytes also express adhesion molecules of the very late activation antigen (VLA) and lymphocyte function–associated antigen (LFA) superfamily of integrins, as well as help recruit and retain lymphocytes to the synovium. The cytokine milieu of the joint induces adhesion molecules such as intercellular adhesion molecule-1 (ICAM-1), VCAM-1, and connecting segment-1 (CS-1) fibronectin on vascular endothelium. These, in conjunction with chemokines and other chemotaxants, attract cells expressing the correct adhesion molecule counter-receptors on their surface.

**Chemokine Receptors.** Synovial T cells in RA express characteristic receptors to specific chemokines. The chemokine receptor CCR5 is the ligand for macrophage inhibitory proteins 1α and 1β and is highly expressed in the infiltrating RA T cells. This particular receptor, along with CXCR3, is preferentially found on Th1 cells, an observation that might explain accumulation of this phenotype in the rheumatoid synovium. In fact, expression of nonfunctional CCR5 alleles that protect from HIV infection might diminish the risk of RA. The chemotactic factor stromal cell–derived factor-1 (SDF-1) is also produced by synovial tissue, and its specific receptor, CXCR4, is displayed by rheumatoid synovial T cells. Other T cell phenotypes recruited to the rheumatoid synovium include Th17 cells that express CCR2 or CCR6 and produce IL-17 and CD4+CD25+ Treg cells that can suppress immune responses (see discussion of T cell subsets
T Cell Receptor Rearrangements. Numerous groups evaluated T cell receptor gene rearrangements for clues related to antigen-specific expansion. In some patients, a pattern emerged suggesting an increased number of T cells expressing \( \beta \beta\) 3, \( \beta\beta\) 14, and \( \beta\beta\) 17, especially in synovial tissue. These particular \( \beta \) genes are structurally related and are unusually susceptible to activation by superantigens. Most studies have neither found evidence for the restricted clonality of T cells in RA synovial fluid, synovial tissue, and blood nor identified expansion of different \( \beta \) or \( \alpha \) genes.

Determinants of T Cell Phenotype. The local accumulation of T cells in the joint is not necessarily related to proliferation induced by particular antigen. Instead, antigen-independent processes related to the expression of chemokines and adhesion molecules on vascular endothelium and circulating lymphocytes help determine the mononuclear cell infiltrate. Although local antigen-specific expansion can occur, it is probably responsible for a relatively small percentage of the T cell infiltrate. Cells that encounter their appropriate antigen in the correct cytokine and antigen-presenting cell environment can potentially activate other local cells through direct cell-cell contact or the elaboration of lymphokines. A high level of telomerase activity is also present in synovial lymphocytes in RA, which is a reflection of their proliferative activity and correlates with the severity of synovial inflammation. However, some data suggest that telomerase activity is actually defective in RA T cells and contributes to early senescence of the immune system. An alternative explanation for T cell activation in RA is that antigen presentation mainly occurs in central lymphoid organs. In this scenario, dendritic cells that encounter antigens in the synovium or other sites migrate to lymphatic tissue and then present antigen to naïve T cells. Once activated, these T cells enter the circulation and can home to the joint, where they can produce cytokines and activate resident cells.

Synovial T Cell Immunoreactivity

The histopathologic appearance of RA, with exuberant infiltration of the synovium with T lymphocytes, is characteristic of tissues with chronic antigen-specific responses. However, the synovium can only respond to inflammation in a limited number of ways and many chronic immune-mediated arthropathies have similar histologic patterns. In fact, the appearance of chronic arthritides that are clearly not mediated by T cells (e.g., chronic tophaceous gout) exhibits many of the same features.

The microheterogeneity of the rheumatoid synovial tissue, with different numbers and proportions of cell lineages in each area, suggests that different antigens might be presented at various locations in the synovium. For instance, type II collagen might be presented to T cells one place, proteoglycans elsewhere, and responses to HSPs or viral antigens in yet another region. Although synovial T cells display an activated phenotype, proliferative and cytokine responses are usually less than normal peripheral blood cells or even autologous peripheral blood T cells from the same patient. Spontaneous and stimulated cytokine production including Th1 factors such as IFN-\( \gamma \) and IL-2 are relatively low. Responses directed toward recall antigens are also deficient, although RA synovial T cells can proliferate briskly to certain HSPs. On the other hand, recent data implicate functional Th17 cells in the synovium, which differentiate in the presence of certain cytokines such as IL-1, IL-6, and/or IL-23 and produce cytokines in the IL-17 family (IL-17A through F).

The mechanisms of decreased responsiveness in the synovial tissue compartment have not been as extensively studied as in synovial fluid or peripheral blood, but they likely include exposure to suppressive factors (e.g., TGF-\( \beta \) or IL-1Ra), abnormal redox potentials that suppress T cell receptor signal transduction, or induction of anergy. Another contributor to local anergy is the relative lack of the co-stimulatory molecule CD80 on HLA-DR+ FLS because coculture of T cells with synoviocytes suppresses subsequent immune responses.

Synovial T cells in RA functionally resemble resting peripheral blood T lymphocytes that have been activated by cytokines rather than antigen. Both synovial and blood T lymphocytes are able to stimulate macrophages to produce TNF in a cell contact-dependent manner. This process is dependent on nuclear factor \( \kappa B \) (NF\( \kappa B \)) and is mechanistically distinct from T cell activation via the T cell receptor, which is independent of NF\( \kappa B \). Therefore the contribution of T cell to the proinflammatory cytokine milieu may be unrelated to antigen-mediated events and could result from passive activation after exposure to the cytokine environment.

Although increased immune reactivity contributes to the etiology of RA, the search for a specific rheumatoid antigen has not been fruitful. Animal models suggest that breakdown of tolerance can be due to a combination of factors early in development. For instance, mutations in the ZAP70 gene can lead to spontaneous T cell–dependent inflammatory arthritis in mice. This protein is intimately involved with transducing T cell receptor signals and activating T lymphocytes. An abnormal ZAP70 gene leads to defective positive and negative selection in the thymus and allows autoreactive T cells to escape. As a result, mice produce autoantibodies (e.g., RF and anti-dsDNA) and develop a severe destructive arthritis. The disease can be transferred by thymocytes or peripheral CD4+ T cells into syngeneic mice with a normal ZAP70 gene.

The cytokine profile of this model is quite similar to RA, and mice with deficient TNF, IL-1, or IL-6 have reduced synovitis. Of interest, IFN-\( \gamma \)-deficiency has no effect, whereas IL-10 deficiency exacerbates disease. These studies demonstrate that minor changes in the T cell receptor complex and signaling can alter T cell selection and induce T cell–dependent arthritis. This is especially relevant in light of the data implicating PTPN22 in RA because this gene is also involved with T cell receptor signaling. Despite the clear T cell dependence of the model, the cytokines implicated are remarkably similar to autoantibody-dependent models of RA. Therefore the cytokine profile in RA (see later) in many ways represents a final common pathway for a variety of autoimmune mechanisms.
Restoring T Cell Tolerance

Assuming that synovial T cell autoreactivity plays a key role in the pathogenesis of RA, one potential therapeutic approach is to restore tolerance. This concept has been somewhat problematic because the underlying defect has not been defined and could include abnormal thymic selection, poorly defined genetic factors that lead to immune hyper-reactivity, or inadequate regulatory T cell function. The lack of a specific antigen identified as pathogenic also increases the complexity of the problem. What is clear, however, is that current approaches do not “cure” RA because cessation of therapy usually leads to a recurrence of disease.

Aggressive treatment of early RA (within the year of clinical disease) with methotrexate and a TNF inhibitor could lead to long-lasting remissions. In many cases, therapy was withdrawn after 1 year of treatment and patients did not flare for at least an additional year. The mechanism of prolonged remission despite discontinuing therapy in early disease is not defined but appears to differ from clinical experience in chronic RA. Additional individualized therapies that enhance regulatory T cell function, alter costimulation, or delete pathogenic T cells are other ways to restore homeostasis and normal immune function. Examples might include inhibitors of co-stimulatory molecules like CTLA4 or ICOS or to potentially delete subpopulations of T cells that express specific T cell receptor that bind to epitopes critical for autoreactivity. However, the latter strategy based on specific T cell Vβ genes over-represented in RA joints has not been successful.

Synovial B Cells

Synovial B cell and plasma cell hyper-reactivity are viewed increasingly as key participants in the perpetuation and initiation phases of RA. This notion has been fueled by the descriptions of novel spontaneous models of arthritis in mice, such as the K/BxN model, in which loss of tolerance leads to autoantibody production, activation of innate immunity, and chronic synovitis. Furthermore, B cell–directed therapies such as anti-CD20 antibody have demonstrated efficacy in RA.

Cytokine Regulation of Synovial B Cells. Although many rheumatoid synovial tissues exhibit a diffuse infiltration with mononuclear cells, a significant percentage also have discrete lymphoid follicles populated by B cells in the sublining region. Follicular dendritic cells, B cells, plasma cells, and T lymphocytes collect in these aggregates. The germinal centers are highly organized structures in which affinity maturation occurs. B cells are present in the aggregates and express the maturation marker CD20, as well as proliferation antigens such as Ki67. The formation of these structures is dependent on several soluble and membrane-bound cytokines including lymphotoxin. B cells accumulate in lymphoid aggregates in RA synovium under the influence of a variety of chemotactic factors including CCL21 and B cell–attracting chemokine-1 (CXCL13).

A member of the TNF superfamily of cytokines known as B lymphocyte stimulator (BLyS, also called BAFF) has also been identified as a key molecule that regulates B cell differentiation. BLys binds to transmembrane activator and CAML interactor (TACI), which is present on both B and T cells. If this system is blocked using recombinant TACI-Ig to absorb the cytokine and a related B cell differentiation factor called APRIL, then the number of B cells is dramatically reduced and antibody production is decreased. The same construct is effective as a therapeutic agent in collagen-induced arthritis, a model that depends on autoantibody production.

BLyS is produced in RA synovium, especially by macrophages. Synovial lining type B synoviocytes can also release BLyS and can be induced by TNF and IFN-γ in vitro. APRIL is localized mainly to dendritic cells in synovial germinal centers. TACI-Ig, which binds both BLyS and APRIL, disrupts the germinal centers and decreases immunoglobulin receptors in SCID mice implanted with rheumatoid synovium. One clinical trial using an anti-BLyS antibody demonstrated minimal benefit in patients with RA, although a second study showed efficacy. Improved benefit might be possible with a biologic that blocks both APRIL and BLyS, although early studies showed that TACI-Ig significantly decreases serum immunoglobulin concentrations but has limited efficacy in RA.

Synovial B Cell Maturation. B cells isolated directly from germinal centers of RA synovium demonstrate a heterogeneous pattern of V-gene usage and rearrangement. The majority of VH genes are not mutated, suggesting that they are recent immigrants from the peripheral blood and are activated locally. For RF-producing cells, shared mutations containing an identical sequence throughout the variable domain of immunoglobulins have been identified in synovial tissue. Preferential utilization of a limited number of VH and DH gene segments and marked preference for a DH reading frame encoding particular hydrophilic residues have also been observed, consistent with antigen-related selection and maturation. Analysis of expressed heavy-chain variable domains supports the notion that the B cell response in RA synovium is oligoclonal. Similarly, B cell clones isolated from either RA synovium or bone marrow with “nurse-like” cells have limited VH usage.

B cells associated with follicular dendritic cells in the rheumatoid synovium can further differentiate and develop additional mutations, suggesting antigen-driven selection. Plasma cells in other areas of the synovium have distinct rearrangements compared with the B lymphocytes associated with dendritic cells. This raises the question of whether the plasma cells arise locally or migrate from the blood. Although plasma cell rearrangements are not always similar to the B cells, groups of plasma cells use similar genes, albeit with distinct mutations. Therefore the plasma cells could derive from synovial B cell clones that mutated and whose progeny proliferate and differentiate.

What is not certain is whether this process represents an ectopic lymphoid organ performing normal functions or whether it is related to autoimmunity. The presence of abundant autoantibody-producing cells in the synovium supports the latter hypothesis, although normal immune responses might also occur in the joint. Cells that produce RFs, anti–type II collagen antibody, and ACPA populate the RA synovium.

Maturation and survival of B cells depend on stromal cells and cells with nurselike properties that support
lymphocyte maturation in the thymus. The synovium of patients with RA also contains nurselike cells, which can increase expression of CD40 and class II MHC proteins on B cells. B cell survival is supported by this population of cells, whereas autoreactive clones evade deletion and produce autoantibodies. A variety of cytokines including GM-CSF, IL-6, and IL-8 are produced by RA nurselike cells, and direct contact with B cells is critical for maximal proliferation and antibody production. Of interest, cultured B cells spontaneously migrate beneath ordinary FLS, which permits them to survive in vitro for prolonged periods of time. The process depends on the interaction of the integrin α4/β1 on B cells with synoviocytes expressing CD106 (VCAM-1). Interference with the B cell–synoviocyte interaction decreases B cell survival and is one potential mechanism by which therapeutic interventions targeted at integrins might suppress autoreactivity and inflammation in RA.

**B Cell Depletion in RA.** Although the role of T cells and cytokines attracted most of the scientific interest in recent years, clinical trials with an anti-CD20 antibody (rituximab) demonstrating efficacy in RA refocused investigators on B cells and autoantibodies. The precise function of CD20 is not well defined; it is expressed on mature B cells but not plasma cells. Rituximab causes a rapid and profound depletion of peripheral blood B cells. In mice, anti-CD20 antibody depletes B cells in blood, spleen, lymph nodes, and bone marrow. The few remaining B cells in the spleen were of the B1 but not B2 phenotype. Peritoneal B cells are only partially depleted, suggesting that some sites are privileged and protect B cells despite adequate drug penetration.

In RA, serial synovial biopsies demonstrate partial B cell depletion after rituximab therapy despite the virtual absence of peripheral B cells. The clinical responses do not always correlate closely with the extent of synovial depletion, although some patients with the most impressive responses appear to have marked declines in the number of B cells in the post-treatment specimens. Surprisingly, B cell depletion does not consistently decrease autoantibody production or cytokine production in the synovium, nor does a change in ACPA or RF levels predict a clinical response. Longer-term biopsy studies suggest that plasma cells are decreased in synovium 24 weeks after the treatment with rituximab and that this correlates with decreased symptoms.

**B Cell Contribution to Synovitis.** The simplest explanation for B cell contributions to RA lies in their ability to produce autoantibodies and differentiate into long-lived plasma cells. However, autoantibody titers do not correlate well with disease activity and RA-associated antibodies are detected in individuals without evidence of synovitis. The synovial biopsy data on synovial autoantibody production also suggest that B cells play a more complex and nuanced role in the disease. For example, plasma cells might be a more relevant source of autoantibodies than CD20+ B cells.

Perhaps more important, B cells are potent antigen-presenting cells that could serve this function either in the inflamed synovium or in central lymphoid organs. In an SCID mouse model using rheumatoid synovial explants, T cell activation, lymphoneogenesis, and cytokine production depended on B cells. In the same model, treatment with TACI-Ig to inhibit APRIL and BLyS decreased synovial inflammation and IFN-γ production in tissues with germinal centers.

B cells can also make a variety of cytokines including LTα, TNF, and IL-6. Although macrophages and synoviocytes might produce greater amounts of proinflammatory factors, production by B cells in strategic locations in the joints and elsewhere might help drive synovitis. A novel subset of regulatory B cells, like regulatory T cells, could decrease inflammatory responses in RA through the production of suppressive cytokines like IL-10 and TGF-β. It is not clear how B cell depletion might select for specific B cell subsets, and it is possible that the regulatory B cells could be relatively enriched.

**Dendritic Cells**

Dendritic cells (DCs) are potent antigen-presenting cells that populate synovial tissue and synovial effusions of patients with RA. DCs generally sample the environment, especially at mucosal and skin surfaces; process antigens; and then migrate to a central lymphatic site where they present the antigens to T cells. An emerging appreciation of DC function in the etiology of RA led to the concept that the cells could be loaded with local autoantigens and migrate to central lymphoid tissues, where they can participate in the initiation of disease. DCs are also potent producers of type I interferons, which has been implicated in numerous autoimmune diseases such as lupus and RA.

Rheumatoid synovium could function like lymphoid tissues in some circumstances with mature DCs expressing CD86, CD83, and DC-LAMP localized in perivascular lymphocytic infiltrates and aggregates. Ultrastructural analysis of the synovium shows the DCs in contact with lymphocytes, where they can present antigens. The presence of DCs is not unique to RA and has been identified in other inflammatory arthritides including gout and spondyloarthropathies.

DCs are recruited to synovium, like other immunocompetent cells, through the generation of chemotactic factors and expression of the appropriate adhesion molecules. The subsequent localization with the synovium and the formation of microarchitecture such as lymphoid aggregates also depends on production of chemokines in the synovial environment. DCs express the chemokine receptor CCR7, which usually permits them to home to lymphoid tissue and orchestrate organization into the appropriate germinal centers. Synoviocytes in RA express two CCR7 ligands (CCL19 and CCL21), which provide a signal for the DCs to remain in the peripheral tissue. Immature DCs and plasmacytoid DCs are also scattered throughout the synovial sublining region. B cell–enriched lymphoid follicles, which are usually organized around follicular DCs (FDCs), are also found in some RA joint tissues. The source of FDCs is not well defined, but cultured FLS can perform FDC functions in vitro and could contribute in vivo as well. Synoviocytes derived from RA patients can bind to germinal-center B cells and suppress B cell apoptosis.

Follicular DC–containing ectopic lymphoid structures in rheumatoid synovium also express high levels of
activation-induced cytosine deaminase, a key gene that participates in immunoglobulin class switching. ACPA-positive plasma cells surround these aggregates, suggesting that the aggregates play a role in autoantibody production. The architecture and ACPA-positive cells persist when tissue is transplanted into SCID mice. Therefore synovial DCs not only help organize the microarchitecture of synovial tissue but also participate in the development of high-affinity IgG autoantibodies.

Cytokines in the rheumatoid synovium such as GM-CSF influence the proliferation and maturation of DCs. Increased numbers are found in the tissue and synovial effusions, where they comprise up to 5% of mononuclear cells. RA synovial tissue DCs do not always behave normally. For example, IL-10 suppresses dendritic cell function, in part by decreasing expression of CD86 and class II MHC molecules. However, RA dendritic cells isolated from the joint are resistant to this effect, possibly because they express lower amounts of the IL-10 receptor. Aside from presenting antigens, DCs also produce cytokines that can influence T cell differentiation in the joint including IL-12 and IL-23, which can enhance the bias toward the Th1 and Th17 (see later) phenotypes; APRIL, which enhances B cell survival; and, as noted earlier, interferons.

**Mast Cells, Polymorphonuclear Leukocytes, and Natural Killer Cells**

Mast cells are present in the synovial membranes of patients with RA and, in some patients, are located at sites of cartilage erosion. Rheumatoid synovial membranes contain more than 10 times as many mast cells than do control synovial samples from patients undergoing surgery for meniscectomy. Mast cells and histamine are also found in a majority of synovial fluid specimens from inflammatory synovitis. A detailed analysis of several indicators of proliferation and the enumeration of synovial mast cells has demonstrated strong positive correlations between the number of mast cells in synovial tissue and the degree of lymphocyte infiltration. Mast cells from RA synovium express significantly higher amounts of the C5a receptor, compared with OA synovium.

Resident mast cells in synovium respond to cytokines that stimulate mast cell growth and chemotaxis. Extracts of mast cells can induce adherent rheumatoid synovial cells to increase production of PGE2 and collagenase, and immunostaining of RA synovial mast cells demonstrates tryptase and TNF. Mast cell–derived heparin has significant effects on connective tissue. In particular, it may modulate the effects of hormones on osseous cells and thereby alter the balance of bone synthesis toward degradation.

The role of mast cells in the initiation phase of synovitis was highlighted in the passive K/BxN model in which their absence prevented disease. It is not certain if the cells are required once synovial inflammation has been established and other cell types such as neutrophils supplant the mast cells in some circumstances. For instance, the production of leukotriene B4 in the same model requires neutrophils but not mast cells in established disease. Treatment of synovial tissue explants cultures with a c-Kit tyrosine kinase inhibitor that blocks stem cell factor signaling and decreased production of TNFThese data suggest that mast cells might contribute to synovial cytokine production in established disease.

More intriguing, recent data suggest that preformed IL-17A is present in granules of synovial mast cells in RA (Figure 69-9). The conventional view of IL-17 is that it mainly comes from activated T cells. Cultured mast cells released the IL-17 when stimulated with TNF or immune complexes. There is also a correlation between the presence of IgG and FcR1ε and mast cell degranulation in rheumatoid synovium. More direct evidence of mast cell involvement in RA comes from an open label clinical trial using the c-Kit inhibitor masitinib in RA, where evidence of clinical efficacy was observed. c-Kit is a receptor tyrosine kinase that binds to stem cell factor and facilitates mast cell growth and survival.

Despite the abundance of neutrophils in RA synovial effusions, only rare polymorphonuclear leukocytes (PMNs) infiltrate the synovium. NK cells have, however, been identified in RA synovium. Cytotoxic NK cells contain large amounts of granzymes, which are serine proteases. One potentially important immunoregulatory role of NK cells is that they can stimulate B cells to produce RFs. A subset of NK cells that express large amounts of CD56 are unusually abundant in RA synovial tissue and fluid. These cells could potentially produce cytokines or enhance proinflammatory cytokine production by T lymphocytes and macrophages. There is little information on invariant natural killer T cells (iNKT), which respond to lipid antigens. However, blocking CD1d and iNKT function decreases severity of collagen-induced arthritis in mice.

**Bone Marrow Cells**

Although most attention has been directed at the synovium, cartilage, and cortical bone in RA, the subcortical bone and bone marrow also contribute to synovial inflammatory responses. Primitive bone marrow mesenchymal cells can traverse cortical bone through pores in murine collagen-induced arthritis before the onset of clinical disease, take up residence in the synovium, and produce mediators that enhance synovitis. This process is TNF dependent and is abrogated in mice that lack TNF receptors. CD34+ mesenchymal cells in rheumatoid bone marrow are more highly activated as judged by their NFκB status and their ability to differentiate into fibroblast-like cells that produce MMPs and proangiogenic factors.

Bone marrow can also contribute other relevant cells including macrophage lineage cells that migrate to the synovium and nurselike cells that support the survival of B cells. Just as the bone marrow can influence the synovium, the reverse is also true. Invasive pannus can rupture through cortical bone and invade the marrow space in some patients. When this occurs, B cell aggregates are especially prominent in the marrow and occur in an environment rich in B cell chemoattractants such as CCL21 and B cell survival factors such as BlyS.

The bone marrow changes in RA could be either secondary to synovitis or a primary event (Figure 69-10). The most common explanation for RA is that synovitis occurs first and that it expands into the marrow through...
Figure 69-9 Interleukin-17A (IL-17A) in mast cells as a source of IL-17A in rheumatoid arthritis (RA). Traditional paradigms suggest that IL-17 is mainly produced by Th17 cells in the RA synovium. However, mast cells can also produce this cytokine. Immunostaining studies showed that preformed IL-17 could be detected in mast cell granules in RA synovium. Note that the combination image on the right demonstrates IL-17A (green) colocalizing with the mast cell marker (red). This suggests a pivotal role of mast cells in the inflammatory response. MCT, mast cell tryptase. (From Hueber AJ, Asquith DL, Miller AM, et al: Mast cells express IL-17A in rheumatoid arthritis synovium, J Immunol 184:3336, 2010.)

soluble factors and direct extension (“outside-in” hypothesis). The notion that bone marrow reactivity might be a primary event and that the activated cells migrate directly to the synovium through cortical bone or via the bloodstream has received increasing attention (“inside-out” hypothesis).

The role of the bone marrow in the pathogenesis of RA is suggested by immunohistologic studies of marrow showing mature B cells and activated T cells forming aggregates, as in the rheumatoid synovium. Functional changes also occur including the ability of marrow mesenchymal cells of RA patients to support B and T cell survival and the production of IL-6 and TNF by marrow cells. Bone marrow mesenchymal cells from RA patients induce B cell growth and support the survival of T cells. Nucleike cells can also be grown from marrow, and marrow monocytes can induce RF production in vitro. There is considerable cross-talk between the synovium and the bone marrow, perhaps via the circulation, diffusion of soluble mediators, or direct migration of cells through cortical pores. The bidirectional interactions between the synovium and the bone marrow suggest that both “inside-out” and “outside-in” mechanisms can participate in RA.

Neural Elements and Rheumatoid Synovium

The synovium in RA is richly innervated with somatic afferent fibers, C-fibers that detect pain, and sympathetic neurons. In addition to providing information to the central nervous system (CNS) on the state of the synovium, peripheral nerves profoundly influence synovial inflammatory responses. For instance, neuropeptides such as substance P and vasoactive intestinal peptide can activate macrophages, increase vascular permeability, and modulate T cell phenotype. Sympathetic fibers can enhance or suppress inflammation depending on the site and timing of stimulation by modulating adrenergic tone. Although vagal branches have not been detected, their primary neurotransmitter acetylcholine is detected in tissues and can suppress synoviocyte cytokine production through the α7 nicotinic receptor. Perhaps most interesting, spinal cord circuits can directly modulate synovial inflammation. Activation of adenosine receptors in the spinal cord or inhibition of spinal p38 can markedly decrease synovial inflammation in joint destruction in rat arthritis models. Intrathecal etanercept also blocks peripheral inflammation almost as well as systemic administration of the TNF blocker.
These studies suggest that patients with “early” RA, as defined by the duration of symptoms, might, in fact, already have chronic disease and that evaluation of truly early disease might require assessment of patients long before the onset of symptoms (if this is even possible). The observation that autoantibodies are produced in RA patients years before the clinical arthritis also supports the notion that a preclinical phase can possibly precede symptomatic synovitis. In addition, profiling cytokine and chemokine levels in stored serum samples of patients who ultimately develop RA show increased levels of chemokines and cytokines, especially in the ACPA-positive population. This reinforces the concept that RA is a continuum that can progress to clinical synovitis. However, the time lines and the environmental stimuli that cause progression are not known.

**Animal Models as Surrogates for Rheumatoid Synovitis**

Animal models of RA are crucial to understand the biology of synovitis. However, they are an imperfect representation at best for this uniquely human disease. The specific inciting agents are known, the time course is compressed, and disease is induced in genetically pure strains. Despite many differences between various models and human disease, there are many similarities for pathogenic influences such as the presence of autoantibodies and, in many cases, the cytokine profile.\(^7\)

---

**Figure 69-10**  Initial synovial lesion in rheumatoid arthritis (RA). Although traditional paradigms suggest that it begins as a synovial disease, the initial insult could be at a distant site. The figure depicts two possible scenarios: (1) RA begins in the synovium and then extends into the bone and bone marrow, and (2) the initial lesion is in the bone marrow, which can extend into the synovium as cells migrate through cortical pores or through the bloodstream. (From Schett G, Firestein GS: Mr Outside and Mr Inside: classic and alternative views on the pathogenesis of rheumatoid arthritis, Ann Rheum Dis 69:787, 2010.)
The time course for cytokine and signal transduction activation is an important variable when interpreting therapeutic interventions. For example, IL-6 gene expression is high during a relatively narrow window in collagen-induced arthritis (Figure 69-11). Activation of signaling molecules such as the MAP kinases in the synovium also follows patterns that are similar to, but not identical to, RA. These issues do not invalidate the value of the models, but they do provide context for understanding why so many interventions are effective in preclinical studies but fail in human clinical trials.

**SYNOVIAL FLUID AND THE SYNOVIAL FLUID CARTILAGE INTERFACE**

**KEY POINTS**

- RA synovial effusions contain neutrophils and mononuclear cells.
- Immune complexes that contain autoantibodies such as RFs or anticitrullinated protein antibodies can fix complement, leading to the generation of chemoattractants.
- Small molecule mediators of inflammation such as prostaglandins and leukotrienes are present in RA synovial fluid.

Synovial effusions accumulate in the joints of most patients with active RA due to a substantial increase in fluid influx that cannot be removed despite an increase in lymphatic flow. There is an inverse relationship between the molecular weight of proteins and their concentrations in minimally inflamed synovial fluid. High-molecular-weight serum proteins gain access more easily to synovial fluid in inflamed joints, and the relatively high concentration of IgG in RA synovial fluid is good evidence for local synthesis of immunoglobulins. Markedly increased permeance of proteins in rheumatoid patients is also evidence of the severe microvascular lesion in rheumatoid synovitis. Newer techniques evaluating the synovial fluid proteome in RA using mass spectroscopy might provide new insights into the key mediators. More than 400 proteins in RA effusions were identified as potential biomarkers of disease activity including C-reactive protein and six members of the S100 family of calcium granule binding proteins. In general, assays for autoantibodies or assessments of exudate versus transudate in synovial fluid have little value; the concentration of serum proteins in effusions is usually about half to two-thirds the level in the blood.

**Polymorphonuclear Leukocytes**

Large numbers of PMNs accumulate in rheumatoid synovial effusions and enter via postcapillary venules in the synovium. Neutrophils adhere to activate synovial microvasculature due to their abundant surface expression of selectins and the β₂ integrins. After adherence, chemotactic agents produced by endothelium and resident synovial cells facilitate exit from the intravascular space into the tissue. Thus considering the survival time of PMNs in synovial fluid, an average rheumatoid effusion containing 25,000 PMNs per mm³ could exceed 1 billion cells each day. The ultimate fate of these cells is usually apoptosis. Neutrophil survival requires expression of the forkhead transcription factor FOXO3. Mice that lack this gene are resistant to inflammatory arthritis due to shortened PMN life span.
High concentrations of chemotactic agents in the synovial fluid in RA recruit a large number of cells to the intra-articular cavity. Few PMNs are seen in the synovium itself; once in the tissue they move rapidly to the synovial fluid, drawn by agents such as C5a generated due to complement activation, leukotriene B$_4$ (LTB$_4$), platelet-activating factor, and chemokines. CXC chemokines including ENA-78 and IL-8 is especially abundant in synovial fluid and can attract neutrophils into the intra-articular space. RA PMNs can also release chemokines such as macrophage-inflammatory protein-3α (MIP-3α) into the milieu that promote migration of additional cells into the joint space.

Once in the joint, neutrophils engage immune complexes through Fc receptors, especially FcRγI and FcRγIII, that activate spleen tyrosine kinase (Syk). This process initiates a signaling cascade that includes the MAP kinases and NFκB, cytoskeletal reorganization, release of granule content, generation of reactive oxygen and nitrogen species, and enhanced phagocytosis. Many of the cells contain immune complexes within phagosomes that include IgG and IgM along with RF and complement proteins such as C1q, C3, and C4. PMNs from synovial fluid in RA release de novo synthesized proteins including matrix proteins such as fibronectin, neutral proteinases, and IL-1. Neutrophils also secrete IL-1Ra and oncostatin M, a member of the IL-6 family. Although the amount of IL-1Ra each neutrophil produces is low compared with macrophages, the sheer number of PMNs allows them to produce large amounts in synovial effusions.

Neutrophil cells also release numerous proteases that can adversely affect the lubricating properties of synovial fluid and the integrity of the cartilage including elastase and trypsin. Neutrophil collagenase (MMP-8) can digest native collagen in cartilage, which then makes it susceptible to degradation by other MMPs such as the stromelysins. Although synoviocytes are generally considered major producers of the MMPs, animal models of arthritis suggest that neutrophil-derived proteases play a significant role in cartilage damage.

Although difficult to assess in humans, animal models demonstrate a role for neutrophils in the inflammatory processes. The passive K/BxN and collagen-induced arthritis models, which depend on autoantibodies to bind to the cartilage and fix complement, require neutrophils for full expression of the disease. Depleting neutrophils with antibodies almost completely prevents synovial inflammation in these models. In the K/BxN model, neutrophils also initiate vascular permeability that permits pathogenic antibodies to gain access to the joint space.

**Synovial Fluid Lymphocytes**

The lymphocyte subsets in synovial fluid differ from that of peripheral blood and synovial tissue. Even though synovial effusions contain an abundance of T cells, the CD4-to-CD8 ratio is actually reversed compared with that of blood or synovial tissue, with an excess of CD8$^+$ cells relative to CD4$^+$ lymphocytes. In addition, synovial tissue is nearly devoid of neutrophils, which often constitute 50% to 75% of synovial fluid cells. The percentage of regulatory T cells (CD4$^+$CD25$^+$) is also higher in synovial fluid than peripheral blood.

Synovial fluid contains T cells, which express high levels of surface HLA-DR antigens. Surface activation antigens are increased on synovial fluid lymphocytes including VLA-1 and CD69. Of CD4$^+$ cells in rheumatoid synovial fluid, most are memory cells and express CD45RO on their surface. Despite the activated phenotype, synovial fluid T cell function is usually deficient compared with peripheral blood cells. For instance, synovial fluid lymphocyte proliferation in response to mitogens or most recall antigens such as tetanus toxoid is significantly less than paired blood T lymphocytes. Mycobacterial antigens and the 60-kD HSP are exceptions because proliferation is actually greater in synovial fluid cells, perhaps because they can also activate TLRs. Cytokine production by synovial fluid T cells in vitro is also low including mitogen-induced expression of IFN-γ and IL-1.

Defective T cell responses by RA synovial fluid mononuclear cells could be due to anti-inflammatory cytokines such as IL-1Ra and TGF-β. Non-specific components of joint effusions such as hyaluronic acid can be toxic and indirectly suppress T cell activation. The mechanism of diminished T cell activation could also be due to defective TCR signaling. Articular T cells have diminished tyrosine phosphorylation of proteins after stimulation, especially the key signal transduction pathway p38 MAP kinase. Furthermore, tyrosine phosphorylation of the TCR ζ chain, an early event in TCR signaling, is low compared with peripheral blood T cells. The hyporesponsiveness of synovial fluid T cells correlates with a significant decrease in the levels of the intracellular redox-regulating agent glutathione. Restoration of intracellular glutathione increases proliferation of RA synovial fluid T cells. Therefore oxidative stress in the inflamed environment can suppress antigen-specific T cell responses.

**Platelets and Platelet Microparticles**

Platelets are a rich source of growth factors and cytokines including platelet-derived growth factor (PDGF). They are present in rheumatoid synovial effusions and can form microparticles. This process is dependent on collagen receptor glycoprotein VI. The microparticles activate cultured synoviocytes and, through an IL-1–dependent mechanism, stimulate cytokine secretion. Platelet depletion also suppresses arthritis severity in animal models.

**Intra-articular Immune Complexes and Complement Fixation**

**Synovial Fluid Immune Complexes**

Complexes containing immunoglobulins are abundant in the blood and synovial fluid of patients with RA, especially containing IgM. The most prevalent antigens in these aggregates contain IgG complexed with RF due to their ability to bind the Fc portion of immunoglobulin. Using more sensitive techniques, circulating immune complexes in RA contain up to 20 polypeptides including albumin, immunoglobulin, complement, type II collagen, fibrinogen, and acute-phase reactants, as well as DNA. These complexes can potentially fix complement, releasing...
chemotactic peptides and factors that activate neutrophils and other inflammatory cells.

**Immune Complexes Embedded in Cartilage**

The aggregates of RF, IgG, and various peptides become embedded into cartilage and other tissues in contact with synovial fluids. Electron microscopic studies demonstrate immunoglobulin complexes with damage to the cartilage matrix in the microenvironment. Immune complexes are absent under areas of cartilage invaded actively by synovial pannus, suggesting that phagocytic cells in the invasive synovium bind to and ingest the immune complexes. This possibility lends credence to the notion that immune complexes deposited in the avascular superficial layers of cartilage in the joint may serve as chemoattractants for the pannus. Immune complexes have been extracted from cartilage of RA and OA patients. Rheumatoid cartilage contains more than 40-fold more IgM and more than 10-fold more IgG than healthy cartilage extracts. IgM RF is found in the majority of RA cartilage extracts but not in OA or healthy control extracts. In addition, more than 60% of the RA cartilage extracts are positive for native and denatured collagen type II antibody.

**Synovial Fluid Complement**

Biologically active products of complement activation accumulate in synovial fluid during acute inflammation. This observation led to the concept that RA represents an intra-articular immune complex disease. This was the prevailing view of RA until the role of T cells and cytokine networks emerged in the 1980s.

The liver is the major source of complement synthesis in humans, and passive transfer of serum proteins into effusions accounts for many of the complement proteins found there. However, synovial tissue also actively produces complement proteins. Macrophages and fibroblasts produce complement proteins under the influence of cytokines such as IFN-γ, IL-1, and TNF. In situ hybridization shows that C2 is expressed in the synovial intimal lining, whereas C3 appears to be produced by synovial sublining macrophages. Analysis of synovial tissue shows that all complement genes from the classic pathway are expressed in RA, as well as in healthy synovium.

Despite the local production of complement components, the activities of C4, C2, and C3 and total hemolytic complement in rheumatoid synovial effusions are lower than in synovial fluids from patients with other joint diseases. Although the most prominent evidence of activation implicates the classic pathway, cleavage products of the alternate pathway including factor B and properdin have also been documented in RA. IgM RF appears to be a more important determinant of complement activation than IgG RF in serum and synovial fluid. Accelerated catabolism of C4 in RA and the presence of C4 fragments in the plasma correlate with titers of IgM RF.

The interactions between PMNs and the complement system are substantial. Neutrophil lysosomal lysates contain enzymes that cleave complement proteins and generate chemotactic activity such as C5a from serum. C5a, in addition to being a principal chemotactic factor in inflammatory effusions, mediates lysosomal release from human PMNs and induces cytoskeletal rearrangement. The chemotactic anaphylatoxins C3a and C5a are often present in rheumatoid effusions, as are the terminal complement components that comprise the C5b-C9 membrane attack complex. The latter is especially interesting because low levels of the complex can activate synoviocytes in vitro.

**Targeting Complement in Rheumatoid Arthritis**

Inhibiting complement proteins has obvious therapeutic potential in RA. In addition to theoretic considerations and evidence of complement consumption, a polymorphism in or near the C5 locus has been associated with increased risk of RA. For instance, extra-articular treatment with a soluble complement receptor (sCR1) inhibits joint swelling in rat antigen-induced arthritis. C5-deficient mice have decreased joint inflammation in collagen-induced arthritis and the passive K/BxN model. Absence of C3 or factor B also inhibits collagen-induced arthritis. Unlike the C5 knockout mice, which have normal antibody responses, the C3- and factor B–null animals have lower levels of anti–type II collagen antibodies. The role of C3 convertase has been explored in more detail using transgenic mice that produce the regulatory protein complement receptor 1–related gene/protein γ (Crry). These mice have no obvious phenotype and are not more susceptible to infection. However, collagen-induced arthritis is suppressed in the Crry transgenic mice.

These data suggest that classic and alternative complement pathways participate in RA. A humanized anti-C5 antibody has been evaluated in a placebo-controlled study. The antibody inhibits C5 activation and function of the C5b-C9 attack complex. Although the monoclonal antibody was well tolerated, there was only modest evidence of clinical efficacy. Similarly, a C5a receptor antagonist did not decrease inflammation in a short-term placebo-controlled synovial biopsy study even though the compound was effective in preclinical models and blood levels were in the therapeutic range.103 C5 alone is not sufficient to sustain inflammation in the rheumatoid joint and it probably participates in a redundant inflammatory network. It is not clear whether targeting other complement components such as C3 convertase would be more successful.

**Arachidonate Metabolites**

**Prostaglandins**

Arachidonic acid metabolites are produced in the inflamed joint via oxidation by cyclooxygenases (COX) to prostaglandins and thromboxanes or by lipoxygenases to leukotrienes. COX inhibitors are clearly effective in RA, although the effect is modest and is due, in part, to spinal analgesic action. Animal models of arthritis are variably sensitive to prostaglandin; indomethacin almost completely prevents adjuvant arthritis in rats but has minimal benefit in collagen-induced arthritis in mice. This serves as a reminder that results in rodent models do not always correlate with human disease.

Stable prostaglandins, especially PGE2, produce vasodilation, increase vascular permeability, and are involved centrally in fever. They also display some anti-inflammatory
activities that could account for limited efficacy. For example, the drug misoprostol, a prostaglandin analogue, has modest but significant anti-inflammatory or immunomodulatory effects. Physiologic concentrations of PGE\textsubscript{2} inhibit IFN-γ production by T cells, HLA-DR expression by macrophages, and T cell proliferation.

Production of prostaglandins in RA depends on both COX-1 and COX-2 (see Chapter 24). The former is constitutively expressed and is responsible for the normal endogenous production of prostaglandins in the joint, as well as in other tissues. COX-2, on the other hand, is an inducible enzyme responsible for increased prostaglandin synthesis in inflamed tissue. Cytokines such as IL-1 and TNF induce COX-2 gene expression by cultured synoviocytes and macrophages. COX-2 mRNA and immunoreactive protein are increased in RA synovium.\textsuperscript{104} Most nonsteroidal anti-inflammatory drugs including indomethacin and ibuprofen inhibit both enzymes. Much of the anti-inflammatory activity and analgesia in RA results from inhibition of COX-2. Targeting prostaglandin receptors such as E2 or E4 is an alternative approach that might have less deleterious effects on the gastric mucosa and cardiovascular disease. More recent data suggest that EP\textsubscript{4} can also act on DCs and T cells to enhance production of Th1 and Th17 cells.\textsuperscript{105}

Prostaglandin I\textsubscript{2}, like PGE\textsubscript{2}, can also contribute to synovial inflammation. For instance, mice lacking the prostaglandin I\textsubscript{2} receptor have significantly decreased arthritis severity in the collagen-induced arthritis model compared with wild type mice even though they make similar amounts of anticollagen antibodies.

**Leukotrienes**

LTBs are a potent proinflammatory produced by neutrophils and chemotactic for neutrophils, eosinophils, and macrophages. They also promote neutrophil aggregation and adherence to endothelium. Peripheral blood PMNs from rheumatoid patients have an enhanced capacity for LTB\textsubscript{4} production compared with normal individuals.\textsuperscript{106} In murine collagen-induced arthritis, LTB\textsubscript{4} antagonists decreased paw swelling and joint destruction, suggesting a pivotal role for this potent chemoattractant. Surprisingly, LTB\textsubscript{4} blockade has minimal efficacy in RA.

**Anti-Inflammatory Arachidonic Acid Metabolites**

Certain arachidonic acid metabolites such as 15-deoxy-delta(12,14)-PGJ(2) can bind to peroxisome proliferators, activate receptors (PPARs), and inhibit cytokine production and inflammation in animal models of arthritis. Cyclopentenone prostaglandins can also inhibit NFKB by blocking one of the key enzymes that activates this pathway, namely IkB kinase-β (IKKB). Lipoxins and resolvins represent a unique class of lipid mediators that help resolve inflammatory diseases. Lipoxins (LX) have a trihydroxytetraene structure and are produced from arachidonic acid via the lipooxygenase pathways. LXA\textsubscript{4} binds with high affinity to a G protein–coupled receptor denoted lipoxin A\textsubscript{4} receptor (ALXR). Activation of ALXR inhibits recruitment of neutrophils by attenuating chemotaxis, adhesion, and transmigration into tissues and by diminishing chemokine and cytokine production. LXA\textsubscript{4} significantly decreased cytokine and MMP expression in FLS through an NFKB-dependent mechanism.

**PERIPHERAL BLOOD LYMPHOCYTE IMMUNE RESPONSES**

Although peripheral lymphocytes are the most accessible, many investigators believe there is greater value studying cells isolated from the primary site of disease. The number of CD4\textsuperscript{+} helper T cells is mildly increased in the circulation of patients with RA, with a concomitant decrease in CD8\textsuperscript{+} lymphocytes (and an increased CD4-to-CD8 ratio). The surface phenotype of circulating T cells in RA suggests activation in some studies, but not in others. For instance, an increased percentage of αβ and γδ TCR–bearing cells might express HLA-DR and the adhesion protein VLA-4 (α\textsubscript{4}β\textsubscript{1}, integrin). The latter is especially critical in that VLA-4 plays an important role in the recruitment of cells to the synovium through interactions with VCAM-1 on endothelial cells. Other markers of activation are not necessarily elevated on RA T cells in the circulation. Therefore peripheral-blood T cells express some, but not all, phenotypic characteristics of activation. It is not clear whether this process occurs in the peripheral or central lymphoid organs or whether cells are activated in the synovium and re-enter the circulation via the synovial lymphatics.

Immunoregulatory dysfunction has been described in RA peripheral lymphocytes. One of the early observations was the inadequate control of EBV-infected B lymphocyte growth due to a defect in T cell function in RA. The abnormal T cell response could be correlated somewhat with disease activity, but it was also noted that the abnormality was present in T cells of some patients with inflammatory arthropathies other than RA.\textsuperscript{107} IFN-γ and IL-2 production is significantly suppressed in these RA lymphocyte cultures under certain conditions.

T cell diversity and maturation are abnormal in RA. Whereas thymic output normally decreases with age, this process appears to be accelerated in RA.\textsuperscript{108} The presence of T cell receptor rearrangement excision circles (TREC) is a measure of thymic release of mature T cells. Using this parameter, the thymic output in RA may decline prematurely. Similarly, telomere attrition suggests inappropriate “aging” of the T cells. This could be due to a primary defect in peripheral T cell homeostasis or due to impaired thymic function with increased T cell turnover due to chronic immune stimulation. This concept is supported by the observation that telomeric length in RA T cells is shorter than in normal controls and more closely resembles older populations.

Activated B lymphocytes are also present in the peripheral blood of patients with RA. The number of circulating B cells that spontaneously produce RF and other autoantibodies is significantly higher in RA compared with normal individuals. B cells that are enriched in autoantibody production are characterized by a surface determinant CD5. This antigen is normally expressed by T cells, but it is also displayed by B cells and a small number of immature B cells in adults. RA patients with normal circulating numbers of lymphocytes show an abnormal
kappa-to-lambda-chain analysis compared with controls, implying oligoclonal B cell proliferation. It is not known whether this reflects expansion of the restricted number of clones capable of producing RF or whether an inciting antigen is something other than IgG and related specifically to RA. On the other hand, normal and RA peripheral blood B cells have equal numbers of B cells that produce IgM anti-type II collagen antibodies. The B cells that accumulate in the synovial fluid, however, produce IgG antibodies that are more likely to be pathogenic.

Attempts to characterize the gene expression profiles in the peripheral blood cells of patients with RA have met with mixed results. In some cases, RNA transcripts can potentially distinguish between RA and psoriatic arthritis and include differential expression of tumor suppressors, MAP kinases, and other proinflammatory proteins. Other limited subsets of genes have been proposed as markers for disease activity or to predict response to targeted therapy. However, additional confirmatory studies are necessary to determine whether these patterns are consistent.

**ROLE OF T CELL CYTOKINES**

**KEY POINTS**

Several subsets of T cells have been implicated in the pathogenesis of RA.

Relatively low levels of T cell cytokines are present in RA synovium.

The T cell cytokines that are present, such as IFN-γ and IL-17, can be produced by Th1 cells or Th17 cells.

Regulatory T cell function, which suppresses activation of other T cells, might be low in RA synovium.

The contribution of T cells to synovial inflammation can be through antigen-independent mechanisms such as direct cell-cell contact with macrophages.

The cytokine milieu in RA is rich in factors produced by many cell lineages. Factors produced by T lymphocytes are surprisingly low in RA, whereas those generated by macrophages and by synovial fibroblasts are markedly increased (Table 69-5). The production and contribution of T cell-derived cytokines can be examined based on functional lymphocyte subsets.

**T Helper Type 1 Cell Cytokines**

Extensive investigation into the cytokine profile of RA suggest a bias toward the Th1 phenotype, which produces cytokines such as IL-2 and IFN-γ and expresses the chemokine receptors CXCR3 and CCR5. Many T cell clones derived from RA synovial tissue produce a Th1 cytokine pattern. Considerable data have accumulated on the relative abundance and function of the prototypic Th1 cytokine, interferon-γ (IFN-γ), which is the most potent inducer of HLA-DR antigens. IFN-γ also induces adhesion molecules on the surface of endothelial cells and can help recruit inflammatory cell accumulation at sites of injury. One of the most important functions of IFN-γ is its capacity to alter the balance of extracellular matrix synthesis and degradation by decreasing collagen synthesis and inhibiting matrix metalloproteinase (MMP) production by cytokine-stimulated cultured FLS.109

Relatively low concentrations of IFN-γ have been detected in RA joints, well below the amounts needed to induce HLA-DR expression on monocytes. The relative lack of IFN-γ in rheumatoid joints has been observed at the level of mRNA using a variety of techniques.

### Table 69-5 Production of Selected Synovial Cytokines in Rheumatoid Arthritis according to Cellular Source

<table>
<thead>
<tr>
<th>Cellular Source</th>
<th>Level of Production in Rheumatoid Arthritis Synovium&lt;sup&gt;+&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T Cells</strong></td>
<td></td>
</tr>
<tr>
<td>Interleukin-2</td>
<td>–</td>
</tr>
<tr>
<td>Interleukin-3</td>
<td>–</td>
</tr>
<tr>
<td>Interleukin-4</td>
<td>–</td>
</tr>
<tr>
<td>Interleukin-6</td>
<td>±</td>
</tr>
<tr>
<td>Interleukin-13</td>
<td>±</td>
</tr>
<tr>
<td>Interleukin-17A, F</td>
<td>–</td>
</tr>
<tr>
<td>Interferon-γ</td>
<td>–</td>
</tr>
<tr>
<td>TNF</td>
<td>–</td>
</tr>
<tr>
<td>Lymphotoxin-α</td>
<td>–</td>
</tr>
<tr>
<td>RANKL</td>
<td>+</td>
</tr>
<tr>
<td>GM-CSF</td>
<td>–</td>
</tr>
<tr>
<td><strong>Macrophages&lt;sup&gt;†&lt;/sup&gt;/Fibroblasts&lt;sup&gt;‡&lt;/sup&gt;</strong></td>
<td></td>
</tr>
<tr>
<td>Interleukin-1</td>
<td>+++</td>
</tr>
<tr>
<td>Interleukin-1Ra</td>
<td>+</td>
</tr>
<tr>
<td>Interleukin-6</td>
<td>+++</td>
</tr>
<tr>
<td>Interleukin-10</td>
<td>+</td>
</tr>
<tr>
<td>Interleukin-12</td>
<td>+</td>
</tr>
<tr>
<td>Interleukin-15</td>
<td>++</td>
</tr>
<tr>
<td>Interleukin-16</td>
<td>+</td>
</tr>
<tr>
<td>Interleukin-18</td>
<td>++</td>
</tr>
<tr>
<td>Interleukin-23</td>
<td>+</td>
</tr>
<tr>
<td>Interleukin-32</td>
<td>+</td>
</tr>
<tr>
<td>Interleukin-33</td>
<td>++</td>
</tr>
<tr>
<td>TNF</td>
<td>++</td>
</tr>
<tr>
<td>M-CSF</td>
<td>+</td>
</tr>
<tr>
<td>GM-CSF</td>
<td>++</td>
</tr>
<tr>
<td>BlyS</td>
<td>++</td>
</tr>
<tr>
<td>LIGHT</td>
<td>++</td>
</tr>
<tr>
<td>RANKL</td>
<td>+</td>
</tr>
<tr>
<td>TGF-β</td>
<td>++</td>
</tr>
<tr>
<td>Chemokines (e.g., IL-8, MCP-1)</td>
<td>+++</td>
</tr>
<tr>
<td>Fibroblast growth factor</td>
<td>++</td>
</tr>
<tr>
<td>Interferon-β</td>
<td>++</td>
</tr>
<tr>
<td><strong>Dendritic Cells</strong></td>
<td></td>
</tr>
<tr>
<td>Interferon-α</td>
<td>+</td>
</tr>
<tr>
<td>Interleukin-12</td>
<td>+</td>
</tr>
<tr>
<td>Chemokines (e.g., CXCL13 and CCL21)</td>
<td>+</td>
</tr>
<tr>
<td><strong>Mast Cells</strong></td>
<td></td>
</tr>
<tr>
<td>TNF</td>
<td>+</td>
</tr>
<tr>
<td>Interleukin-17A</td>
<td>+</td>
</tr>
</tbody>
</table>

*+, absent or very low concentrations; +, present.
†Tissue macrophages or type A synoviocytes.
‡Tissue fibroblasts or type B synoviocytes.
BlyS, B lymphocyte stimulator; GM-CSF, granulocyte-macrophage colony-stimulating factor; MCP-1, monocyte chemoattractant; M-CSF, macrophage colony-stimulating factor; RANKL, receptor activator of NFκB ligand; TGF-β, transforming growth factor-β; TNF, tumor necrosis factor.
Immunohistochemical analysis clearly demonstrates IFN-γ in small numbers of RA synovial T cells, although the percentage is far less than in chronically inflamed tonsils.

Another major Th1 cytokine, IL-2, is a T cell–derived cytokine that serves as an autocrine or paracrine T cell growth factor. Although it was originally reported to be present in synovial fluid using biologic assay, more specific immunoassays showed that IL-2 is detected in only a small percentage of RA synovial effusions and synovial tissues and, when present, is only found in low concentrations. TNF, GM-CSF, and IL-6 can be expressed by Th1 and Th2 cells. All three are abundant in synovial fluid and produced by RA synovial tissue. However, the primary sources of these cytokines in the rheumatoid joint are macrophages and fibroblasts rather than T cells.

Targeted therapeutics to block Th1 function or recruitment to RA synovium is an attractive concept. Several CCR5 antagonists have been developed because this chemokine receptor serves as a binding protein for HIV and might be used to decrease Th1 cytokine levels in the joint. Some of these have demonstrated efficacy in animal models of arthritis. However, the role of IFN-γ is actually quite complex because administration of the cytokine does not exacerbate disease and, if anything, results in improvement in some patients. IFN-γ knockout mice or IFN-γ receptor deficiency can exacerbate collagen-induced arthritis in mice and serve as a reminder that the effects of cytokines can be highly variable depending on the model and timing of expression.

T Helper Type 2 Cell Cytokines

Although relatively low levels of Th1 cytokines are readily detected, Th2 cytokine levels are exceedingly low in RA. Using immunoassays, IL-4 and LTα are generally not detected in RA synovial fluid. In situ hybridization also shows little or no IL-4 mRNA in RA synovial tissue. When sensitive nested reverse-transcriptase polymerase chain reaction (RT-PCR) techniques are used on synovial biopsies, Th2 cytokines IL-4 and IL-13 are absent in RA, whereas both IFN-γ and IL-12 (a cytokine that induces T cell maturation toward the Th1 phenotype) are usually present. The Th2 cytokine IL-10, which has potent anti-inflammatory activities, is expressed in RA synovium. However, macrophages rather than T cells are the major producers in RA. Some data suggest that synovial fluids in early synovitis have high concentrations of Th2 cytokines such as IL-4 and IL-13 and that this pattern distinguishes patients who progress to RA. As the disease evolves in these patients, the Th2 signature diminishes.

T Helper Type 17 Cytokines

The proinflammatory cytokine produced by Th17 cells, IL-17, exists as six isoforms (IL-17A through F). IL-17A and, to a certain extent, IL-17F mimic many of the activities of IL-1 and TNF with respect to FLS function including induction of collagenase and cytokine production. IL-17A is present in modest, but functionally relevant, concentrations in RA synovial effusions. More important, T cell–derived IL-17A in synovial tissue can synergize with IL-1 and TNF by activating synoviocytes to produce matrix MMPs and other proinflammatory cytokines. IL-17 receptors IL-17RA and IL-17RC are expressed by synoviocytes and, when engaged, can activate the transcription factor NFκB and initiate an inflammatory cascade. In addition to its effect on mesenchymal cells, IL-17 can participate in bone erosion by enhancing osteoclast activation. Bone resorption in an in vitro model using synovial explants and bone shows that blockade of IL-17, IL-1, and TNF is more effective than blocking the individual factors.

Immunohistochemistry initially suggested that IL-17 is mainly present in sublining T cells, although it could also be derived from mast cells. In some studies, IL-17 is present in the synovium of a minority of RA patients. Because immunoreactive IL-17 can be detected near the erosive front of pannus, it could also participate in extracellular matrix destruction. Animal models of arthritis demonstrate that IL-17 inhibition is anti-inflammatory and protects animals from bone and cartilage destruction. IL-1, IL-23, and/or TGF-β in the joint can potentially enhance Th17 cell differentiation, although the precise cytokine environment for this process in humans and mice can differ. BlyS also has the potential to enhance Th17 cell differentiation and promote inflammatory arthritis. All of these cytokines are present in RA synovium, thereby providing an excellent milieu to enhance the generation of these cells.

The role of the IL-17 family is clear in animal models of arthritis, where blocking IL-17A in particular has dramatic effects on inflammation and matrix destruction. The data in humans are less robust, perhaps because the segregation of human T cells into clear subsets is not as well defined. Early clinical studies with anti-IL-17A antibodies in RA are encouraging, although the benefit is modest compared to TNF inhibitors.

Regulatory T Cells (Tregs)

The role of classically defined inducible regulatory T cells (CD4+CD25+CD127−) in RA remains poorly defined. Deficiency of this subset, which produces IL-10 and TGF-β and regulates other T cells through cell contact, could contribute to autoimmunity. Studies of RA peripheral blood demonstrate normal numbers of Tregs, although the CD4+CD25+ subset appears to accumulate in rheumatoid synovial effusions. However, peripheral blood Treg responses in vitro to stimuli such as anti-CD3 and anti-CD28 antibody displayed an anergic phenotype and were unable to suppress T cell or monocyte cytokine production. Of interest, this abnormality might be reversed when patients are treated with TNF inhibitors. The effect of treatment on synovial Tregs was also observed in a synovial biopsy study. The number of FoxP3+ cells was relatively low in T cell infiltrates identified in rheumatoid synovial biopsies. Interestingly, the number decreased further after intra-articular injection with corticosteroids.

These data suggest that abnormal Treg function could be secondary to cytokine imbalance in RA rather than a primary event. Treatment can have a significant effect on the number and function of Tregs, which makes evaluation of their role in patient samples difficult to interpret. Nevertheless, enhancing or restoring Treg activity as a therapeutic intervention could potentially downregulate other T cells and cytokine production.
Enhancing the number or activity of Tregs has been used to treat animal models of arthritis. Antigen-induced arthritis is exacerbated when CD4+CD25+ cells are depleted and suppressed when they are passively transferred to affected animals. Treg function is low in some inflammation models. Methotrexate also increases Treg function in collagen-induced arthritis. Administration of neuropeptides such as vasoactive intestinal peptide (VIP) appears to suppress collagen-induced arthritis by enhancing the Treg function in synovium and lymph nodes.

**T Helper Cell Cytokine Imbalance in RA**

The relative abundance of Th1 cells suggests that the synovium resembles a Th1-like delayed-type hypersensitivity reaction and/or a Th17 autoimmune environment. Th2 cytokines and cellular responses that normally suppress Th1 activation are nearly absent, thereby raising the possibility that the lack of T cell activation along the Th2 pathway in RA contributes to disease perpetuation. For example, addition of exogenous IL-10 or IL-4 to cultures of synovial tissue cells or synovial tissue explants suppresses synthesis of proinflammatory cytokines and MMPs by cultured RA synovial tissue explants. The inhibitory action of IL-4 might be mediated by decreased c-Jun and c-Fos expression, which is required for efficient production of MMPs and cytokines. In addition, IL-10 and IL-4 increase the release of other anti-inflammatory cytokines such as IL-1Ra by synovial cells. Although IL-10 protein is present in RA synovial fluid and the gene is expressed by synovial tissue cells, in vitro studies of cultured synovial cells suggest that not enough IL-10 is produced to suppress IFN-γ production.

The notion that Th1 and Th17 cytokines initiate and perpetuate arthritis, whereas Th2 cytokines are suppressive, is supported by studies in animal models. For instance, IL-4 and IL-10 were administered individually or in combination in collagen-induced arthritis. The cytokines had modest or no benefit when used separately, but together the effect was impressive. Clinical improvement correlated with decreased synovial IL-1, TNF, and cartilage destruction. Anti–IL-10 antibody therapy in collagen-induced arthritis accelerates disease. The complexity of cytokine networks in inflammatory arthritis is underscored by studies on the role of IL-12 in collagen-induced arthritis. In early arthritis, IL-12 administration increases the incidence of collagen-induced arthritis, whereas anti–IL-12 is beneficial. However, in late disease, IL-12 administration suppresses arthritis and anti–IL-12 causes an exacerbation. As noted above, administration of anti–IL-17A antibody to block the prototypical Th17 cytokine has only modest benefit. Alternative approaches that inhibit both IL-17A and F, such as soluble receptor constructs, might increase efficacy. The role of Th17 cells is clearer in other inflammatory diseases, most notably psoriasis, where IL-17A blockade demonstrates a robust response.

Although the notion that enhancing Th2 cytokines is attractive, a clinical trial using IL-10 in RA did not demonstrate significant clinical benefit or improvement in histologic evidence of synovial inflammation. It is possible that combinations of Th2 cytokines will be required to coordinate a maximum effect.

**Activation of Synovial Cells by Cell-Cell Contact with T Lymphocytes**

Even though T cell activation is relatively modest in rheumatoid synovium, direct cell-cell contact permits these cells to participate in synovial cytokine networks and matrix destruction. Membranes prepared from activated T cells can directly stimulate macrophages and FLS to produce cytokines and MMPs. The membrane constituents that regulate this process vary, depending on the particular culture conditions, but include adhesion molecules such as LFA-1 and membrane-bound TNF. Hence a T cell displaying these proteins can potentially contribute to macrophage and fibroblast activation in an antigen-independent fashion. One of the best-characterized consequences of this pathway is the ability of T cells to enhance synovial macrophage TNF production in a contact-dependent manner after exposure to macrophage-derived IL-15.

The concept that lymphocytes can activate cells in the environment through direct contact suggests an unanticipated role for T cells in RA. The traditional paradigm assumes that T cells in the joint respond to a pathogenic stimulus and subsequently drive an antigen-specific response. However, cell-cell contact influences can be antigen-independent and only require colocalization of memory T cells with synoviocytes or macrophages. Because T cells with a memory phenotype accumulate into the joint due to the release of chemoattractants, there is no requirement for a specific arthritogenic antigen to initiate the process. Instead, activation of innate immunity by nonspecific stimuli permits subsequent ingress of the correct T cell phenotype to engage resident synovial lining cells. The combination of adaptive immune responses and antigen-independent stimulation could contribute to the diversity of individual responses to targeted therapy and the relative paucity of complete remissions.

**ROLE OF MACROPHAGE AND FIBROBLAST CYTOKINES**

**KEY POINTS**

- Macrophage and fibroblast cytokines are abundant in RA synovium.
- Cytokine networks involve proinflammatory cytokines such as IL-1, TNF, IL-6, IL-15, IL-18, GM-CSF, and IL-33. These and many other factors help perpetuate synovial inflammation.
- Chemokines that recruit inflammatory cells into the joint are commonly produced by macrophages and fibroblasts.
- Anti-inflammatory cytokines such as IL-1Ra and IL-10 are produced in rheumatoid synovium, although in amounts insufficient to suppress proinflammatory cytokine function or production.

Virtually every macrophage and fibroblast proinflammatory mediator investigated in the RA synovium is abundant. In this section, some of the major cytokines and effectors produced in the joint are described, with an emphasis on the prevalence of macrophage and fibroblast products as
major forces that perpetuate RA. Macrophages, in particular, are the most vigorous producers of cytokines. The role of macrophage and fibroblast cytokines in the pathogenesis of RA was initially suggested by studies involving patient samples, as well as preclinical experiments in animal models.

**Proinflammatory Macrophage and Fibroblast Cytokines**

**Interleukin-1 Family**

The IL-1 family is a ubiquitous group of polypeptides with a wide range of biologic activity; they include IL-1α, IL-1β, IL-18, IL-33, and IL-1Ra, which is a natural inhibitor of IL-1 (see Suppressive Cytokines and Cytokine Antagonists later for a description of IL-1Ra). Abundant animal data indicate that IL-1 can serve as a key regulatory factor in inflammatory arthritis. For instance, recombinant IL-1β induces the accumulation of PMNs and mononuclear leukocytes in the joint space and the loss of proteoglycan from articular cartilage when injected directly into rabbit knee joints. Transgenic mice that overexpress IL-1 also develop inflammatory arthritis, whereas mice that lack the natural IL-1 antagonist IL-1Ra have increased susceptibility to collagen-induced arthritis. In most cases, IL-1 blockade in animal models modestly decreases synovial inflammation while markedly diminishing bone and cartilage destruction.

**Interleukin-1.** Synovial macrophages are the most prolific source of IL-1 in the joint, and nearly half of all macrophages in the RA synovium express IL-1β. Immunohistologic studies confirm this, with especially abundant IL-1 protein in synovial lining macrophages adjacent to type B synoviocytes and in sublining macrophages near blood vessels. The IL-1 in the lining can subsequently activate type B synoviocytes to proliferate and secrete a variety of mediators. A broad range of stimuli are capable of inducing IL-1 production by macrophages; for example, immunoglobulin Fc fragments and, to a lesser extent, immune complexes, can generate IL-1 production by rheumatoid synovial macrophages. Collagen fragments can induce IL-1 production, and type IX collagen, which has been found only in articular cartilage and localized into intersections of collagen fibrils, is a potent inducer of IL-1 by human monocyte. Within the rheumatoid joint, IL-1 induces fibroblast proliferation; stimulates the biosynthesis of IL-6, IL-8, and GM-CSF by synovial cells; and enhances collagenase and prostaglandin production. It increases glycosaminoglycan release in human synovial fibroblast cultures, although the effect of IL-1 on the production of intact proteoglycan molecules by intact articular cartilage explants can be the opposite. IL-1 induces a number of adhesion molecules on FLS and endothelial cells including VCAM-1 and ICAM-1 and enhances bone resorption.

IL-1 has been implicated in RA, and inhibition of this mediator using various IL-1 targeted biologic agents has modest anti-inflammatory activity. Even combinations of IL-1 and TNF-directed therapy does not provide benefit beyond TNF blockade alone. In contrast, diseases with a well-defined role for IL-1 such as Still’s disease, familial cold autoinflammatory syndrome, Muckle-Wells syndrome, and crystal diseases such as gout have a much better clinical response. These data suggest that IL-1 plays a modest role in the clinical manifestations of RA. Its contributions to cartilage and bone destruction are still uncertain, but preclinical models suggest that the protective effects of TNF blockade are mediated through IL-1.

One explanation for the relatively modest benefit of IL-1 in RA relates to its signaling mechanisms. IL-1 activates NFκB through the kinase MyD88, which is the same pathway as TLRs. When IL-1 signaling is blocked, it is possible that TLR ligands in the synovium including exogenous ones such as peptidoglycan or endogenous ones such as heat shock proteins can provide the stimulus required to overcome IL-1 blockade. This concept was tested in the passive K/BxN model where mice lacking the IL-1 receptor had markedly decreased arthritis severity. When small amounts of the TLR4 ligand lipopolysaccharide were administered, robust synovitis ensued. If a similar system occurs in RA, then IL-1 blockade would be unable to control synovitis as long as TLR signaling remains intact.

**Interleukin-18.** In addition to IL-1α and IL-1β, a homologous protein in the IL-1 family known as IL-18 has been implicated in RA. This cytokine was originally defined by its ability to bias the immune response toward the Th1 phenotype, especially in the presence of IL-12. In collagen-induced arthritis, IL-18 inhibition significantly attenuates disease. Of particular interest, the same effect was observed in IFN-γ knockout mice, indicating that other non–Th1-related activities of IL-18 might be important. Subsequent studies showed that IL-18 induces GM-CSF, nitric oxide production, and TNF expression by synovial macrophages. IL-18 is expressed by RA synovial tissue, especially by synovial fibroblasts and macrophages, and its production is markedly increased by TNF and IL-1β. A natural inhibitor, the IL-18 binding protein can potentially be used as a therapeutic agent to block both the proinflammatory effects and pro-Th1 effects of IL-18. One potential concern for IL-18 as a target for RA is the fact that the IL-1 convertase inhibitor had only modest benefit. IL-18, like IL-1β, is processed by this enzyme in order to produce biologically active cytokine. Therefore an IL-1 convertase inhibitor theoretically should block both IL-1β and IL-18 release by macrophages.

**Interleukin-33.** IL-33 is a novel cytokine that signals through the ST2 surface receptor. Like high-mobility-group box 1 (HMGB1), it is an “alarmin” that provides a danger signal due to tissue damage and necrosis with release of intracellular contents. Blocking IL-33 decreases inflammation in several animal models of arthritis including collagen-induced arthritis and antigen-induced arthritis. Mast cells are activated by IL-33 and can subsequently release their content. Thus mast cell–dependent models such as passive K/BxN are less severe in ST2−/− mice. More intriguing, its ability to trigger mast cells could play a role in the release of mast cell products such as TNF and IL-17A in RA synovium. IL-33 can also bias T cells toward the Th2 phenotype, so the fact that it is expressed in RA synovium and synovial effusions raises some questions about its precise function in long-standing disease.
**Tumor Necrosis Factor**

The TNF superfamily is an extended group of related genes that play a major role in inflammation, immune responses, cell survival, and apoptosis. At least 19 members of the family have been identified, with TNF identified as the eponymous member. Each cytokine has its own preference for cell surface receptor, although there is some promiscuity of receptor binding and functional overlap. The TNF superfamily members exhibit conserved amino acid sequences, suggesting a single ancestral gene. Many include type II membrane protein characteristics and can be released from cell surfaces after proteolytic cleavage. A C-terminal conserved domain called the TNF-homology domain is also shared by several superfamilial members. The active forms of the proteins are homotrimers, except for LTα and LTβ, which can form either heterotrimers or homotrimers. Several members have been discussed earlier in the relevant sections. In this section, one of the “founding” members of the superfamily is discussed due to its critical role in synovial inflammation.

TNF is a pleiotropic cytokine that has been implicated as a key proinflammatory cytokine in RA and detected in rheumatoid synovial fluid and serum. It is produced as a membrane-bound protein that is released from the cell surface after proteolytic cleavage by TNF convertase, a membrane MMP. IL-1 and TNF have many similar activities including the ability to enhance cytokine production, adhesion-molecule expression, proliferation, and MMP production by cultured synoviocytes. In some systems, the effects of these two agents are synergistic. Although they share many functions and signal transduction pathways, IL-1 and TNF use distinct surface receptors and intracellular signaling pathways.

The efficacy of TNF inhibitors in RA demonstrates its critical role in the disease; heterogeneity of the rheumatoid process is also apparent because only about one-third of patients have a dramatic response to TNF inhibitors. Efficacy requires continuous therapy because cessation typically leads to a flare of disease. Perhaps more interesting, evidence is beginning to accumulate, suggesting that early aggressive therapy with anti-TNF agents can induce long-term remissions even after therapy is withdrawn. This exciting notion suggests that interventions in the earliest stages of disease could prevent the establishment of chronic synovitis.

TNF, like IL-1, stimulates collagenase and PGE₂ production by human synovial cells, induces bone resorption, inhibits bone formation in vitro, and stimulates resorption of proteoglycan and inhibits its biosynthesis in explants of cartilage. In situ hybridization and immunohistochemical studies show that TNF is primarily produced by synovial macrophages in RA.

Animal models have also supported the general role played by TNF in inflammatory arthritis. For instance, overexpression of TNF in transgenic mice leads to an aggressive and destructive synovitis. In fact, the arthritis also spontaneously occurs in transgenic mice that express only a membrane-bound form of TNF on T cells. TNF blockade is an effective anti-inflammatory agent in many animal models of arthritis, although the effects on bone and cartilage destruction are less prominent than with IL-1 inhibitors and are likely mediated through downstream inhibition of IL-1 production.

TNF inhibition in RA significantly decreases extracellular-matrix destruction as measured by radiographic progression. It is not clear why the bone-protective effects are more prominent in humans than in the animal models. TNF blockade is also more effective in animal models when combined with an IL-1 inhibitor, supporting the additive or synergistic relationship between the two cytokines in animal models. However, this has not been observed in RA.

The mechanism of action for TNF blockers is distinct from other biologic agents. For instance, individuals with an inadequate response to a TNF inhibitor are still likely to respond to either rituximab or abatacept. This supports the notion that multiple independent pathways can contribute to the pathogenesis of RA and the heterogeneous nature of the disease. Surprisingly, clinical responses do not always correlate with protection of the extracellular matrix. Patients with little or no clinical improvement in signs and symptoms of RA still have significant delay or arrest of joint damage. This observation supports the contention that inflammation and destruction could have distinct pathogenic mechanisms such as inhibition of osteoclast maturation, even though synovial inflammation continues unabated.

**Interleukin-6 Family**

IL-6 is a complex cytokine produced by many cell types including T cells, monocytes, and FLS. Originally defined by its B cell–stimulating properties, it induces immunoglobulin synthesis in B cell lines, is involved in the differentiation of cytotoxic T lymphocytes, and is a major factor in the regulation of acute-phase response proteins like C-reactive protein by the liver. The IL-6 receptor includes a common chain (gp130) that is shared with other cytokines and an IL-6-specific chain (IL-6R). IL-6R can be shed from cell surfaces, bind to IL-6, and deliver it to cells that lack IL-6 receptors by combining with gp130. IL-6 signaling proceeds through the Janus kinases (JAKs), especially JAK1, and phosphorylation of STAT3.

A striking correlation between serum IL-6 activity and serum levels of acute-phase reactants such as C-reactive protein, α₁-antitrypsin, fibrinogen, and haptoglobin occurs in patients with RA. Very high levels of IL-6 are present in RA synovial fluid, and synovial cells in culture from diverse inflammatory arthropathies produce IL-6. In situ hybridization of synovial tissue also shows IL-6 mRNA in the intimal lining, and immunohistochemistry studies show IL-6 protein in the lining and sublining regions. Although many synovial macrophages express the IL-6 gene, the majority of IL-6 appears to be produced by type B synoviocytes.

The pivotal role of IL-6 in RA has been demonstrated by clinical trials using a monoclonal antibody that binds to IL-6R. The clinical responses are similar to TNF inhibitors including protective effects on bone and cartilage damage. Surprisingly, IL-6R antibody is also effective in patients who have not responded to TNF blockade. This observation is surprising considering that TNF is considered upstream of IL-6 in the cytokine cascade and that TNF inhibition markedly decreases IL-6 levels in RA. Neutropenia, increased
liver enzymes, and altered blood lipids have been observed with IL-6 inhibition, although the relative importance of these observations is still uncertain. JAK inhibitors have also demonstrated efficacy in RA; this might be due, in part, to the effect on IL-6 signaling.

Cytokines with structural similarity to IL-6 and that share surface-receptor subunits have also been implicated in RA. Several of these, IL-11, leukemia inhibitory factor (LIF), and oncostatin M, are expressed by rheumatoid synovium and can be detected in synovial effusions. The biologic effects of these factors are complex and can be either protective (e.g., by increasing expression of protease inhibitors such as tissue inhibitors of metalloproteinase [TIMP]) or proinflammatory (e.g., by increasing expression of chemokines or MMPs) depending on the culture conditions or the specific model evaluated. This dichotomy among the family members is demonstrated by the fact that IL-11 administration ameliorates collagen-induced arthritis, whereas antibodies to oncostatin M are protective.

**Interleukin-12 Family**

The IL-12 family includes a group of cytokines that play a key role in the differentiation of T cells and inflammation. IL-12 is a heterodimeric cytokine with two subunits (e.g., p35 and p40) encoded on separate genes and produced by antigen-presenting cells. IL-23 and IL-27 have similar heterodimeric structures to IL-12, with p29/p40 and p28/EB13 components, respectively. In the context of antigen presentation, IL-12, IL-23, and IL-27 can bias T cell responses toward the Th1 phenotype. IL-23 can also play a pivotal role in the production of Th17 cells, along with IL-1 and TGF-β. The IL-12 family of cytokines are generally produced by macrophages in the rheumatoid joint but are also produced by other antigen-presenting cells such as dendritic cells. Although clinical data defining the role of these cytokines are still lacking, there are anecdotal case reports of RA patients with malignancy treated with recombinant IL-12 developing a flare of disease. Animal models including adjuvant arthritis in rats and collagen-induced arthritis in mice are partially ameliorated by neutralization of IL-12, IL-23, or IL-27 depending on the timing of treatment. In some situations, however, IL-27 has anti-inflammatory effects.

**Interleukin-15**

IL-15 is an IL-2–like cytokine that regulates numerous immunologic functions relevant to RA including T cell chemotaxis and proliferation, production of immunoglobulins by B cells, and the generation of NK cells. Although IL-15 can serve as an IL-2–independent mechanism for activating T cells, its role in RA may be related to its key role in TNF regulation. Macrophages are the primary source of IL-15 in RA, and the cytokine is able to induce a cell-contact mechanism of macrophage TNF production that requires T cells. Although T lymphocytes, or at least their membranes, are required for this process, the macrophages actually produce the TNF. This network provides a potential mechanism whereby local IL-15 production in the synovium can lead to autocrine production of TNF in a T cell–dependent but antigen-independent fashion. IL-15 has been demonstrated in RA synovial macrophage cells. Soluble IL-15 receptors can function as an IL-15 inhibitor, and when used in vivo can decrease joint inflammation in collagen-induced arthritis. Preliminary clinical trials in RA using an anti-IL-15 antibody had a signal for efficacy, although probably not sufficient to continue developing this particular biologic agent.

**Interleukin-32**

IL-32 is a novel cytokine that activates NFκB and induces the production of several proinflammatory cytokines and chemokines including TNF, IL-1, IL-6, and IL-8. It has been implicated in Crohn’s disease by virtue of its ability to markedly enhance caspase 1 activation and IL-1 production in cells that have been exposed to muramyl dipeptides. More recently, IL-32 was demonstrated by immunohistochemistry in synovial tissues of patients with RA, especially in synovial-lining macrophage-like cells. The level of IL-32 expression correlated with the presence of other cytokines implicated in RA including TNF, IL-1, and IL-18. Injection of IL-32 into the joints of naïve mice causes a robust transient synovitis. The synovial response could be partially abrogated by anti-TNF antibodies, suggesting that IL-32 induces this cytokine in vivo. These data suggest that IL-32 might be upstream from several proinflammatory mediators in RA and could represent a therapeutic target.

**Colony-Stimulating Factors**

Granulocyte-macrophage colony-stimulating factor (GM-CSF) supports the differentiation of bone marrow precursor cells to mature granulocytes and macrophages. As with other major colony-stimulating factors, GM-CSF also participates in normal immune responses. It is a potent macrophage activator including the induction of HLA-DR expression, IL-1 secretion, intracellular parasite killing, and priming for enhanced release of TNF and PGE2. Neutrophil function is also regulated by GM-CSF, which enhances antibody-dependent cytotoxicity, phagocytosis, chemotaxis, and the production of oxygen radicals.

RA synovial fluid contains GM-CSF, which is produced by RA synovial tissue cells. The major source in the synovium is macrophages, although IL-1– or TNF-stimulated FLS also express the GM-CSF gene. Its ability to induce HLA-DR gene expression on macrophages might be of particular importance in RA. GM-CSF, or IFN-γ, is the major DR-inducing cytokine in RA synovial effusions. Collagen-induced arthritis in mice is less severe in animals that lack a functional GM-CSF gene or are treated with anti–GM-CSF antibody, which supports the hypothesis that GM-CSF is an important proinflammatory mediator. Anti-GM-CSF antibodies are in clinical development. However, there is some concern that pulmonary alveolar proteinosis might be an adverse event because GM-CSF–deficient mice and autoimmune diseases marked by spontaneous production of anti-GM-CSF antibodies develop this syndrome. A recent phase IIa clinical trial demonstrated potential clinical benefit from an anti-GM-CSF antibody in RA, although long term efficacy and safety still must be evaluated.

Macrophase colony-stimulating factor (M-CSF) is also expressed by RA synovium and is present in synovial
effusions. Its primary pathogenic role in RA probably relates to its osteoclast-differentiating capacity. This factor cooperates with RANKL to facilitate bone erosions.

Chemokine Families

Chemokines are a family of related chemoattractant peptides that, with the assistance of adhesion molecules, summon cells into inflammatory sites. They are generally divided into families including CC, CXC, and CX3C based on the position of characteristic cysteine residues.

Many chemokines have been identified in the rheumatoid joint. IL-8, a CXC chemokine that was originally characterized as a potent chemoattractant for neutrophils, along with immune complexes and other chemotactic peptides such as C5α, contributes to the large influx of PMNs into the joint. Immunohistochemical analysis of synovial tissue demonstrates IL-8 protein in sublining perivascular macrophages, as well as in scattered lining cells. Cultured synovial tissue macrophages constitutively produce IL-8, and FLS express the gene if they are stimulated with IL-1 or TNF. IL-8 accounts for about 40% of the neutrophil chemotactant activity in synovial fluid. In addition, IL-8 activates neutrophils through G protein–coupled receptors and is a potent angiogenesis factor.

Many other chemoattractant proteins are implicated in RA. Macrophage-inhibitory protein-1α, macrophage–inhibitory protein-1β, macrophage chemotactant protein-1 (MCP-1), and regulated on activation, normally T cell expressed and secreted (RANTES)—members of the CC subfamily—are produced by RA synovium, as are many other CC chemokines.

CXCL16 and epithelial neutrophil-activating peptide-78 (ENA-78) are CXC chemokines and are also abundant along with many others in that family. The former, which binds to CXCR6 on T cells, can contribute to the recruitment of lymphocytes into the synovium. ENA-78 accounts for about 40% of the chemotactic activity for neutrophils in RA synovial fluid. Concentrations of these chemokines are higher in RA synovial effusions compared with noninflammatory arthritides such as OA. Although the chemokines can also be detected in the blood, the levels are considerably lower than in the joint, thereby providing a gradient that signals cells to migrate into the synovium.

As noted earlier, lymphocyte-specific factors might contribute to the germinal center architecture of RA. The CXC factor B–cell–activating chemokine-1 (BCA-1; CXCL13) binds to specific CXCR5 receptors on B cells. CXCL13 is expressed in the RA synovial tissues, especially by follicular dendritic cells in germinal centers and likely accounts for B cell migration to these regions. CCL21 and several other factors participate in the anatomic organization of germinal centers, marginal zones, and other regions of lymphoid follicles. Another chemokine, SDF-1, is expressed by synoviocytes and endothelial cells and can play a major role as a chemoattractant for T cells in synovium via its receptor CXCR4. Unlike other chemokine receptors that can bind multiple members of the family, CXCR4 is highly specific for SDF-1 and is expressed by memory CD4+ lymphocytes.

Chemokines have attracted considerable attention as therapeutic targets in order to prevent recruitment of immune cells into the synovium. Numerous preclinical models support the use of chemokine blockers. For instance, antibodies to fractalkine (CX3CL1) suppress murine collagen-induced arthritis, even though anti–type II collagen antibody production was not affected. Anti-CXCL16 antibody also decreases clinical arthritis in the same model. One problem is that the system is highly redundant and several different chemotactic proteins can bind to the same receptor. No clinical improvement was observed in a study using anti-MCP-1 antibody, perhaps because the antibody also altered the kinetics of MCP-1 metabolism. Anti-IL-8 antibodies have met with limited success in psoriasis. More promising results were observed with antibody to IP-10 (CXCL10), which met its primary efficacy end point in a phase II study in RA.

Alternatively, chemokines bind to G protein–coupled receptors and can be targeted to block multiple factors. For instance, CCR1 antagonists, which blocks RANTES and MIP-1α, has been evaluated in clinical trials. Although one compound significantly decreased synovial infiltration by CCR1-expressing cells, no significant improvement was observed. Chemokine receptor blockade has other levels of complexity such as CCR2-deficient mice, which develop more severe arthritis in some models. Several CCR2 antagonists, which block MCP-1, have not demonstrated significant efficacy in RA. These data suggest that the chemokine system including the receptors is both redundant, complex, and, in some cases, perhaps protective in arthritis.

Chemokines and other chemotactic factors (such as C5α and LTβ3) can signal through a variety of mechanisms, although many pathways converge on phosphoinositide-3-kinase (PI3K). Of the several PI3K isoforms, PI3Kγ is relatively specific for chemokine signal transduction. This provides an opportunity to block multiple chemokine receptors simultaneously. Proof of concept for this approach was provided by studies in P13Kγ knockout mice, which have less synovial inflammation in passive and active collagen-induced arthritis than wild-type mice. A small molecule P13Kγ inhibitor provided similar benefit. Therefore targeting shared intracellular pathways could potentially overcome some of the limitations presented by the complex chemoattractant system.

Platelet-Derived Growth Factor and Fibroblast Growth Factor

PDGF is a potent growth factor that is both chemoattractant and mitogenic for fibroblasts and induces collagenase expression. It is the most potent stimulator of long-term growth of synovial cells in culture. PDGF is expressed in vascular endothelial cells, platelets, and other synovial sublining cells in rheumatoid synovium, compared with healthy tissue. Multiple isoforms of this molecule have been identified (PDGF A through D), all of which have been detected in RA synovial membranes. PDGF D has been identified as an especially potent stimulator of MMP-1 expression in cultured synoviocytes. PDGF is a potent activator of the P3K pathway, which leads to Akt phosphorylation and supports synoviocyte survival.

Fibroblast growth factors (FGFs) are a family of peptide growth factors with pleiotropic activities. In rheumatoid patients, it is likely that heparin-binding growth factor, the
precursor of acidic fibroblast growth factor, is a major mitogen for many cell types and stimulates angiogenesis. Interactions between FGF and proteoglycans is required for biologic activity. It is a potent angiogenic factor, inducing capillary endothelial cells to invade a three-dimensional collagen matrix and form capillaries. Synoviocytes can also be induced to increase expression of RANKL by FGF, thereby enhancing osteoclast activation and bone resorption. FGF is present in RA synovial fluid, and the genes are expressed by synovial cells. Synovial fibroblasts express FGF receptors and proliferate after exposure to the growth factor.

Suppressive Cytokines and Cytokine Antagonists

The proinflammatory cytokine network in RA is offset by a variety of suppressive and anti-inflammatory factors that attempt to reestablish homeostasis. Low production of these suppressive cytokines could potentially contribute to the perpetuation of the synovitis. Many cytokine antagonists or natural immunosuppressives represent potential therapeutic targets for the treatment of inflammatory diseases.

Interleukin-1 Receptor Antagonist

IL-1Ra is a naturally occurring IL-1 inhibitor that binds directly to types I and II IL-1 receptors and competes with IL-1 for the ligand-binding site. Interaction of IL-1Ra with the IL-1 receptors does not result in signal transduction, and, in contrast to IL-1α or -β, the receptor-ligand complex is not internalized after it binds to the IL-1 receptor. Even though IL-1Ra has high affinity for the IL-1 receptor, it is a relatively weak inhibitor because IL-1 can activate cells even if only a small percentage of IL-1 receptors is occupied. Because of this, a substantial excess of the inhibitor is required to saturate the receptor and thereby block IL-1–mediated stimulation (usually 10- to 100-fold excess of IL-1Ra). Recombinant IL-1Ra inhibits a variety of IL-1–mediated events in cultured cells derived from the joint including the induction of MMP and prostaglandin production by chondrocytes and synoviocytes. It can block synovitis in rabbits induced by direct intra-articular injection of recombinant IL-1.

IL-1Ra is present in rheumatoid synovial effusions; much of it is produced by neutrophils and macrophages. Immunohistochemical studies of rheumatoid synovium reveal IL-1Ra protein, especially in perivascular mononuclear cells and the synovial intimal lining. The IL-1Ra protein and mRNA can be detected in synovial macrophages and, to a lesser extent, in type B synoviocytes (Figure 69-12). The presence of IL-1Ra in synovium is not specific to RA because OA synovial tissue also contains IL-1Ra, albeit in lesser amounts. Normal synovium contains little, if any, IL-1Ra protein. Synovial cell culture experiments show that the amount of IL-1Ra is insufficient to antagonize synovial IL-1.

Interleukin-10

IL-10 is an immunosuppressive cytokine that was originally characterized as an inhibitor of T cell cytokine production. Its immunosuppressive actions might be important in pregnancy to suppress an immune response directed against paternal MHC antigens, and it might regulate susceptibility to some parasitic infections. As noted previously, IL-10 protein is present in RA synovial fluid, and the gene is expressed by synovial tissue macrophages. Serial synovial biopsies in RA patients who were treated with recombinant IL-10 did not show any significant histologic improvement, and clinical responses were not impressive in a limited study.

Transforming Growth Factor-β

TGF-β is a key member of the TGF superfamily, which includes the bone morphogenic proteins that signal through intracellular signaling molecules known as Smads. It is widely distributed in different tissues and produced by many cells including T cells, monocytes, and platelets. It suppresses the production of collagenase and induces the expression of TIMP. TGF-β accelerates the healing of incisional wounds and induces both fibrosis and angiogenesis in experimental animal models. Substantial amounts of TGF-β are present in synovial fluid, although it is mainly present in an inactive, latent form, and the mRNA can be detected in RA synovial tissue. Although typically considered an immunosuppressive cytokine with wound-healing properties, the role of TGF-β in RA is complex.

In RA, TGF-β is one of the factors responsible for blunted responses of T cells that have been exposed to synovial fluid. TGF-β also downregulates IL-1 receptor expression on some cell types including chondrocytes. When it is injected directly into the knees of animals,
fibrosis and synovial lining hyperplasia develop. In streptococcal cell wall arthritis, parenteral administration or systemic gene therapy with the TGF-β gene ameliorates the disease. However, intra-articular administration of anti-TGF-β antibody decreases arthritis in the injected joint but not in the contralateral joint in the same model. Although mainly considered anti-inflammatory, TGF-β also plays a key role in the development of autoimmunity through the differentiation of T cells into the Th17 phenotype.

**Soluble Cytokine Receptors and Binding Proteins**

Soluble cytokine receptors and binding proteins can absorb free cytokines and prevent them from engaging functional receptors on cells. Although these obviously could inhibit cytokine action, they also could act as carrier proteins that protect cytokines from proteolytic degradation or deliver them directly to cells such as the IL-6 receptor.

TNF receptors are normally expressed as membrane-bound proteins and can be released from the cell surface after proteolytic cleavage. Soluble p55 and p75 TNF receptors have been detected in RA synovial fluid, sometimes in high concentrations. Soluble TNF receptor levels can be considerably higher than the concentration of TNF in blood or synovial fluid and probably explain why biologically active TNF is difficult to detect in RA synovial fluid despite the presence of immunoreactive protein. Synovial membrane mononuclear cells have increased surface expression and mRNA levels of both TNF receptors compared with OA synovial tissue cells or peripheral blood cells.

Many other soluble receptors and binding proteins are produced in RA, albeit in concentrations too low to effectively suppress the proinflammatory cytokine milieu of the joint. For instance, the IL-1 type II receptor is present in RA synovial fluid, along with lesser amounts of the type I receptor. These soluble receptors can bind to IL-1 or IL-1Ra in synovial effusions. Soluble receptors to IL-15 and IL-17 have been characterized, and an IL-18 binding protein also can inhibit cytokine activity. In some cases, a soluble receptor can protect a cytokine from degradation or transport it to the cells, as with the IL-6 receptor.

**Perpetuation of Synovitis by Macrophage-Fibroblast Cytokine Networks**

Mapping the cytokine profile in RA led to targeted biologic therapy to inhibit cytokines and supports the concept that cytokine networks play a key role. The network pathway is not autonomous and, except for certain patients with very early disease, discontinuation of anti-cytokine therapy allows the disease to flare. Nevertheless, paracrine and autocrine cytokine networks in the synovial intimal lining contribute largely to inflammatory arthritis in RA (Figure 69-13).

Several cytokines that have been identified in the synovium or synovial fluid can participate in this system and might explain lining cell hyperplasia, HLA-DR and adhesion-molecule induction, and synovial angiogenesis. The list of potential candidates in this highly redundant system is extensive. Several of these including IL-1, TNF, and IL-6 now have well-defined roles. The first two, along with numerous other factors (e.g., IL-15, IL-18, IL-32) are produced by synovial macrophages and stimulate synovial fibroblast proliferation and secretion of IL-6, GM-CSF, and chemokines, as well as effector molecules such as MMPs and prostaglandins. GM-CSF, which is produced by both synovial macrophages and IL-1β− TNF−stimulated synovial fibroblasts, can, in turn, induce IL-1 secretion to form a positive feedback loop. GM-CSF, especially in combination with TNF, also increases HLA-DR expression on macrophages. Macrophage and fibroblast cytokines could also indirectly contribute to the evidence for local T cell and B cell activation including RF production. Newly implicated members of the macrophage-fibroblast cytokine network such as IL-33 can activate other cell lineages such as mast cells that have attracted increased attention in recent years.

The cytokine profile recruits other immune and inflammatory cells into the synovium through the production of chemokines that select specific cell lineages for admission and retention in the synovium. Many of these chemokines including the CXC and CC families are produced by macrophages and fibroblasts and attract neutrophils, macrophages, and specific subpopulations of T and B cells. The sublining chemokine profile including CCL13 and CCL20 helps organize these newly infiltrated cells into organized lymphoid structures in some patients. Other factors such as IL-12 and IL-23 differentiate CD4+ T cells into the Th1 phenotype to produce relatively small amounts of IFN-γ and other relevant cytokines. IL-1 and TGF-β produced mainly by the lining can, along with IL-23, support the production of Th17 cells that release the potent IL-17 family members into the local milieu. All of this occurs in the presence of inhibitor factors, soluble receptors, and binding proteins that are overwhelmed by the inflammatory drive. Other cytokines, such as RANKL and M-CSF, activate osteoclasts that remodel bone.

Even though they do not cause RA, per se, cytokines clearly orchestrate the rheumatoid process. For individual patients, the pivotal cytokine or cytokines that must be blocked could be different, and even this could vary with the stage of disease. Ultimately, understanding the genetic
predisposition, the environmental triggers, and the specific patterns of cytokine production could help determine the correct combination of cytokine inhibitors that will be effective.

**SIGNAL TRANSDUCTION AND TRANSCRIPTION FACTORS**

**KEY POINTS**

| Complex intracellular signaling mechanisms regulate cytokine production and actions in RA synovium. |
| NFκB, MAP kinases, AP-1, JAK, Syk, and several other pathways are potential therapeutic targets in RA. |

Intracellular signal transduction systems transmit information from the environment to the interior of a cell, which can then respond appropriately to stress. The remarkable diversity of signaling pathways and transcription factors provides a selective mechanism for orchestrating activation and repression for appropriate genes. Many of the inflammatory responses observed in RA synovium including the activation of cytokine, proteases, and adhesion-molecule genes can be traced to specific transcription factors and signal transduction pathways.

**Nuclear Factor κB**

NFκB is a ubiquitous transcription factor that plays a key role in the expression of many genes central to RA including IL-1, TNF, IL-6, and IL-8. NFκB normally resides as an inactive heterodimer or homodimer in the cell cytoplasm associated with an inhibitory protein called IκB that regulates the DNA binding and subcellular localization of NFκB proteins by masking a nuclear localization signal. Extracellular stimuli such as cytokines or TLR agonists initiate a signaling cascade leading to activation of two IκB kinases (IKKα and IKKβ), which phosphorylate IκB at specific NH2-terminal serine residues. Phosphorylated IκB is then selectively ubiquitinated and degraded by the 26S proteasome. This process permits NFκB to migrate to the cell nucleus, where it binds its target genes to initiate transcription.

NFκB is abundant in rheumatoid synovium, and immunohistochemical analysis demonstrates p50 and p65 NFκB proteins in the nuclei of cells in the synovial intimal lining. Although the proteins can also be detected in OA synovium, NFκB activation is much greater in RA because of phosphorylation and degradation of IκB in RA intimal lining cells. Nuclear translocation of NFκB in cultured FLS occurs rapidly after stimulation with many proinflammatory cytokines (e.g., IL-1, TNF, IL-17) or TLR ligands (e.g., peptidoglycan, LPS). NFκB is also a major survival factor for cells and is responsible for resistance to apoptosis in some cell lineages after activation. This is especially important in intestinal inflammation, where IKKβ deficiency can actually increase mucosal damage because the lack of NFκB can increase cell death.

The relevance of NFκB to inflammatory arthritis has been tested in several animal models. Synovial NFκB is rapidly activated, often long before clinical arthritis is evident. Adjuvant arthritis in rats is ameliorated using intra-articular gene therapy with the dominant negative IκBα construct that blocks the IKK pathway, and streptococcal cell-wall arthritis is blocked with decoy oligonucleotides or a dominant negative IκB adenovirus. NFκB inhibition is associated with decreased synovial cellular infiltration, as well as increased apoptosis. The role of this transcription factor in murine collagen-induced arthritis has been demonstrated using selective IκB inhibitors, which suppress arthritis and joint destruction.

**Activator Protein-1**

Like NFκB, AP-1 regulates many genes implicated in RA including TNF and the MMPs. AP-1 activity can be induced by extracellular signals including cytokines, growth factors, tumor promoters, and the Ras oncoprotein. AP-1 includes members of the Jun and Fos families of transcription factors that form Jun homodimers, Jun-Jun heterodimers, or Jun-Fos heterodimers. Multiple Jun and Fos family members (c-Jun, JunB, JunD, c-Fos, FosB, Fra-1, Fra-2) are expressed in different cell types that mediate the transcription of both unique and overlapping genes. AP-1–driven gene expression is greatly enhanced when one of its components, especially c-Jun, is phosphorylated by the c-Jun N-terminal kinase (JNK).

AP-1 proteins and nuclear binding are elevated in RA synovium, especially in the nuclei of cells in the intimal lining layer. c-Jun and c-Fos proteins are also expressed in the sublining inflammatory infiltrate, albeit to a lesser degree. Localization of AP-1 to the intimal lining correlates with the site where most protease and cytokine genes are overexpressed in RA. AP-1 proteins are usually not detected in normal synovium, although modest amounts have also been detected in OA.

Cytokines such as IL-1 and TNF and TLR ligands probably contribute to the activation of AP-1 in RA synovium. These factors are potent inducers of AP-1 nuclear binding in cultured FLS. The specific Jun family members that constitute AP-1 in synoviocytes have a clear effect on function. For instance, c-Jun increases the production of proinflammatory mediators, whereas JunD suppresses cytokine and MMP production. AP-1 decoy oligonucleotides suppress collagen-induced arthritis and inhibit cytokine production by synovial tissue.

**Mitogen-Activated Protein Kinases**

MAP kinases, which are signal transduction enzymes activated in response to cellular stress, are composed of parallel protein-kinase cascades that regulate cytokine and MMP gene expression. There are three different families of MAP kinases known as JNK, p38, and extracellular signal-regulated kinase (ERK). MAP kinases phosphorylate selected intracellular proteins, including transcription factors, which subsequently regulate the expression of various genes by transcriptional and post-transcriptional mechanisms. MAP kinases are activated by phosphorylation at conserved threonine and tyrosine residues by a cascade of dual-specificity kinases. These are, in turn, activated by MAP kinase kinase kinases. The relative hierarchy
of the individual MAP kinases depends on the cell type and inflammatory stimulus.

The MAP kinases are widely expressed in synovial tissue and are activated in rheumatoid synovium. Phosphorylated ERK, p38, and JNK can be detected by immunohistochemistry or Western blot analysis. All three kinases and their upstream regulators are constitutively expressed by cultured FLS and can be activated within minutes after exposure to cytokines. They regulate production of proinflammatory cytokines and MMPs.

p38 inhibitors are effective anti-inflammatory agents in murine collagen-induced arthritis and rat adjuvant arthritis, possibly by decreasing the production of proinflammatory cytokines. In addition, p38 inhibitors block TNF and IL-6 production by cultured macrophages and synoviocytes, as well as cultured synovial tissue cells. Spinal p38 also plays a major role in pain processing, and inhibitors have potential for analgesic, as well as anti-inflammatory, action. Recent studies suggest that p38 in the central nervous system can regulate peripheral inflammation because intrathecal administration of a p38 inhibitor suppresses inflammation and joint destruction in rat adjuvant arthritis.

Given the putative role of p38 in RA, it is surprising that clinical efficacy of the selective p38 inhibitors is modest, at best. Clinical response rates are rarely greater than 40%, compared with the 60% range for TNF blockers. Perhaps more interesting, an initial decrease in acute-phase reactants is only transient despite adequate drug levels. The explanations for the dissociation among the animal, in vitro, and human data are uncertain, and it is clear that the patterns of kinase and cytokine activation differ. However, p38 regulates several anti-inflammatory pathways in addition to traditional proinflammatory mediators. It is possible that the beneficial effects are mitigated by the loss of these negative feedback loops.

The other two MAP kinases are also potential therapeutic targets. The ERK pathway, which regulates cell growth and some inflammatory mediators, can be blocked by inhibiting its upstream kinases MEK1 and MEK2. One clinical trial with a small molecule inhibitor saw minimal benefit, although the placebo response rate was high. Two of the three JNK isoforms (JNK1 and JNK2) regulate a variety of genes in inflammation including TNF and metalloproteinases. Using a JNK inhibitor, marked protection of bone destruction was observed in the adjuvant arthritis model, along with decreased synovial AP-1 activation and collagenase gene expression. No benefit was observed in the JNK2–/– mice, and only modest cartilage protection was seen in the JNK2–/– animals.

As an alternative to blocking MAP kinases themselves, upstream kinases might be targeted. MKK3 and MKK6, which regulate p38, are activated in the rheumatoid synovial intimal lining. MKK3 knockout and MKK6 knockout mice have markedly decreased joint inflammation in the passive K/BxN model. MKK4 and MKK7 are the main kinases that modulate JNK function and are also activated in the rheumatoid synovium. Only MKK7 is required for cytokine-stimulated JNK activation and MMP expression in cultured synoviocytes. Going even further upstream is also possible, and targeting TAK1, which plays a role in NFκB and JNK activation, could have broader effects.
In contrast to DCs, where IRF7 is the primary IRF that regulates IFN responses, IRF3 is the pivotal factor that is responsible for the IFN signature. Overexpression of some IRF3-driven genes, most notably IFN-β, could have a beneficial impact in inflammatory arthritis. Mice with collagen-induced arthritis injected with IFN-β or transduced fibroblasts expressing IFN-β have less severe disease compared with controls including decreased bone and cartilage destruction. A clinical trial of IFN-β in RA showed no benefit. An alternative approach might be to inhibit IKK-related kinases to block the chemokines associated with the IFN response and concomitantly treat with low levels of exogenous IFN-β.

**Spleen Tyrosine Kinase and Other Signaling Pathways**

Several other signaling molecules have been implicated in RA and are potential therapeutic targets. For example, spleen tyrosine kinase (Syk) is involved in immunoreceptor signaling in a variety of cell types. For example, Syk plays a critical role in Fc receptor signaling in macrophages and mast cells. It also participates in B cell activation after ligation of the B cell receptor. A small molecule Syk inhibitor has demonstrated efficacy in patients with RA, although significant benefit was not observed in individuals that did not respond to TNF blockers. c-Kit blockade has also been evaluated in an open-label study that provided a positive signal. PI3 kinases, especially the gamma and delta isoforms, are also attractive therapeutic targets due to their role in innate immunity and cell recruitment. There is no shortage of targets, and many others including upstream MAP kinase regulators, Ras, IL-1 associated kinases (IRAKs), sphingosine kinase-1 (SK-1), and Bruton’s tyrosine kinase (BTK) will be evaluated in the future.

**CELL SURVIVAL AND DEATH IN RHEUMATOID SYNOVIIUM**

**KEY POINTS**

- Reactive oxygen and nitrogen in RA joints contribute to a toxic environment that can damage cells and increase inflammation.
- Deficient apoptosis, or cell death, can contribute to the accumulation of cells in rheumatoid synovium.
- Abnormalities of key regulatory genes such as the p53 tumor suppressor can enhance accumulation of cells in the joint.
- Inducing apoptosis can potentially suppress synovial inflammation and joint destruction.

Studies defining the life cycle of cells have opened a new door to understanding the pathogenesis of neoplastic and inflammatory diseases. Although most investigators previously focused on cell proliferation as a mechanism of synovial hyperplasia, increasing attention has been paid to the other side of the equation (i.e., whether insufficient cell death could also contribute to this process). In this section, the role of oxidative damage, programmed cell death, and permanent changes in the genome are discussed because they can alter the natural history of RA.
Reactive Oxygen and Nitrogen

Oxidative stress in the joints of RA patients results from increased pressure in the synovial cavity, reduced capillary density, vascular changes, an increased metabolic rate of synovial tissue, and locally activated leukocytes. The generation of reactive oxygen species can also be facilitated by repetitive ischemia-reperfusion injury in the joint. Tissue injury releases iron and copper ions and heme proteins that are catalytic for free-radical reactions. Electron transport chains are also disrupted in the mitochondria and endoplasmic reticulum, leading to leakage of electrons to form superoxide.

Evidence for increased production of reactive oxygen species in RA patients includes elevated levels of lipid peroxidation products, degradation of hyaluronic acid by free radicals, decreased levels of ascorbic acid in serum and synovial fluid, and increased breath pentane excretion. Moreover, the levels of thioredoxin, a marker of oxidative stress, are significantly higher in synovial fluid from RA patients compared with other forms of arthritis. Peripheral blood lymphocyte DNA from RA patients contains significantly increased levels of the mutagenic 8-oxohydrodeoxyguanosine, which is a product of oxidative damage to DNA, pointing to the genotoxic effects of oxidative stress.

Nitric oxide (NO) production is also high in rheumatoid synovial tissue. Low levels of NO are constitutively produced by endothelial or neuronal syntheses, and this is substantially increased by inducible NO synthase after stimulation by cytokines or bacterial products. The nitrite levels in synovial fluid are elevated in RA patients, indicating local NO production. In addition, the urinary nitrate-to-creatinine ratio is increased and inducible NO synthase is present in the synovium.

Apoptosis

Programmed cell death, or apoptosis, removes cells safely from living tissue and permits remodeling, or cell deletion without causing an inflammatory response. Apoptosis is a normal process that is tightly regulated and can be initiated by withdrawal of hormones and growth factors. It is evident in the elimination of autoreactive cells such as thymocytes in the thymus gland and the loss of cells after DNA damage or toxic exposure. It also plays a critical role in immune response by deleting activated T cells and terminating an inflammatory response by rapidly removing neutrophils.

Genes Regulating Apoptosis

The accumulation of cells in RA results from a balance of cell recruitment, cell egress, local proliferation, and local death. Any imbalance can potentially lead to synovial hyperplasia. T cell apoptosis in RA synovial effusions, for instance, is significantly less than lymphocytes from crystal-induced arthropathy. High expression of the antiapoptotic molecule Bcl-2 is found in lymphoid aggregates and protects synovial T cells from programmed cell death. Resistance to apoptosis in vitro increases if RA T cells are cocultured with FLS. The specific adhesion molecules involved are not defined, although the integrin-binding RGD motif (arginine-glycine-asparagine) blocks the protective effects of synoviocytes.

Fas and its TNF superfamily counter-receptor Fas ligand (FasL) are potent regulators of cell death for many cell types including synovial T cells and synoviocytes. Fas is expressed by rheumatoid synovial fluid T cells, and the number of Fas’ cells in the peripheral blood of RA patients is greater than in healthy controls. Anti-Fas antibody, which cross-links Fas on cell surfaces, rapidly causes apoptosis in synovial fluid B and T lymphocytes in RA, although peripheral-blood T cells are more resistant. Another member of the TNF superfamily, TNF-related apoptosis-inducing ligand (TRAIL) binds to two receptors (DR4 or DR5) to induce caspase-dependent apoptosis. DR5 is expressed in RA FLS but not OA cells, and apoptosis can be induced by either TRAIL or agonistic anti-DR5 antibody.

Studies of apoptosis in RA synovial tissue show only a small number of apoptotic nuclei in the intimal lining and sublining. Electron microscopic studies show rare cells that exhibit the typical findings of programmed cell death. Lymphoid aggregates containing high levels of Bcl-2 have few apoptotic cells. Macrophage apoptosis is also low because they express high levels of the caspase 8 inhibitor FLICE-like inhibitory protein (FLIP), which can inhibit Fas-mediated apoptosis.

The mechanisms for inducing apoptosis in FLS can involve several pathways including induction of JNK and AP-1 activation, inhibition of the kinase Akt, or suppression of NFκB. However, it is comparatively difficult to induce apoptosis in cultured synoviocytes. p53, which typically induces cell-cycle arrest and either DNA repair or apoptosis, is also expressed in the synovial lining and sublining. However, one of the main effectors of p53-mediated apoptosis, p53 upregulated modulator of apoptosis (PUMA), is only present in low concentrations in the synovium and cultured synoviocytes. Synoviocytes are resistant to p53-mediated apoptosis even when cells are forced to over-express the protein using genetic methods. Fas is constitutively expressed by cultured synoviocytes, and programmed cell death is initiated in a minority of cells when it is cross-linked by anti-Fas antibody. Synoviocyte apoptosis can be initiated by oxidative stress such as hydrogen peroxide or by exposure to nitric oxide.

The relative paucity of apoptosis in RA can also be partially explained by patterns of gene expression that favor cell survival. Senrin-1, a ubiquitin-like protein, regulates the cell survival by modifying proteins involved in apoptosis. Senrin-1 is expressed in RA synovium, especially at sites of cartilage invasion, and protects cells from Fas-mediated death. A second protein, phosphatase and tensine homolog on chromosome 10 (PTEN) was originally defined as a key factor that protects from tumorigenesis through antagonism of PI3K, Akt, and many other proliferative pathways. Underexpression of PTEN in RA has been described in rheumatoid synovial intimal lining, as well as cultured FLS.

Therapeutic Interventions That Increase Apoptosis

Fas-induced death has some clinical relevance and has been used successfully in murine collagen-induced arthritis using anti-Fas antibodies and adenovirus encoding for Fas ligand. Anti-Fas antibody also induces synovial cell death in RA synovial tissue explanted in SCID mice. In the SCID
mouse model using RA synovial explants, anti-DR5 antibody decreased cartilage erosion. Similarly, adenoviral transfer of TRAIL in a rabbit model of arthritis decreases synovial inflammation. The importance of apoptosis as a regulator of inflammation was confirmed in murine collagen-induced arthritis, where genetic DR5 deficiency exacerbated the disease. Blocking FOXO30 in PMNs is also an effective method of deleting these inflammatory cells and suppressing inflammatory arthritis in mice.

The Bcl2 homology 3 (BH3) domain-only proteins are potent inducers of apoptosis. The challenge is how to get these proteins expressed in the target cell. One of these, Bim, was engineered into a cell membrane permeable protein (TAT-Bim) and evaluated in the passive K/BxN model. The construct decreased arthritis severity in prophylactic and therapeutic treatment protocols and was associated with apoptosis of cells, mainly in the myeloid lineage. Therefore targeted cell death of individual lineages can potentially decrease inflammatory arthritis.

Other molecules that regulate apoptosis have also demonstrated potential utility in animal models. For instance, NFκB blockade in streptococcal cell-wall arthritis induces synovial apoptosis and suppresses arthritis. p53 gene therapy in rabbit antigen-induced arthritis induces synovial apoptosis and decreases inflammation. The pleiotropic activities of p53 were demonstrated in collagen-induced arthritis because p53−/− mice with the disease developed increased inflammation and greater joint destruction in association with decreased apoptosis. Joint damage was mediated by increased expression of collag enase genes in the knockout mice, most likely because p53 directly suppresses MMP gene transcription. However, p53 knockout mice with passive models of arthritis have normal disease severity. This suggests that the protective effects of p53 are partly due to effects on adaptive immunity.

**Tumor Suppressor Genes**

The p53 tumor suppressor is a key regulator of DNA repair and cell replication. p53 protein expression is significantly greater in the rheumatoid synovium compared with OA and normal tissue. Of interest, p53 protein can also be detected in RA synovium from patients with very early RA. However, its expression is much lower in other inflammatory arthropathies such as reactive arthritis, which might reflect greater DNA damage and oxidative stress in RA.

Somatic mutations in the p53 gene occur in RA synovium could contribute to the unusual phenotype of RA synoviocytes and inadequate apoptosis in synovial tissue. Transition mutations, which are characteristic of damage induced by reactive oxygen or nitric oxide, account for more than 80% of the base changes. Some of the mutant p533 genes exhibit dominant negative characteristics and suppress the function of the wild-type allele. Microdissection studies identified mutant islands with oligoclonal expansion, and the loss of p53 function in a region of RA synovium was associated with increased IL-6 gene expression in the same location. The data suggest that mutations do not cause RA but, instead, are the result of long-standing oxidative stress. The gene alterations can then potentially increase the aggressive nature of the synovium and alter the natural history of RA.

Abnormalities in other genes have also been reported in RA. For instance, synovial T cells in RA have an increased incidence of mutations in the HPRT1 gene. Although not functionally important, these synovial T cells act as a marker for oxidative damage that occurs in the synovial milieu. Some of these abnormal lymphocytes can also be detected in the peripheral blood, suggesting that articular T cells can migrate out of the joint.

Microsatellite instability, which is marked by mutations in mononucleotide and dinucleotide repeat sequences in noncoding DNA, is also significantly greater in RA than OA synovial tissue. Occasional mutations in a coding region microsatellite in the WISP-3 gene, which can regulate type II collagen and aggrecan expression, have been identified in RA synovium. However, similar mutations were observed in OA, suggesting that these are not specific. Mutations in mitochondrial genes have also been described in RA, most likely due to oxidative damage.

Evaluation of DNA mismatch repair (MMR) genes in rheumatoid synovium suggests that the balance of two genes that protect against mutations might contribute to the pattern of DNA damage in RA, with relatively high levels of MSH3 and low levels of MSH6 after reactive nitrogen stress. The former repairs large insertions and deletions, whereas the latter repairs single-base abnormalities. Because most mutations detected in RA involve single bases, the changes in MMR enzyme levels favor these limited mutations rather than more substantial ones.

**BLOOD VESSELS IN RHEUMATOID ARTHRITIS**

**KEY POINTS**

Angiogenesis is a dynamic process in RA that provides nutrients to expanding synovium.

Angiogenic factors such as IL-8, FGF, and VEGF can enhance blood vessel proliferation in the synovium.

Microvascular endothelium in the synovium expresses adhesion molecules that guide circulating cells into the joint under the influence of chemoattractants.

Blood vessels play an active role in such inflammatory processes, not only as a means of selecting which cells should enter the tissue but also as a determinant of tissue growth and nutrition through the proliferation of new capillaries. Understanding the structure and function of the microvasculature provides insights into how a highly catabolic tissue such as the rheumatoid synovium can flourish.

**Angiogenesis in Rheumatoid Arthritis: Feeding the Starved Synovium**

The importance of luxurious new capillary growth early in the development of synovitis has been recognized for many years. The absolute number of blood vessels is increased in RA synovium (Figure 69-15), with a rich network of sublining capillaries and postcapillary venules in histologic sections stained with endothelium-specific antibodies. These blood vessels, however, are not necessarily normal, with a predominance of straight, branching morphology compared...
angiogenic factors are regulated by the hypoxia-sensing protein HIF-1α. Low oxygen tension also leads to HIF-1α-induced transcription of VEGF, a specific endothelial cell mitogen that is present in high concentrations in rheumatoid synovial fluid and tissue. Elevated serum concentrations in early disease correlate with subsequent radiographic progress. VEGF also stimulates expression of collagenase, which can degrade the extracellular matrix to make room for the advancing vasculature and pannus. VEGF expression is especially high in the synovial intimal lining, and the angiogenesis factor is also produced by cultured FLS that have been exposed to hypoxia and IL-1. HIF-1α has other functions that regulate inflammation; selective deficiency in myeloid lineage cells suppresses inflammation in the passive K/BxN model of arthritis.

**Fig. 69-15** Human rheumatoid synovial membrane stained with antibody to von Willebrand factor to delineate blood vessels. Virtually all of these blood vessels formed in response to angiogenic stimuli after the rheumatoid process had been initiated. (Courtesy Dr. Paul-Peter Tak.)

with tortuous vessels in psoriatic arthritis synovium. The blood vessels in the inflamed synovium were less mature, probably due to increased DNA damage and decreased recruitment of pericytes. Of interest, the immature vasculature is selectively depleted by TNF blockers in RA.

**Hypoxia**

The mass of tissue outstrips angiogenesis in RA as determined by the number of blood vessels per unit area and causes local tissue ischemia. Synovial fluid oxygen tensions are remarkably low, lactate measurements are frequently high, and the pH can be as low as 6.8. The mean rheumatoid synovial fluid PO2 in samples from rheumatoid knees is approximately 30 mm Hg and occasionally less than 15 mm Hg. Another cause of diminished blood flow is increased positive pressure exerted by synovial effusions within the joint, a process that obliterates capillary flow while producing ischemia-reperfusion injury in the joint. Altered vascular flow may not be the only cause of hypoxia in joints; oxygen consumption of the rheumatoid synovium is 20 times normal.

Hypoxia is a potent stimulus for angiogenesis, and many angiogenic factors are regulated by the hypoxia-sensing protein HIF-1α. Low oxygen tension also leads to HIF-1α-induced transcription of VEGF, a specific endothelial cell mitogen that is present in high concentrations in rheumatoid synovial fluid and tissue. Elevated serum concentrations in early disease correlate with subsequent radiographic progress. VEGF also stimulates expression of collagenase, which can degrade the extracellular matrix to make room for the advancing vasculature and pannus. VEGF expression is especially high in the synovial intimal lining, and the angiogenesis factor is also produced by cultured FLS that have been exposed to hypoxia and IL-1. HIF-1α has other functions that regulate inflammation; selective deficiency in myeloid lineage cells suppresses inflammation in the passive K/BxN model of arthritis.

**Angiogenic Factors**

VEGF can bind to two receptors with tyrosine kinase domains: VEGF-R1/Flt-1 and VEGF-R2. VEGF-R1 regulates inflammatory responses in macrophages such as IL-6 and phagocytosis. VEGF-R1+ mice are resistant to arthritis in the HTLV1 pX model, which is marked by unregulated proliferation of synovial cells. Small-molecule VEGF-R inhibitors also suppress acute models of inflammation such as carrageenan paw edema and mouse collagen-induced arthritis. Therefore targeting this receptor with a small molecule might suppress the angiogenic and proinflammatory actions of VEGF.

In addition to the hypoxia-driven stimulus for blood vessel growth, the inflammatory cytokine milieu of the joint also encourages angiogenesis. Several proinflammatory factors expressed by the rheumatoid joint including IL-8, FGF, and TNF are angiogenic. Many of these cytokines further enhance angiogenesis by increasing expression of angiopoietins (Ang-1 and Ang-2) by synoviocytes, which can then bind to their tyrosine kinase receptor, Tie-1, on RA capillary endothelial cells. Additional angiogenesis factors derived from activated adhesion molecules on the surface of endothelial cells such as soluble E-selectin and soluble VCAM are released in RA synovium and contribute to vascular proliferation. Limited quantities of some antiangiogenic mediators that inhibit capillary proliferation such as platelet factor-4 and thrombospondin are also produced by the joint.

Vascular remodeling is an active process that involves the continuous creation and resorption of blood vessels. In RA, new capillaries that form under the influence of proangiogenic factors can be identified by the expression of integrins such as αvβ3. Endothelial proliferation is especially prominent in synovial tissue regions containing VEGF. Synovial blood vessel involution can also be detected as evidenced by apoptosis of the endothelium in other synovial locations. The ratio of proliferating and involuting blood vessels is significantly higher in RA than OA or normal synovium.

**Targeting Angiogenesis**

The importance of new blood-vessel formation in inflammatory arthritis was elegantly demonstrated in the collagen-induced arthritis model. The disease was markedly attenuated in animals pretreated with an angiostatic compound similar to fumagillin, which is derived from Aspergillus. This compound is cytotoxic to proliferating, but not resting, endothelial cells. In addition, there was regression of established arthritis if treatment was initiated well into the course of the disease. Hence angiogenesis is essential for the establishment and progression of inflammatory arthritis because of the need for blood vessels to either recruit leukocytes or provide nutrients and oxygen to starved tissue.

Targeting HIF-1α with a small molecule inhibitor that blocks nuclear translocation and VEGF induction was effective in adjuvant arthritis. Other small molecule inhibitors of the VEGF-R1 with either antibodies or kinase inhibitors have been successfully tested in preclinical models of arthritis as well. One of these, vatalanib, decreased inflammatory knee arthritis in rabbits.
Several other antiangiogenesis approaches are effective in animal models of arthritis. For instance, thrombospondin 1 overexpression significantly decreased blood vessel density, inflammation, and joint destruction in rat collagen-induced arthritis. Direct intra-articular administration of a cyclic RGD peptide was used in a rabbit model to block $\alpha_4\beta_1$ integrin. As with RA synovium, $\alpha_4\beta_1$ is expressed by proliferating blood vessels in inflamed rabbit synovial tissue. The cyclic peptide decreased joint inflammation, increased endothelial cell apoptosis, and suppressed bone and cartilage destruction.

The ability of RGD to bind selectively to proliferating blood vessels was also used to home a proapoptotic agent to synovial neovascularature in murine collagen-induced arthritis. The cyclic RGD peptide was administered systemically and accumulated in inflamed synovium but not normal joints or other organs. Apoptosis was induced in synovial blood vessels and arthritis regressed. The potent angiogenesis inhibitor endostatin has been tested in the SCID mouse model, and it decreased synovial explant inflammatory cell infiltration and capillary density. Despite the compelling rationale for antiangiogenic therapy, an anti-$\alpha_4$ antibody showed minimal efficacy in a clinical trial, perhaps because other pathways are more important in the synovium.

### Adhesion Molecule Regulation

Endothelial cells activated by cytokines and other mediators express adhesion molecules that bind to counter-receptors on mononuclear cells and neutrophils from the circulation and facilitate their recruitment from the blood (see Chapter 25). Several categories of vascular adhesion molecules exist. The selectins (E-, L-, and P-selectin) are a family of adhesion molecules whose primary ligands are carbohydrates, especially sialyl Lewis$^a$, and related oligosaccharides. A second family is integrins, which are heterodimers that include an $\alpha$- and a $\beta$-chain. The counter-receptors depend on the specific combination of these chains and are frequently proteins in the immunoglobulin supergene family or extracellular matrix proteins. Several novel peptides have been described that selectively bind to the blood vessels of human synovial explants in SCID mice and have potential utility as inhibitors of cell adhesion, specifically to joint tissue.

### Integrins and Ligands

As one might expect, adhesion molecule expression is increased in the RA synovium due to the rich cytokine milieu. Immunohistochemical techniques localize high levels of ICAM-1 to sublining macrophages, macrophage-like synovial lining cells, and fibroblasts compared with normal tissue. Significant amounts are also present on the majority of vascular endothelial cells. Cultured FLS also constitutively express ICAM-1, which can be markedly increased by TNF, IL-1, and IFN-$\gamma$. ICAM-1 and the other ICAM family members can bind to cells expressing the $\beta_2$ integrins, especially neutrophils.

Adhesion of $\alpha_4\beta_1$ (VLA-4)–expressing mononuclear cells such as memory T cells or monocytes to cytokine-activated endothelial cells can be mediated by VCAM-1 or CS-1 fibronectin. A role for VLA-4 in arthritis has been suggested by a number of experimental observations. In adjuvant arthritis in rats, anti-$$\alpha_4$$ antibody decreases lymphocyte accumulation in the joint but not lymph nodes, suggesting that VLA-4 is more important in recruitment to inflamed sites than to noninflamed sites. T lymphocytes isolated from the synovial fluid and synovial membrane of RA patients exhibit increased VLA-4–mediated adherence to VCAM-1, relative to autologous peripheral blood lymphocytes. These studies also suggest that leukocytes expressing functionally activated VLA-4 are selectively recruited to inflammatory sites in RA. Anti-$\alpha_4$ antibody has potential utility in RA, but enthusiasm is mitigated by effects on host defense observed with natalizumab in multiple sclerosis.

Moderate amounts of VCAM-1 are expressed in RA synovial blood vessels. Surprisingly, the intimal lining is the location of the most intense staining with anti–VCAM-1 antibodies on histologic sections. Even normal synovial tissue expresses VCAM-1 in the lining, albeit less than in RA. Cultured FLS constitutively express small amounts of VCAM-1, and the level is increased by a variety of macrophage and T cell–derived cytokines. VCAM-1 also contributes to T cell adhesion to high endothelial venules in frozen sections of RA synovium. The other VLA-4 counter-receptor, CS-1–containing forms of FN, is restricted to inflamed RA vascular endothelium and the synovial intimal lining.

The integrin $\alpha_4\beta_1$, which also binds to VCAM-1, is a specific adhesion molecule involved in lymphocyte homing to Peyer’s patches. Most intraepithelial and lamina propria lymphocytes express $\alpha_4\beta_1$; this molecule is rarely identified in other lymphoid tissues. The expression of $\alpha_4\beta_1$ on peripheral blood lymphocytes from patients with RA is low (similar to normal individuals), but up to a quarter of synovial fluid lymphocytes, mostly CD8$^+$ T lymphocytes, express this adhesion molecule and provide an interesting link between arthritis and the gut.

### Selectins

E-selectin expression is also elevated in rheumatoid synovium, although the increase is less dramatic than for the integrins and their counter-receptors. This might be due, in part, to the kinetics of E-selectin expression on endothelial cells, which is transient after stimulation with cytokines.

### Therapeutic Potential of Blocking Adhesion Molecules

The therapeutic potential for antiadhesion therapy has been studied in the SCID mouse model. Labeled human peripheral mononuclear cells were injected into engrafted mice, and migration into the tissue was examined. If the mice were treated with TNF, ICAM-1 expression and trafficking into synovium were significantly increased. Anti–ICAM-1 antibody blocked leukocyte migration into the explant under these conditions. In another study, tonsil mononuclear cells also migrated into the RA synovial grafts in SCID mice. RA clinical trials using anti–ICAM-1 therapy have been reported using anti–ICAM-1 antibody or antisense ICAM-1 oligonucleotides, although minimal significant clinical benefit was observed. In addition, mice
lacking E- and P-selectin actually had accelerated disease in the collagen-induced arthritis model. This paradoxical result serves as a reminder of the complexity of the inflammatory process.

CARTILAGE AND BONE DESTRUCTION

KEY POINTS

Cartilage degradation and bone destruction in RA are mediated by distinct mechanisms and cell types.

Several classes of proteases including metalloproteinases, serine proteases, cathepsins, and aggrecanases are produced by intimal lining cells in RA, especially FLS.

Synovial lining cells, especially FLS, can invade and damage cartilage in RA.

Bone destruction is mediated by osteoclasts that are activated under the influence of RANKL and other cytokines produced by RA synovium.

Cartilage Destruction and the Pannus-Cartilage Junction

In RA, the cartilage is initially covered by a layer of tissue composed of mesenchymal cells. In the established lesion, macrophage-like and fibroblast-like cells penetrate into cartilage matrix (Figure 69-16). Invasive pannus is more commonly found in metatarsophalangeal joints, compared with hip and knee joints in which a layer of resting fibroblasts appeared to separate pannus from cartilage, perhaps explaining why erosions occur more often around small joints.

FLS from the intimal lining are major effectors of cartilage destruction in RA. They produce prodigious amount of proteases, bind to cartilage, and invade into the extracellular matrix. The pivotal role of synoviocytes in cartilage destruction was demonstrated in arthritis models using cadherin-11 blockade to disrupt the intimal lining.

Cartilage destruction is markedly attenuated even though bone erosions progress.

Other cells in the joint, especially neutrophils and cells from the pannus that burrow directly into cartilage, could also be responsible for cartilage, whereas osteoclasts are responsible for bone erosions. More primitive mesenchymal cells isolated directly from the cartilage-pannus junction express phenotypic and functional features of both synoviocytes and chondrocytes and have also been described in the synovium.

Cartilage is destroyed in RA by both enzymatic and mechanical processes. The enzymes induced by factors such as IL-1, IL-17, TNF, phagocytosis of debris by synovial cells, and mechanical trauma degrade the matrix proteins. Early in synovitis, proteoglycans are depleted from the tissue, most likely due to the catabolic effect of cytokines such as IL-1 on chondrocytes with the production of MMPs and aggrecanases, and this leads to mechanical weakening of cartilage. As proteoglycans are depleted, cartilage loses elasticity and becomes susceptible to mechanical fragmentation and fibrillation. Eventually the tissue loses functional integrity concurrent with its dissolution by collagenases and stromelysins. Some of the MMPs responsible for this process are also derived from the chondrocytes themselves. Multiple MMPs, especially stromelysin and collagenase levels, are expressed in RA cartilage, and in situ hybridization studies confirm the presence of the mRNA within chondrocytes. Hence the cartilage is under attack from a multitude of sources: It is bathed in protease-rich synovial fluid, is under extrinsic attack from the invasive pannus, is damaged from within by chondrocytes, and is fragmented by mechanical forces.

Enzymes released by PMNs in synovial fluid including neutrophil collagenase and multiple serine proteases also contribute to cartilage loss. Immune complexes containing RFs are embedded in the superficial layers of cartilage and can attract and activate neutrophils. Electron microscopic examinations of articular cartilage in RA reveal evidence of breakdown of collagen and proteoglycan due to superficial activity of joint fluid enzymes. In a rabbit model of arthritis in which IL-1 was injected directly into the joint, the degree of cartilage damage as measured by proteoglycan levels in synovial fluid correlated best with the stromelysin concentrations in synovial effusions (presumably derived from synoviocytes). Neutrophil depletion of animals did not interfere with subsequent destruction of extracellular matrix, suggesting that MMPs derived from the synovium are more important.

By and large, most animal studies indicate that IL-1 is a key regulator of matrix degradation in arthritis. Although TNF blockade has clear anti-inflammatory effects, chondroprotection is less prominent. The joint destruction observed in TNF-dependent models often requires IL-1. Recent data suggest that IL-17 and TLR ligands can also contribute to joint destruction directly or by synergizing with IL-1 and TNF.

The rate-limiting step in cartilage loss is the cleavage of collagen because proteoglycans are degraded soon after inflammation begins. MMPs, released into the extracellular space and active at neutral pH, are probably responsible for most of the effective proteolysis of articular-cartilage proteins, but other classes of enzymes may contribute to joint...
destruction. Enzymes such as cathepsins B, D, G, K, L, and H may play a role within and outside cells in degrading noncollagenous matrix proteins. Serine proteinases (e.g., elastase and plasmin) and aggrecanases are doubtless involved as well.

Proteases: Mediators of Joint Destruction

Matrix Metalloproteinases

The MMPs are a family of enzymes that participate in extracellular-matrix degradation and remodeling (see Chapter 8). They are usually secreted as inactive proenzymes, and their proteolytic activity requires limited cleavage or denaturation to reveal a zinc cation at the core. Their activation can be mediated by other proteases including trypsin, plasmin, or tryptase. The substrates for MMPs are varied but quite specific for individual members of the family. Collagenases degrade native collagen types I, II, III, VII, and X, whereas gelatinases are able to degrade denatured or cleaved collagen. Stromelysins have broader specificity and can digest proteoglycans in addition to proteins. They also process procollagenase to the active form, thereby serving as a positive-feedback signal for matrix destruction. Some MMPs such as TNF convertase (TACE) are responsible for the processing and release of cytokines from the cell surface. Many different families of proteinases are found in the joint (Table 69-6), but the MMPs are thought to play a pivotal role in joint destruction.

Regulation of MMP Production. The cytokine milieu has the capacity to induce the biosynthesis of MMPs by synovial cells and alter the balance between extracellular matrix production and degradation. IL-1 and TNF, in particular, induce MMP gene expression by many cells, especially FLS and chondrocytes. The two cytokines are additive or synergistic when used in combination. Many other cytokines and TLR ligands implicated in rheumatoid synovitis can also induce MMP expression including IL-17, LIF, LPS, and peptidoglycans.

MMP induction is mediated by both an increase in gene transcription and mRNA stabilization. Culture medium from rheumatoid synovium stimulates cartilage degradation in vitro, and this is mainly due to IL-1. IL-6 does not induce MMP production by synovial cells but instead increases the production of TIMP-1, a naturally occurring inhibitor of MMPs. TGF-β inhibits collagenase synthesis and enhances the production of TIMP by fibroblasts and chondrocytes. TGF-β also increases collagen production, shifting the balance from destruction to matrix repair.

Although multiple upstream regulatory sequences are involved in MMP gene transcription, the dominant element in the promoter is AP-1. Other regulatory sites such as an NFκB-like region can also contribute to collagenase expression. AP-1 activity is markedly increased in FLS by proinflammatory cytokines, and its transcriptional activity is mediated by increased expression of components such as c-Jun. The MAP kinases are especially important for this activity, and JNK is the most efficient upstream activator. Glucocorticoids markedly inhibit MMP gene expression by blocking AP-1.

Collagenases and stromelysins have the capacity to degrade virtually all the important structural proteins in the extracellular tissues within joints. Collagenase-1 (MMP-1) cleaves through the triple-helical collagen molecule at a single glycine-isoleucine bond approximately three-quarters of the distance from the NH2-terminus. This enzyme degrades only the interstitial helical collagens (e.g., types I, II, III, and X). It has little or no activity against types IV, V, and IX and other nonhelical collagens or denatured collagen; the degradation of the latter is primarily accomplished by the gelatinases. MMP-1, however, is a relatively inefficient enzyme, whereas collagenase-3 (MMP-13) has more favorable kinetics. Neutrophil collagenase, or MMP-8, is constitutively stored in neutrophil granules and is released into the milieu after degranulation. Of note, rodents lack the collagenase-1 gene, whereas the collagenase-3 gene is preserved. This is especially important to note when evaluating effects of MMP inhibitors in animal models.

MMP Expression in Synovium. The collagenase-1 and collagenase-3 genes are produced by RA synovial tissue, and the latter is highly expressed by chondrocytes in cartilage. In situ hybridization studies show that the primary location of collagenase-1 gene expression in the synovium, like many other MMPs, is the intimal lining, especially in fibroblast-like cells. Subchondral bone is another region in which proteinase expression occurs in RA and could participate in bone resorption. Increased MMP gene expression is an early feature of RA and occurs during the first few weeks of clinically active disease. High expression of collagenase-1, as well as gelatinases such as MMP-2, early in disease correlates with rapidly progressive erosions. Similarly, increased blood levels of the proenzymes are also associated with more severe disease.

Stromelysin-1 (MMP-3) and the other members of the stromelysin family have no activity against most native collagens but effectively degrade type IV collagen, fibronectin, laminin, proteoglycan core protein, and type IX collagen. Stromelysin removes the NH2-terminal propeptides from type I procollagen and is integrally involved in the activation of procollagenase. Like collagenase, stromelysin gene expression occurs mainly in the synovial intimal lining (Figure 69-17). Despite the putative importance of this enzyme in matrix destruction, stromelysin knockout mice are susceptible to collagen-induced arthritis and develop as much joint destruction as mice with functional stromelysin. This observation led to decreased interest in stromelysin inhibitors to treat diseases such as RA.

Table 69-6  Key Proteases and Inhibitors in Rheumatoid Arthritis Synovium

<table>
<thead>
<tr>
<th>Protease</th>
<th>Inhibitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metalloproteinases</td>
<td>TIMP family; α2-macroglobulin</td>
</tr>
<tr>
<td>Collagenase-1</td>
<td></td>
</tr>
<tr>
<td>Collagenase-3</td>
<td></td>
</tr>
<tr>
<td>Stromelysin-1</td>
<td></td>
</tr>
<tr>
<td>92-kD gelatinase</td>
<td></td>
</tr>
<tr>
<td>Serine proteinases</td>
<td>SERPINs; α2-macroglobulin</td>
</tr>
<tr>
<td>Trypsin</td>
<td></td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td></td>
</tr>
<tr>
<td>Tryptase</td>
<td></td>
</tr>
<tr>
<td>Cathepsins</td>
<td>α2-macroglobulin</td>
</tr>
<tr>
<td>Cathepsin B</td>
<td></td>
</tr>
<tr>
<td>Cathepsin L</td>
<td></td>
</tr>
<tr>
<td>Cathepsin K</td>
<td></td>
</tr>
</tbody>
</table>

SERPINs, serine protease inhibitors; TIMP, tissue inhibitor of metalloproteinases.
MMP inhibitors are effective in animal models of rheumatoid arthritis and can suppress bone destruction, as well as the inflammatory synovitis. In models of osteoarthritis, deletion of MMP genes such as stromelysin do not necessarily improve outcomes. Clinical trials in RA using nonselective inhibitors have had minimal success and significant side effects, possibly related to decreased matrix turnover. Inhibitors of TACE (which can also block other MMPs) actually appear to increase disease activity in RA, perhaps due to increased levels of membrane-bound TNF. One of the most consistent side effects experienced by patients treated with MMP inhibitors is increased joint stiffness thought to result from deposition of fibrous tissue without sufficient protease activity to permit removal of matrix proteins. This observation has been replicated in rats, which provides an opportunity to determine if highly selective MMP inhibitors will have a better risk-benefit ratio.

Cysteine Proteases: The Cathepsins

Cathepsins are an extensive family of cysteine proteases that have broad proteolytic activity including activity on types II, IX, and XI collagen and proteoglycans. Like MMPs, the cathepsins are regulated by cytokines and by proto-oncogenes such as Ras. IL-1 and TNF induce cathepsin L expression in cultured FLS. In situ hybridization studies demonstrate expression of cathepsin B and L in RA synovium, especially at sites of erosion. A ribozyme that cleaves cathepsin L decreases FLS invasion and cartilage destruction in the SCID mouse model with implanted cultured synoviocytes.

Cathepsin K has been implicated in bone resorption by osteoclasts. This protease is unique among the cathepsins because it can degrade native type I collagen. It is expressed in RA synovial tissue by both macrophages and fibroblasts and is present in significantly higher concentrations than in OA. Serum levels of cathepsin K correlate with the extent of radiographic damage. A potential role of cathepsins as mediators of bone destruction in arthritis was confirmed in studies in which a cysteine protease inhibitor significantly decreased joint damage in the rat adjuvant arthritis model. In the TNF-transgenic mouse model, lack of the cathepsin K deficiency decreased, but did not eliminate, bone erosions.

Aggrecanases

Aggrecan is a major proteoglycan component of articular cartilage. Because of its large size and negative charge, it contains a considerable amount of water, which increases compressibility. Two proteolytic sites are available on aggrecan in its globular domain. One site is susceptible to MMP cleavage, whereas the other, located 32 amino acids toward the C-terminus, is the site for cleavage by a family of enzymes known as aggrecanases. The two sites can be identified in tissues using monoclonal antibodies after cleavage when specific neoepitopes are revealed.

Normal cartilage contains a surprising amount of aggrecan neoepitope, suggesting continuous matrix turnover. The level of aggrecanase cleavage product increases with age. Two aggrecanase genes, aggrecanase-1 and aggrecanase-2, have been cloned and are members of the "a disintegrin and metalloproteinase with thrombospondin motif" (ADAMTS) family of proteins (ADAMTS-4 and ADAMTS-5, respectively). They are expressed in OA and RA cartilage, and their proteolytic activity can be detected in synovial fluids. Especially high levels of the neoepitope are present in arthritic cartilage. IL-1 increases aggrecanase expression in cartilage explants, as well as cultures of chondrocytes. Aggrecanase-1 and aggrecanase-2 are
Inhibitors of Protease Activity

\( \alpha_2 \)-Macroglobulin (\( \alpha_2 \)M) accounts for more than 95% of collagenase inhibitory capacity in serum. The mechanism of inhibition by \( \alpha_2 \)M involves hydrolysis of the protease by the protease of a susceptible region in one of the four polypeptide chains of \( \alpha_2 \)M (sometimes called the “bait”), with subsequent trapping of the proteins within the interstices of the \( \alpha_2 \)M. Ultimately, the protease is covalently linked to a portion of the \( \alpha_2 \)M molecule. The serine protease inhibitors (SERPINs) are also abundant in synovial effusions and plasma and can serve a dual purpose of directly blocking serine protease function and indirectly decreasing MMP activity by preventing serine proteases from activating MMP proenzymes. One SERPIN, \( \alpha_1 \)-antitrypsin, has been well characterized in synovial fluid and is frequently inactivated after oxidation by reactive oxygen species.

A family of proteins that specifically block MMP activity, called TIMPs, has been cloned and characterized. The TIMP proteins block proteinase activity by binding directly to MMPs in a 1:1 molar ratio. TIMP generally binds only to the active enzyme, although exceptions such as TIMP-2 can interact with a progelatinase (MMP-2). The inhibitors bind to MMPs with extremely high avidity. Even though the interaction does not result in new covalent bonds, it is essentially irreversible.

TIMP proteins are present in RA synovial fluid in excess. It is, in fact, difficult to detect free active collagenses and stromelysin because they are usually complexed with the inhibitors. The majority of MMP, however, is in the proenzyme form. Immunohistochemical and in situ hybridization studies have localized the TIMPs in hyperplastic synovial lining cells in rheumatoid synovium, but not in the cells of normal synovium. TIMP gene expression is not significantly altered by IL-1 or TNF but is increased by IL-6, oncostatin M, and TGF-\( \beta \). TIMP-3 knockout mice have significantly more synovial inflammation and TNF production in antigen-induced arthritis, perhaps because it is not available to inhibit TACE. Similarly, TIMP-1 or TIMP-3 gene transfer limits rheumatoid FLS invasion into cartilage in a SCID mouse model. The function of these genes can extend beyond protease inhibition and include a number of paracrine functions, as well as induction of apoptosis when expressed intracellularly in cultured synoviocytes.

Given the important role of MMPs in tissue destruction, the relative balance between MMPs and TIMPs ultimately determines the fate of the extracellular matrix. The ratio in RA, with its more destructive potential, favors degradation, whereas OA has a lower MMP-to-TIMP ratio. The levels of TIMP gene expression are similar in the two diseases and may well be maximal. The higher ratio in RA results from increased MMP production. This balance between protease and inhibitor can be modified in vivo with drug therapy. For instance, intra-articular corticosteroid injections markedly decrease synovial collagenase, stromelysin, and TIMP gene expression. In contrast, chronic low-dose methotrexate therapy specifically decreases collagenase but not TIMP-1 mRNA. Suppressed collagenase gene expression suggests that a low collagenase-to-TIMP ratio is one mechanism of decreased tissue destruction observed in patients treated with methotrexate.

Regulation of Bone Destruction

Osteoclasts are the major cells responsible for bone degradation. RANKL, which was originally described for its role in T cell–dendritic cell interactions, as well as lymphocyte and lymph node development, is perhaps the single most important factor that modulates bone resorption. Osteoclast development is complex and involves the differentiation of monocytes under the influence of cytokines such as M-CSF in combination with RANKL. Subsequent osteoclast activation can involve several pathways, most of which also depend on the presence of RANKL. Its receptor, known as RANK, is expressed by the osteoclast precursors. RANKL is produced by many cell types including activated T cells and FLS.

Abundant evidence implicates this powerful mechanism in bone destruction caused by inflammatory arthritis. For instance, administration of OPG, a RANKL decoy receptor, to rats with adjuvant arthritis inhibits bone destruction but has almost no effect on inflammation or clinical signs of arthritis.\(^{202}\) RANKL-\( ^{-}\) mice are also protected from bone erosions in the passive K/BxN model of arthritis, although cartilage destruction still occurs. Animal models of arthritis point to IL-17 as a mediator of osteoclast generation. Genetic deficiency of IL-17 or anti-IL-17 antibodies have remarkable bone-sparing effects in these experiments.

RANK, RANKL, and OPG (as well as M-CSF and IL-17) have been detected in the synovium and synovial fluid of patients with RA. The ratio of RANKL to OPG is significantly higher in RA synovial effusions than in either OA or gout, which is consistent with the more destructive nature of RA. Osteoclasts expressing tartrate-resistant acid phosphatase (TRAP), capable of forming resorption lacunae, can be generated from cultured RA synovial cells (Figure 69-18). This activity is blocked by the addition of exogenous OPG. RA synoviocytes and synovial membrane T cells that display RANKL can also induce differentiation of osteoclasts from peripheral blood cells.\(^{201}\)

The functional relevance of the RANK-RANKL system was confirmed in RA studies in which an anti-RANKL antibody denosumab decreased bone erosions. As predicted from animal model studies, the antibody had no effect on inflammation or clinical signs of synovitis.\(^{202}\)

A second system that regulates bone remodeling involves the Wingless (Wnt) proteins. Several members of this family bind to receptors that regulate osteoblast differentiation through effects on \( \beta \)-catenin. Wnt signaling is modulated by many other proteins, most notably the Dickkopf (DKK) family. DKK-1 in particular blocks binding of Wnts to its receptors. In TNF-transgenic mice with arthritis, bone erosions are blocked by inhibiting either TNF or DKK-1. With the latter, proliferative bone lesions such as osteophytes formed instead.\(^{203}\) Therefore DKK-1 is a master switch that determines the fate of bone in inflammatory
lesions. When present, bone destruction is favored; when absent, bone formation occurs.

**Tissue Repair**

Extracellular matrix turnover in RA has been likened to wound healing due to the critical role of collagen production, proteases, and protease inhibitors. Remodeling the matrix by removing damaged proteins is a key element in early repair. Subsequently, the balance shifts to protease inhibition, production of cytokine inhibitors, removal of inflammatory cells through apoptosis, and release of anti-inflammatory eicosanoids such as lipoxins to suppress inflammation. Neutralization of oxidants via glutathione reductase or superoxide dismutase further limits tissue damage.

This process then permits either a return to normal architecture or scar formation. TGF-β, in particular, appears to play a key role in that it increases collagen deposition, suppresses MMP expression, and enhances production of the TIMPs. Although TGF-β levels in the joint are substantial, they are not sufficient to overcome the impressive array of MMPs expressed in synovitis. The repair process is insufficient in RA, perhaps because of persistent T cell activation or autonomous activation of other cell lineages. However, strategies to shift from tissue damage by enhancing endogenous mechanisms might not only suppress symptoms but also enhance appropriate remodeling of the matrix to restore homeostasis.

Because the invasive rheumatoid synovium exhibits some properties similar to neoplastic diseases, the possibility that the tissue contains immature cells or embryonic genes that regulate repair has been explored. The embryonic growth factors from the wingless (wnt) and frizzled (fz) gene families have been demonstrated in RA synovium. Normally, these proteins participate in bone marrow progenitor differentiation and limb bud mesenchyme. Wnt5a and Fz5, in particular, are markedly elevated in RA tissues and cultured synoviocytes. When normal fibroblasts are transfected with the WNT5A gene, cytokine expression such as IL-6 increases significantly. Antisense WNT5A and dominant negative WNT5A vectors diminish cytokine expression by synoviocytes.

These data raise the possibility that immature mesenchymal cells populate the synovium in RA, either as a primary event or as a repair mechanism. Similar primitive mesenchymal cells circulate in the peripheral blood of RA and normal individuals, and in collagen-induced arthritis they infiltrate the synovium before clinically apparent synovial inflammation.

Restoring homeostasis and tissue repair in RA is therefore a complex process that involves the ingress or dedifferentiation of mesenchymal cells that can remodel the matrix. In addition to TGF-β, the function of these cells is modulated by the bone morphogenic proteins (BMPs). The BMPs are members of the TGF-β superfamily and, like TGF-β, signal through the Smad pathway. Several members including BMP-2 and BMP-7 are expressed in the joint and facilitate repair, although inappropriate release can also enhance joint damage or lead to ankylosis or enthesophyte formation. BMP function is also regulated by a family of inhibitors such as Noggin, which can surprisingly limit cartilage damage when overexpressed in murine antigen-induced arthritis. Modulating the relative balance and timing of BMP expression could ultimately be used to either modify the destructive influence of synovitis or regenerate damaged tissues.

**SUMMARY**

The etiology and pathogenesis of RA remains a complex problem, although the level of understanding has progressed considerably in recent years. Both T cell–dependent and T cell–independent processes contribute to disease initiation and perpetuation. Moreover, disease mechanisms might differ at various stages of the process. These hypotheses have unveiled many novel therapeutic targets and interventions that might lead to significant clinical benefit. Such was the case with the TNF inhibitors, B cell depletion, T cell co-stimulation, and most recently IL-6 inhibition, which have joined the pharmacopoeia for the treatment of RA. Early observations that defined the cytokine profile in arthritis and that delineated the biology of macrophage cytokines led to this breakthrough. Similarly, it is possible that understanding of apoptotic pathways, abnormalities in tumor-suppressor genes, the function of the susceptibility genes, signaling pathways, B cell function, or T cell differentiation will lead to new therapies.

**Selected References**


Full references for this chapter can be found on www.expertconsult.com.
References


KEY POINTS

- Rheumatoid arthritis is a symmetric inflammatory polyarticular arthritis that mainly affects the small joints of the hands and feet.
- Larger joints can be involved, usually later and in a symmetric fashion.
- Cartilage destruction and bone erosions are common, especially in rheumatoid factor-positive or anticitrullinated protein antibody-positive patients.
- Uncontrolled synovitis can lead to severe deformity, loss of function, and increased mortality due to accelerated atherosclerosis.
- Early therapy with aggressive treatment goals improves long-term outcomes in rheumatoid arthritis.
- Systemic manifestations include rheumatoid nodules, pulmonary disease, vasculitis, serositis, and eye disease.

EPIDEMIOLOGY AND THE BURDEN OF DISEASE

The prevalence of rheumatoid arthritis (RA) in most populations is around 1% in most populations, with an incidence in women twice that in men. This number was based on many studies of population samples, which varied among the surveys from 0.3% to 1.5%. The prevalence of RA in some populations might be changing, however, as suggested by more recent data on incidence rates in different decades. The incidence of RA in Rochester, Minnesota, decreased by 50% between 1950 and 1974. Differences between incidence and prevalence are enhanced by the realization that as the population ages, the prevalence of RA may increase or stay the same, regardless of short term trends in incidence simply because individuals with RA are living longer.

The incidence of RA increases during adulthood, except among men in their 40s through 60s. In Olmsted County, Minnesota, the increased incidence with increasing age continues until age 85, after which the incidence declines. In a 10-year extension of this study, the age-adjusted and sex-adjusted incidence per 100,000 population decreased from 62 in the decade 1955 to 1964 to 32.7 in the decade 1985 to 1994. The decrease was more prominent in women than in men, and the average age at onset of the disease shifted upward. Perhaps more intriguing were cyclic patterns of incidence within decades, suggesting the influence of environmental factors. One explanation for the decline in incidence and the shift toward older age at onset is a birth cohort effect, the greatest impact of which is seen early in life. A recent update from the Olmsted County, Minnesota, cohort of RA patients from 1955-2007 examined trends in the incidence and prevalence of RA from 1995-2007. During the more recent time period, the incidence of RA in women, but not in men, increased moderately. Causes for this trend reversal in women were not determined but might involve environmental factors. Current incidence rates using this same population-based study of Minnesota patients revealed that the lifetime risk of RA among U.S. adults is 3.6% for women and 1.7% for men.

Throughout the world, pockets of ethnic groups have a much higher incidence of RA. Native Americans constitute one of these groups. In one geographic area between 1986 and 1994, non–Native American populations had an RA prevalence of 1.1% to 0.9%, whereas the prevalence among Algonquian Indians in the same region ranged from 2% to 2.1%, and disease onset was seen 12 years earlier in the Native American population. Among Pima Indians, who bear a very high incidence of RA, a decline in incidence has been correlated with a decrease in seropositivity for rheumatoid factor (RF). The highest likelihood of seropositivity was noted in Pima Indians born at the turn of the 20th century, and seropositivity has decreased ever since that time. This provides additional supportive evidence for a birth cohort effect.

Although newer, more effective therapy for rheumatoid patients has led to reduced morbidity and disability from the disease, dollar costs for RA, which recently surpassed the cost per patient for diabetes, are still substantial. In a panel of individuals with RA in San Francisco followed for 15 years, medical care costs for RA averaged $5,919 per year, and additional costs of $2,582 were incurred for medical but non-RA reasons. More than half of these costs were generated by hospitalization, with some patients bearing costs greater than $85,000/year while their function declined. In another cohort of 4258 patients with RA followed for 17,085 patient-years, lifetime direct medical care costs were estimated to be $93,296.

RISK FACTORS

A predisposition to RA appears to be multifactorial based on the following: (1) Relatively few identical twins have RA (about 15%), even though concordance for the disease is much more likely in twins than in the normal population; (2) despite the powerful influence of the “shared epitope” on HLA-DRB chains in predisposing to the severity of disease, this susceptibility allele is not a risk factor in certain population studies; and (3) the combination of many gene...
polymorphisms confers a modestly increased risk for disease. A reasonable hypothesis is that the genetic predisposition to RA involves a propensity to autoimmune responses, but that repeated exposure to environmental agents is ultimately responsible for tipping the balance from subclinical autoimmunity to diseases such as RA. Many of the risk factors for RA are discussed in Chapter 69, especially genetic associations, environmental exposures, and the role of autoantibodies.

CLINICAL PRESENTATIONS OF EARLY RHEUMATOID ARTHRITIS

In the Northern hemisphere, the onset of RA is more frequent in winter than in summer. In several series, onset of RA from October to March in the Northern hemisphere was found to be twice as frequent as in the other 6 months. The appearance of RF or anticitrullinated protein antibodies (ACPAs), also referred to as anticyclic citrullinated protein antibodies (anti-CCPs), often precedes symptoms of arthritis in patients. Approximately half of patients with RA have specific serologic abnormalities several years before the onset of symptoms. A finding of an elevated serum level of immunoglobulin (Ig)M-RF or anti-CCP in a healthy person correlates with increased risk of developing RA. This is especially important in light of the new criteria for RA classification in early RA because symptoms can be minimal for some of those who meet the criteria for diagnosis (see later).

Patterns of Onset

Insidious Onset

RA has an insidious, slow onset over weeks to months in 55% to 65% of cases. The initial symptoms may be systemic or articular. In some people, fatigue, malaise, swollen hands, and diffuse musculoskeletal pain may be the first nonspecific symptoms, with joints becoming involved later. Involvement of tendon sheaths early in the process can focus attention on periarticular structures. In retrospect, the patient often can identify one joint that was involved first, quickly followed by the others. Asymmetric initial presentations (often with increased symmetry developing later in the course of disease) are common. The reason for the symmetry of joint involvement compared with other forms of arthritis, such as the seronegative spondyloarthropathies, is unknown.

Morning stiffness is a cardinal sign of inflammatory arthritis that can appear even before pain and may be related to the accumulation of edema fluid within inflamed tissues during sleep. Morning stiffness dissipates as edema and products of inflammation are absorbed by lymphatics and venules and returned to the circulation by motion accompanying the use of muscles and joints. To be specific for joint inflammation, morning stiffness (e.g., “difficulty moving around”) should persist for at least 30 to 45 minutes before disappearing. A similar “gel” phenomenon can occur if a person is inactive during the day.

It is rare for symptoms to remit completely in one set of joints while developing in another. This quality of arthritis sets RA apart from rheumatic fever or palindromic rheumatism, in which a true migratory pattern of arthritis is common. A subtle, early change in RA is the development of muscle atrophy around affected joints. Muscle efficiency and strength become diminished. As a result, weakness develops that can be out of proportion to pain. Opening doors, climbing stairs, and doing repetitive work rapidly become more demanding. A low-grade fever without chills is rarely present. Depression and anxiety can accentuate symptoms. A small but significant weight loss is common and reflects the catabolic effects of cytokines and associated anorexia.

Acute or Intermediate Onset

Among patients with RA, 8% to 15% have an acute onset of symptoms that peak within a few days. Rarely, a patient can pinpoint the onset of symptoms to a specific time or activity. Symptoms mount, with pain developing in other joints, often in a less symmetric pattern than in patients who have an insidious onset. Acute-onset RA is difficult to diagnose, and sepsis or vasculitis should be ruled out. Fever, suggesting an infectious process, can rarely be a prominent sign. An intermediate type of onset, in which symptoms develop over days or weeks, occurs in 15% to 20% of patients. Systemic complaints are more noticeable than in the insidious type of onset.

Joint Involvement

The joints most commonly involved first in RA are the metacarpophalangeal (MCP) joints, the proximal interphalangeal (PIP) joints, the metatarsophalangeal joints, and the wrists (Table 70-1). Larger joints generally become symptomatic after small joints. Synovitis in large joints is likely to remain asymptomatic for a longer time than in smaller ones, and a biopsy specimen of an asymptomatic knee often shows histologic evidence of synovitis. One anatomic study correlated the area, in square centimeters, of synovial membrane with that of hyaline cartilage in each joint. Joints with the highest ratio of synovium to articular

| Table 70-1 | Distribution of Joints Involved in Attacks Based on Cumulative Experience with 227 Patients |
|---|---|---|
| Joint Involvement | % Patients (Mean) | % Patients (Range) |
| MCP, PIP | 91 | 74-100 |
| Wrists | 78 | 54-82 |
| Knees | 64 | 41-94 |
| Shoulders | 65 | 33-75 |
| Ankles | 50 | 10-67 |
| Feet | 43 | 15-73 |
| Elbows | 38 | 13-60 |
| Hips | 17 | 0-40 |
| Temporomandibular | 8 | 0-28 |
| Spine | 4 | 0-11 |
| Sternoclavicular | 2 | 0-6 |
| Peri-articular sites | 27 | 20-29 |

MCP, metacarpophalangeal; PIP, proximal interphalangeal.

cartilage correlated positively with those most frequently involved in the disease (see Table 70-1).17

Early Synovitis: Which Patients Develop Rheumatoid Arthritis?

Distinguishing early RA from other inflammatory arthropathies can be challenging. In its earliest stages, RA might involve only a few joints and may not show the typical symmetric distribution. What diagnostic clues can be used to determine who will progress to classic RA, and who will develop an alternative inflammatory arthritis such as one of the spondyloarthropathies or have a spontaneous remission? The implications for disease management are obvious because early treatment potentially could limit or prevent joint damage and possibly permit long-term remission or even cure. Because 30% to 40% of patients with early inflammatory synovitis have spontaneous remission, accurate identification of patients with RA is essential to avoid undertreatment and overtreatment.

Some of these questions have been addressed by the Leiden Early Arthritis Clinic, which evaluates patients with symptoms of less than 2 years’ duration (most patients have symptoms for less than 6 months). In this cohort, only about 20% of patients met criteria for RA when initially evaluated by a rheumatologist.18 One-third of patients defied categorization and were considered to have “undiagnosticated arthritis.” When this group of patients was followed for 1 year, 27% ultimately developed RA, and 40% remained undiagnosticated. Clinical features that were more commonly seen among patients who developed RA included greater numbers of joints involved (mean of seven joints vs. four joints), longer duration of morning stiffness (90 minutes vs. 60 minutes), and the presence of autoantibodies. These features were insufficient individually to permit early diagnosis of RA, although a composite scoring system has been proposed.19 The predictive value of this system approaches 90%.

Additional studies from the Leiden Early Arthritis Clinic show that assessment and management of patients with symptoms of less than 12 weeks’ duration by a rheumatologist was associated with decreased joint destruction and increased likelihood of disease-modifying antirheumatic drug (DMARD)-free remission.20 Treatment of early inflammatory synovitis with methotrexate therapy delays progression to RA, but disease ultimately progresses to RA if the medication is discontinued21 (see Chapter 42).

Among predictive features most likely to be useful in patients, serum autoantibodies might be the most important. ACPAs, in particular, are strongly associated with the evolution of undifferentiated arthritis into RA and progression to erosive disease. In addition, the diversity of citrullinated peptides recognized by ACPAs increased during the period preceding onset of disease in patients who progressed to RA, suggesting that epitope spreading plays a role in the evolution of disease.22 Other autoantibodies have also been used in a diagnostic algorithm for patients with very early synovitis (symptoms of <3 months’ duration), including RFs, anticitrullinated protein, and anti-RA33.23 Through stepwise analysis of each antibody, RA could be diagnosed in 72% of patients and confirmed by subsequent clinical course and development of RA.

Other Patterns of Disease Onset or Variants of Disease

Palindromic Pattern

Palindromic rheumatism was described by Hench and Rosenberg in 1942. Pain usually begins with pain in one joint or in periarticular tissues; symptoms worsen for several hours to a few days and are associated with swelling and erythema. Then, in reverse sequence, symptoms resolve, leaving no residua. Table 70-2 lists joints involved in a series of 227 patients. An intercritical period, similar to that of gout, is asymptomatic. Half of patients with palindromic rheumatism go on to develop RA, particularly those with HLA-DR4. In a compilation of patients from nine series, only 15% became asymptomatic after at least 5 years with a palindromic syndrome (see Table 70-2).24 In the remainder, multiple joints became involved, swelling did not subside completely between attacks, and tests for RF became positive. Neither the characteristics of joint fluid nor the pathologic findings of synovial biopsy specimens allow the prediction that RA will evolve from palindromic rheumatism,25 although it may be worthwhile to measure ACPAs in these people.26 Those who do not develop RA rarely have constitutional symptoms, and involved joints show no erosion because the synovitis does not become chronic. A more recent study of long-term outcomes in 60 patients diagnosed with palindromic rheumatism revealed that two-thirds of patients developed chronic arthritis, and the risk for chronic arthritis remained for longer than 10 years.27 Of 51 patients with palindromic rheumatism, 41 experienced marked improvement in frequency and duration of attacks during treatment with antimalarials.28 The use of antimalarials might reduce the risk of progression to RA.

### Table 70-2  Evolution of Patients with Palindromic Rheumatism in Nine Series*

<table>
<thead>
<tr>
<th>No. of Cases</th>
<th>Remission or Cure (%)</th>
<th>Persistent PR (%)</th>
<th>RA (%)</th>
<th>Other Diseases (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total or average</td>
<td>653</td>
<td>15 ± 14</td>
<td>48 ± 20</td>
<td>33 ± 17</td>
</tr>
</tbody>
</table>

*In nine series of patients (653 total), the number undergoing remission or cure, remaining palindromic, evolving toward rheumatoid arthritis (RA), or developing another disease is expressed as an average percentage of the total patient population.

PR, Palindromic rheumatism; RA, rheumatoid arthritis

Insidious Onset in Older Individuals

Older individuals (≥65 years old) who develop RA often present with stiffness, limb girdle pain, and diffuse swelling of the hands, wrists, and forearms. Clinical onset that mimics polymyalgia rheumatica or remitting seronegative synovitis with pitting edema (RS3PE) also can occur in the elderly. Individuals with onset at age 60 years or older are less likely to have subcutaneous nodules or RF at the onset of disease, despite the high prevalence of RF in the general population in this age group. Generally, elderly individuals who develop RA tend to have a more benign course than younger people; the frequency of positive tests for RF is lower, but a strong association with HLA-DR4 has been noted. Onset is slow, but stiffness is often incapacitating.

In a study of patients with RA of less than 15 months’ duration, older patients had higher scores for joint space narrowing and osteophytes at baseline than patients younger than 55 years. However, no evidence suggested that older patients had more rapid progression of damage, indicating that osteoarthritis was responsible for a significant portion of the damage noted at the onset of disease.

Arthritis Robustus

Arthritis robustus is not so much an unusual presentation of disease as an unusual reaction of patients to the disease. Most patients are men whose disease is characterized by proliferative synovitis, often with deformity, which seems to cause little pain and even less disability. Patients are athletic and invariably keep working (often at physical labor). Periodic articular osteopenia is unusual, whereas new bone proliferation at joint margins near significant erosions of bone and cartilage is common. Bulky subcutaneous nodules develop. Subchondral cysts also develop, presumably from excessive pressure caused by synovial fluid within a thick joint capsule during muscular effort.

Rheumatoid Nodulosis

Whether rheumatoid nodulosis is a variant subset of RA or a different entity has not been clarified. The clinical picture includes recurrent pain and swelling in different joints, radiologic subchondral bone cysts, and subcutaneous rheumatoid nodules. In one series of 16 patients followed over 12 years, 6 had an aggressive course indistinguishable from classic erosive polyarticular RA. In 7 patients, cholesterol crystals were found in fluid from the olecranon bursae. Second-line drugs helped articular disease but did not help other components of the process.

COURSE AND COMPLICATIONS OF ESTABLISHED RHEUMATOID ARTHRITIS

Involvement of Specific Joints: Effects of Disease on Form and Function

The effects of rheumatoid synovitis on joints are a complex function of the intensity of the underlying disease, its chronicity, and the stress put on individual joints by the patient. Most well-documented observations of specific joint involvement and of complications of the disease were reported in the decades before 1980. Since then, these observations have been refined, but few new data have become available. Despite advances in our understanding of the pathophysiology of RA, including delineation of the cellular and enzymatic pathways that destroy joints, guidelines for the practicing physician—so that the probability that an individual patient would go on to develop erosive disease requiring aggressive treatment can be determined—are only in early stages of development. The spectrum of the clinical course of RA can range from patients who have mild pauciarticular synovitis, with negative serum autoantibodies and few radiographic changes, to those who have unrelenting pain, synovitis, joint damage, and extra-articular manifestations.

Hands and Wrist

The hand and wrist should be considered together because they form a functional unit. Data have linked disease of the wrist to ulnar deviation of the MCP joints. The hypothesis is that weakening of the extensor carpi ulnaris muscle leads to radial deviation of the wrist as the carpal bones rotate (the proximal row in an ulnar direction and the distal ones in a radial direction). In response to this, ulnar deviation of the fingers (a “zigzag” deformity) keeps the tendons to the phalanges in a normal line with the radius. Other factors, including the tendency for a power grasp to pull the fingers into an ulnar attitude and inappropriate intrinsic muscle action, are also involved (Figure 70-1). Erosion of bone or articular cartilage is not essential for the development of ulnar deviation (Figure 70-2). Significant but reducible ulnar deviation can result from repeated synovitis or muscle weakness in the hands (e.g., in systemic lupus erythematosus, in Parkinson’s disease).

Dorsal swelling on the wrist within the tendon sheaths of the extensor muscles is one of the earliest signs of disease. Typically, the extensor carpi ulnaris and extensor digitorum communis sheaths are involved. Rarely, cystic structures resembling ganglia are early findings of RA.

As synovial proliferation develops within the wrist, pressure increases within the relatively nondistensible joint spaces. Proliferative synovium develops enzymatic machinery sufficient to destroy ligaments, tendons, and the articular disk distal to the ulnar head. Pressure and enzymes combine to produce communications among radiocarpal, radioulnar, and midcarpal joints. The integrity of the distal radioulnar joint is lost. The ulnar collateral ligament, stretched by the proliferative synovium of the radioulnar joint, finally ruptures or is destroyed, and the ulnar head springs up into dorsal prominence, where it “floats” and is easily depressed by the examiner’s fingers (piano key styloid).

On the volar side of the wrist, synovial protrusion cysts develop; they can be palpated, and their origins can be confirmed by arthrography. The thick transverse carpal ligament provides significant resistance to decompression, however, and the hyperplastic synovium can compress the median nerve, causing carpal tunnel syndrome, often bilaterally.

Progression of disease in the wrist may be characterized by radiographic loss of joint space and bone or by ankylosis (Figure 70-3A). Ultrasound of the wrist correlates with function and classic signs of inflammation and is a
complementary tool in the evaluation of wrist arthritis in RA (Figure 70-3B). Early detection of carpal bone involvement by RA is also possible with magnetic resonance imaging (MRI), which reveals early synovial proliferation and carpal bone erosions. Bony ankylosis is associated with increased duration and severity of the disease and is found in joints that have been immobilized by pain, inflammation, treatment, or all of these.

The hand may have many joints involved in RA. A sensitive index of hand involvement is grip strength, which simultaneously stresses multiple hand joints. Muscular contraction causes ligamentous tightening around joints, compressing inflamed synovium. The immediate result is weakness, with or without pain; the reflex inhibition of muscular contraction due to pain may be a primary factor...
Three types of deformity have been described for the thumb:
1. MCP inflammation leads to stretching of the joint capsule and a boutonnière-like deformity.
2. Inflammation of the carpometacarpal joint leads to volar subluxation during contracture of the adductor hallucis.
3. After prolonged disease of both MCP joints, exaggerated adduction of the first metacarpus, flexion of the MCP joint, and hyperextension of the DIP joint result from the patient’s need to provide a means to pinch.

One of the most common manifestations of RA in the hands is tenosynovitis in flexor tendon sheaths; this can be a major cause of hand weakness. Tenosynovitis manifests on the volar surfaces of the phalanges as diffuse swelling between joints or as a palpable grating within flexor tendon sheaths in the palm and may occur in half of RA patients.

It is particularly important to diagnose de Quervain’s tenosynovitis of the extensors of the thumb because this condition causes severe discomfort and yet is easily treated. Pain originating from these sheaths can be shown by Finkelstein’s test, that is, ulnar flexion at the wrist after the thumb is maximally flexed and adducted.

Frequently, rheumatoid nodules develop within tendon sheaths and may “lock” the finger painfully into fixed flexion or cause “trigger” fingers. When they are chronic or recurrent, it may be necessary to inject the tendon sheath or, if that fails, to remove it surgically.

**Elbows**

RA rarely manifests with severe pain in the elbow, perhaps because the elbow is a stable hinge joint. Nevertheless, involvement of the elbow is common, and if lateral stability at the elbow is lost as the disease progresses, disability can be severe.

The frequency of elbow involvement varies from 20% to 65%, depending on the severity of disease in the patient populations studied. One of the earliest findings, often unnoticed by the patient, is loss of full extension. Because the elbow is principally a connecting joint between the hand and the trunk, the shoulder and the wrists can compensate partially for the loss of elbow motion.

**Shoulders**

RA of the shoulder not only affects synovium within the glenohumeral joint but also involves the distal third of the clavicle, various bursae and the rotator cuff, and multiple muscles around the neck and chest wall. Severe shoulder pain is often bilateral and can lead to sleep disorders because of difficulty finding a comfortable position. Involvement of the rotator cuff in RA also has been recognized as a principal cause of morbidity. The function of the rotator cuff is to stabilize the humeral head in the glenoid. Weakness of the cuff results in superior subluxation. Rotator cuff tears or insufficiency from other causes can be shown by shoulder arthrography or MRI. In a series of 200 consecutive patients with RA studied by arthrography, 21% had rotator cuff tears, and an additional 24% had evidence of frayed...
RA. Histories reveal that 55% of patients have jaw symptoms at some time during the course of their disease. Radiographic examination reveals structural alterations in 78% of the joints examined. An overbite can develop as the mandibular condyle and the corresponding surface of the temporal bone, the eminentia articularis, are eroded. Physical examination of the rheumatoid patient should include palpation of the temporomandibular joint for tenderness and auscultation for crepitus. Occasionally, patients have acute pain and an inability to close the mouth, necessitating intra-articular glucocorticoid therapy to suppress the acute process.

Temporomandibular joint abnormalities are common in nonrheumatoid populations. The only specific findings for RA in the temporomandibular joint are erosions and cysts of the mandibular condyle detected by CT or MRI. No correlation has been noted between clinical and CT findings of the temporomandibular joint in RA.

Cricoarytenoid Joints

The cricoarytenoid joints are small diarthrodial joints with an important function: They rotate with the vocal cords as the vocal cords abduct and adduct to vary the pitch and tone of the voice. Careful histories may reveal hoarseness in 30% of rheumatoid patients. This hoarseness is not disabling in itself, but the cricoarytenoid joints may become inflamed and immobilized, with the vocal cords adducted to the midline, causing inspiratory stridor. Autopsy examinations have shown cricoarytenoid arthritis in almost half of patients with RA, suggesting that much significant disease of the larynx may be asymptomatic. Although CT scans detected laryngeal abnormalities in 54% of patients with moderately severe RA, no symptoms suggested that these abnormalities would be found. In contrast, findings on indirect laryngoscopy, which detected mucosal and gross functional abnormalities (including rheumatoid nodules), were abnormal in 32% of the same patients and correlated with symptoms of sore throat and difficult inspiration. It follows that the latter examination should be performed in symptomatic rheumatoid patients. Asymptomatic cricoarytenoid synovitis occasionally may lead to aspiration of pharyngeal contents, particularly at night.

Sternoclavicular and Manubriosternal Joints

Sternoclavicular and manubriosternal joints, both possessing synovium and a cartilaginous disk, are often involved in RA. Because of their relative immobility, few symptoms are reported. Patients occasionally describe pain in the sternoclavicular joints, however, while lying on their sides in bed. When symptoms do occur, the physician must be concerned about superimposed sepsis. CT or MRI is useful for careful delineation of the sternoclavicular joint. Manubriosternal involvement is almost never clinically important, although by tomographic criteria, it is common in RA.

Cervical Spine

In contrast to other nonsynovial joints, such as the manubriosternal joint or the symphysis pubis, diskovertebral joints in the cervical spine often manifest osteochondral destruction in RA and on lateral radiographs may be found

**Figure 70-5** Abnormalities of the shoulder in rheumatoid arthritis. The Grashey posterior oblique view of a shoulder shows severe glenohumeral joint space narrowing with marginal erosion and cystic changes of the humeral head adjacent to the greater tuberosity (lower arrow). Elevation of the humeral head with respect to the glenoid indicates chronic rotator cuff tear. Tapering of the distal end of the clavicle is seen, along with widening of the acromioclavicular joint (upper arrow). (Courtesy Dr. Barbara Weissman.)

One likely mechanism behind tears is that the rotator cuff tendon insertion into the greater tuberosity is vulnerable to erosion by the proliferative synovitis that develops there. Previous injury and aging may predispose to the development of tears. Sudden tears may be accompanied by pain and inflammation so great as to suggest infection.

Standard radiographic examinations of the shoulder in RA reveal erosions and superior subluxation (Figure 70-5). Arthograms, in addition to showing tears of the rotator cuff, can show diffuse nodular filling defects, irregular capsular attachment, bursal filling defects, adhesive capsulitis, and dilation of the biceps tendon sheath (perhaps unique to RA). High-resolution computed tomography (CT) or MRI may provide much of this information without the need for invasive techniques.

Marked soft tissue swelling of the anterolateral aspect of the shoulders in RA may be caused by chronic subacromial bursitis rather than by glenohumeral joint effusions. In contrast to rotator cuff tears, bursal swelling is not associated with decreased range of motion or pain. Synovial proliferation within the subdeltoid bursa might explain the resorption of the undersurface of the distal clavicle seen in this disease. Rarely, the shoulder joint may rupture, with symptoms resembling those of obstruction of venous return from the arm.

Temporomandibular Joints

The temporomandibular joint is commonly involved in RA. Histories reveal that 55% of patients have jaw tendons.

In contrast, findings on indirect laryngoscopy, which detected mucosal and gross functional abnormalities (including rheumatoid nodules), were abnormal in 32% of the same patients and correlated with symptoms of sore throat and difficult inspiration. It follows that the latter examination should be performed in symptomatic rheumatoid patients. Asymptomatic cricoarytenoid synovitis occasionally may lead to aspiration of pharyngeal contents, particularly at night.

Sternoclavicular and Manubriosternal Joints

Sternoclavicular and manubriosternal joints, both possessing synovium and a cartilaginous disk, are often involved in RA. Because of their relative immobility, few symptoms are reported. Patients occasionally describe pain in the sternoclavicular joints, however, while lying on their sides in bed. When symptoms do occur, the physician must be concerned about superimposed sepsis. CT or MRI is useful for careful delineation of the sternoclavicular joint. Manubriosternal involvement is almost never clinically important, although by tomographic criteria, it is common in RA.

Cervical Spine

In contrast to other nonsynovial joints, such as the manubriosternal joint or the symphysis pubis, diskovertebral joints in the cervical spine often manifest osteochondral destruction in RA and on lateral radiographs may be found
probable vertebral artery compression. In the latter, paresthesias may occur in the shoulders or arms during movement of the head.

Physical findings suggestive of atlantoaxial subluxation include loss of occipitocervical lordosis, resistance to passive spine motion, and abnormal protrusion of the axial arch felt by the examining finger on the posterior pharyngeal wall. This process may produce microfractures of the vertebral end plates, disk herniation, and degeneration of disk cartilage. The atlantoaxial joint is prone to subluxation in several directions and is summarized later.

- The atlas can move anteriorly on the axis (most common). This results from laxity of the ligaments induced by the development of proliferative synovial tissue in adjacent synovial bursae or by fracture or erosion of the odontoid process.
- The atlas can move posteriorly on the axis. This can occur only if the odontoid peg has been fractured from the axis or destroyed.
- The atlas can sublux vertically in relation to the axis (least common). This results from destruction of the lateral atlantoaxial joints or of bone around the foramen magnum. Vertical (superior) migration of the odontoid can develop from unattended anterior or posterior subluxation.

The earliest and most common symptom of cervical subluxation is pain radiating up into the occiput. Two other serious, but less common, clinical patterns include slowly progressive spastic quadriplegia with painless sensory loss in the hands and transient episodes of medullary dysfunction associated with vertical penetration of the dens and probable vertebral artery compression. In the latter, paresthesias may occur in the shoulders or arms during movement of the head.

Physical findings suggestive of atlantoaxial subluxation include loss of occipitocervical lordosis, resistance to passive spine motion, and abnormal protrusion of the axial arch felt by the examining finger on the posterior pharyngeal wall. Radiographic views (lateral, with the neck in flexion) reveal more than 3 mm of separation between the odontoid peg and the axial arch. In symptomatic patients, films in flexion should be taken only after radiographs (including an open-mouth posteroanterior view) have ruled out an odontoid fracture or severe atlantoaxial subluxation. Studies have indicated that CT is useful for showing spinal cord compression by loss of posterior subarachnoid space in patients with C1 to C2 subluxation. MRI has proved particularly valuable in determining pathologic anatomy in this syndrome (Figure 70-7).

Neurologic symptoms often have little relationship to the degree of subluxation and may be related to individual variations in the diameter of the spinal canal. Symptoms of spinal cord compression that demand intervention include altered consciousness, syncope, and loss of sphincter control. Dysphagia, vertigo, convulsions, hemiplegia, dysarthria, nystagmus, and peripheral paresthesias are additional symptoms that require immediate attention.

Some of these symptoms may be related to compression of the vertebral arteries, which must wind through foramina in the transverse processes of C1 and C2, rather than to compression of the spinal cord.

The progression of peripheral joint erosions parallels cervical spine disease in RA. The two coincide in severity and timing; cervical subluxation is more likely to develop
Thoracic, Lumbar, and Sacral Spine

The thoracic, lumbar, and sacral portions of the spine are usually spared in RA. Exceptions include the apophyseal joints; rarely, synovial cysts at the apophyseal joint can impinge as an epidural mass on the spinal cord, causing pain, neurologic deficits, or both.

Hips

The hip is less frequently involved in early RA than in juvenile RA. Hip joint involvement must be ascertained by careful clinical examination; symptoms of hip synovitis include pain in the lower buttock or, more commonly, the groin. Pain on the lateral aspect of the hip is often a manifestation of trochanteric bursitis rather than true hip joint synovitis.

About half of patients with well-established RA have radiographic evidence of hip disease. In contrast to osteoarthritis, in which the femoral head usually migrates superiorly, symmetric thinning of the cartilage in RA leads to axial migration. The femoral head may collapse and be reabsorbed, and the acetabulum is remodeled and pushed medially, leading to protrusio acetabuli (Figure 70-8). Significant protrusion occurs in about 5% of all patients with RA.

Loss of internal rotation on physical examination correlates best with radiographic findings. Similar to the situation in other weight-bearing joints, the femoral head may develop cystic lesions that communicate with the joint space.

Knees

In contrast to the hips, synovial inflammation and proliferation in the knees are readily shown on physical examination. Early in knee disease, often within 1 week after the onset of symptoms, noticeable quadriceps atrophy leads to the application of more force than usual through the patella in patients with erosion of the hands and feet. In a series of patients with RA referred for hip or knee arthroplasty, 61% had radiographic evidence of cervical spine instability.

Is mortality increased in patients with atlantoaxial subluxation? Neurologic signs do not necessarily develop in patients with large subluxations. When signs of cervical cord compression do appear, however, myelopathy progresses rapidly, and 50% of patients die within 1 year. These patients are at risk for these complications if they sustain small falls, whiplash injuries, and general anesthesia with intubation. Cervical collars can be prescribed for symptomatic relief. Operative stabilization may be considered if symptoms are progressive.

Some data support the hypothesis that early C1-to-C2 fusion for atlantoaxial subluxation before the development of superior migration of the odontoid decreases the risk for further progression of cervical spine instability. The incidence of sustained neurologic deterioration related to surgery may be 6%; this emphasizes the importance of a skilled surgical team and of careful assessment of each patient. In many cases, surgical intervention in asymptomatic patients is riskier than conservative management despite the dire appearance of imaging studies.

Vertical atlantoaxial subluxation is important and may follow anterior or posterior subluxation. Symptoms associated with this collapse of the lateral support system of the atlas occur in patients with severe erosive disease. Neurologic findings include decreased sensation in the distribution of cranial nerve V, sensory loss in the C2 area, nystagmus, and pyramidal lesions.

Figure 70-7 Rheumatoid arthritis of the cervical spine. T2-weighted sagittal image shows low signal periodontoid pannus (P). Odontoid process appears irregular secondary to erosion (arrow). The atlantodental distance shows mild widening (solid line). Vertical subluxation can also be seen without signs of cord compression. The anterior subarachnoid space is compromised by disk protrusions at multiple levels. Erosions (arrowheads) are seen at the vertebral end plates at the C6-C7 level. (Courtesy Dr. Barbara Weissman.)

Figure 70-8 Bilateral protrusio acetabuli in rheumatoid arthritis. The medial acetabular margins protrude into the pelvis. Severe accompanying cartilage loss is evident. (Courtesy Dr. Barbara Weissman.)
to the femoral surface. Another early manifestation of knee disease in RA is loss of full extension—a functional loss that can become a fixed flexion contracture unless corrective measures are undertaken.

Flexion of a knee with a moderate to large effusion markedly increases intra-articular pressure. The increased intra-articular pressure may cause an outpouching of posterior components of the joint, producing a popliteal or Baker’s cyst. This can generate pressures so high in the popliteal space that it may rupture down or dissect into the calf or, less often, superiorly into the posterior thigh. Rupture occurs posteriorly between the medial head of the gastrocnemius and the tendinous insertion of the biceps. Clinically, popliteal cysts and their complications have several manifestations. An intact popliteal cyst may compress superficial venous flow from the lower leg, producing dilatation of superficial veins, edema, or both.\(^{45}\) Rupture of the joint posteriorly with extravasation of joint fluid into the calf may resemble acute thrombophlebitis with swelling and tenderness and may produce systemic signs of fever with leukocytosis. One helpful sign in identifying cyst rupture may be the appearance of a crescentic hematoma beneath one of the malleoli of the ankle.\(^{46}\) Although arthrography clearly defines the abnormal anatomy of a Baker’s cyst, this invasive procedure has been replaced by ultrasonography and, when necessary, MRI (Figure 70-9).

**Ankles and Feet**

Ankle involvement is usually mild in RA, but damage can occur in severe progressive forms of the disease. Clinical evidence for ankle involvement consists of cystic swelling anterior and posterior to the malleoli. Much of the stability of the ankle depends on the integrity of the ligaments holding the fibula to the tibia and these two bones to the talus. In RA, inflammatory and proliferative disease may loosen these connections by stretching and eroding the collagenous ligaments, causing erosions (Figure 70-10). The result is incongruity, which progresses to pronation deformities and eversion of the foot.

The Achilles tendon is a major structural component and kinetic force in the foot and ankle. Rheumatoid nodules develop in this collagenous structure, and spontaneous rupture of the tendon has been reported when diffuse granulomatous inflammation is present.\(^{47}\) The subtalar joint controls eversion and inversion of the foot on the talus; patients with RA invariably have more pain while walking on uneven ground; this is related to the relatively common subtalar joint involvement in RA. Progressive eversion at the subtalar joint, combined with foot pain, leads to a lateral subluxation beginning in the midfoot and the development of a rocker-bottom deformity. Midfoot disease leads to collapse of the arch, which contributes to difficulty walking because of pain.

More than one-third of patients with RA have significant disease in the feet (Figure 70-11). Metatarsophalangeal (MTP) joints are often involved, and gait is altered as pain develops during push-off in striding. Studies have shown that MTPs are the initial site of erosion in many patients. Downward subluxation of the metatarsal heads occurs soon after the MTP joints become involved, producing “cock-up” toe deformities of the PIP joints. Hallux valgus and bunion or callus formation occur if disease continues. Cystic collections representing outpouchings of flexor tendon sheaths often develop under the MTP joints.\(^{48}\)

---

**Figure 70-9** Magnetic resonance imaging (MRI) of the knee in rheumatoid arthritis. Sagittal fast spin echo T2-weighted fat-suppressed image allows excellent contrast. Synovial fluid is shown in white as a posterior fluid collection. (Courtesy Dr. Barbara Weissman.)

**Figure 70-10** Rheumatoid arthritis of the ankle. Diffuse loss of cartilage space can be seen with erosion of the fibula (arrows). Scalloping along the medial border of the distal fibula is designated the fibular notch sign and is a characteristic finding in rheumatoid arthritis. The hindfoot is in valgus alignment.
Patients with subluxation of metatarsal heads can develop pressure necrosis of the plantar surfaces. Alternatively, those who have subluxation of MTP joints often develop ulceration over the PIP joints that protrude dorsally (hammer toes). The net result is increased pressure on the MTP joints with a sensation described as “walking on marbles” by many patients. Changes caused by the progress of disease and destruction in the foot include intermetatarsal joint liga-
ment stretching in response to inflammation, spreading of the forefoot, anterior migration of the plantar fat pad, and dorsal subluxation of toes followed by plantar subluxation of the metatarsal heads. Concurrently, hallux valgus results in progressive overlap of the second and third toes on top of the great toe.

DIP joints of the foot are rarely affected in RA, but a functional rigid hallux caused by muscle spasms of the great toe intrinsic muscles in an effort to relieve pressure on the lesser metatarsal heads can be extremely painful and may require surgical intervention. Another cause of foot pain in rheumatoid patients is the tarsal tunnel syndrome. In a group of 30 patients with RA, erosions in the feet visible on radiographs and foot pain (4 patients; 13%) were shown by electrodiagnostic techniques to produce slowing of medial or lateral plantar nerve latency, or both.

**Extra-articular Complications of Rheumatoid Arthritis**

Generally, the number and severity of extra-articular features vary with the duration and severity of the disease. Several of these features may be related to extra-articular foci of an immune response, based on evidence of independent and qualitatively different production of RF in the pleural space, pericardium, muscle, and even meninges. Patients with systemic immune responses have true rheumatoid disease, not just RA. Other unusual proteins and protein complexes in the circulation of patients with active rheumatoid disease include antiphospholipid antibodies, circulating immune complexes, and cryoglobulins. Extra-articular manifestations of RA are associated with excess mortality.

**Rheumatoid Nodules**

The mature rheumatoid nodule has a central area of necrosis rimmed by a corona of palisading fibroblasts that is surrounded in turn by a collagenous capsule with perivascular collections of chronic inflammatory cells. The earliest nodules, nests of granulation tissue, have been identified measuring less than 4 mm. These nodules grow by accumulating cells that expand centrifugally, leaving behind central necrosis initiated by vasculopathy and compounded by protease destruction of the connective tissue matrix. Nodules occur in 15% to 20% of patients with definite arthritis or RA. They occur most often on extensor surfaces or pressure points, such as the olecranon process and the proximal ulna (Figure 70-12A), as well as on tendons (Figure 70-12B). They are subcutaneous and vary in consistency from a soft, amorphous, entirely mobile mass to a hard, rubbery mass attached firmly to the periostuem.

The appearance of nodules in unusual sites can lead to confusion in diagnosis; they sometimes can appear identical to other types of nodules such as tophi. Sacral nodules may be mistaken for bedsores if the overlying skin breaks down. Occipital nodules also occur in bedridden patients. In the larynx, rheumatoid nodules on the vocal cords may cause progressive hoarseness. Nodules can even be found in the heart and lungs (see later). Nodules on the sclera can produce perforation of this collagenous tissue. Multiple reports describe rheumatoid nodule formation within the central nervous system, involving leptomeninges more frequently than parenchyma. Some patients develop rheumatoid nodules within vertebral bodies, resulting in bone destruction and signs of myelopathy.

Careful histologic study of early lesions suggests that development of the nodule is mediated by affected small arterioles and resulting complement activation and terminal vasculitis. This immunologic response is linked to proliferation of resident histiocytes and fibroblasts and to an influx of macrophages from the circulation. Proliferation of cells and the supporting scaffold of connective tissue are mediated by cytokines expressed in patterns similar to those found in rheumatoid synovium. The cytokine profile of the rheumatoid nodule suggests a T helper-1 (Th1) gene expression pattern similar to that of synovial tissue, including expression of tumor necrosis factor (TNF), interleukin (IL)-1, IL-12, and interferon (IFN)-γ, but not IL-2. More recently, gene expression of IL-17 family members in rheumatoid nodules showed that IL-17A gene expression was minimal in contrast to that of synovial tissue. The absence of IL-17A could be due to decreased IL-23 expression in the nodule. Data from studies using monoclonal antibodies against receptors for complement C3b and C3bi, monocytes, activated macrophages, and HLA-DR molecules suggest that mononuclear phagocytes are constantly being recruited into peripheral layers and subsequently migrate into the palisade to constitute most of the cell population in this area. Other studies using cytochemical markers (nonspecific esterase and CD68 for macrophages, and prolyl hydroxylase for fibroblasts) indicate that a mixture of macrophages and fibroblasts makes up the cellular content of nodules. This evidence fits with data from nodule tissue in organ culture; similar to synovial tissue, cells in the palisading region have...
punched-out bone lesions and almost always occur in the presence of elevated serum urate concentration. Nodules of multicentric reticulohistiocytosis contain large, lipid-filled macrophages. Numerous proliferative disorders that affect cutaneous tissue, including erythema elevatum diutinum, acrodermatitis chronica atrophicans, bejel, yaws, pinta, and leprosy, can resemble rheumatoid nodules. A rheumatoid nodule, particularly when it occurs on the face, may simulate basal cell carcinoma.

Bone Density

The skeleton has two anatomically and functionally separate components, cortical and trabecular bone, which respond differently to systemic and local diseases and to drugs. RA can be associated with generalized osteopenia and osteoporosis owing to the effects of drugs (especially corticosteroids); cytokine-induced and receptor activator of nuclear factor κB (NFκB) ligand (RANKL)-induced activation of osteoclasts; and the fact that certain groups of patients with the disease, especially postmenopausal women, have other risk factors that enhance the potential for bone loss. Risks for hip fracture and vertebral compression fracture can be high. Bone densitometry should be performed routinely in patients with RA; treatment with bisphosphonates should be considered as an adjunct to therapy.

Because postmenopausal women are at greater risk for osteoporosis, this group should be treated aggressively. Minimizing steroid use is one method that can be used to decrease the risk of osteoporosis in this group and in other
patients with RA. Two-phase loss of bone seems to be induced by glucocorticoids: a rapid first phase, when 12% of bone mass disappears in the first 6 to 12 months of therapy, followed by a subsequent chronic phase, which has a slower rate of bone loss. It is encouraging that axial bone loss in patients with RA induced early by glucocorticoids can be reversed. The evaluation, biology, and management of osteoporosis are discussed in Chapter 101. In the relationship between RA and bone, the focus is, appropriately, on osteoporosis; however, diffuse loss of bone in RA, whether or not it is related to glucocorticoid therapy, leads to a high incidence of stress fractures of long bones in RA. The fibula is the most common fracture site. Acute leg pain in a thin, elderly rheumatoid patient, even without a history of trauma, should generate suspicion of a stress fracture. Geodes (i.e., subchondral cysts developed by synovial penetration of the cortex or subchondral plate and subsequent proliferation) weaken bone and can predispose bone to fracture.

Muscle

Clinical weakness is common in RA, but is it caused by muscle involvement in rheumatoid inflammation, or is it a reflex weakness response to pain? Most rheumatoid patients have muscle weakness, but few have muscle tenderness or elevated muscle enzymes in the blood. In an early autopsy series, focal accumulations of lymphocytes and plasma cells with some contiguous degeneration of muscle fibers were found in all rheumatoid patients—a condition termed nodular myositis. More recent studies have pointed to multiple types of muscle disease in RA, although clinically relevant active myositis is uncommon:

- Diminution of muscle bulk with atrophy of type II fibers
- Peripheral neuropathy, usually due to a mononeuropathy multiplex
- Steroid myopathy
- Active myositis and muscle necrosis with foci of endomyosial mononuclear cell infiltration
- Chronic myopathy resembling a dystrophic process, probably the end stage of inflammatory myositis

In biopsy specimens, atrophy of type II fibers is most common. Evidence of myositis and focal necrosis is found occasionally in biopsy specimens of patients with active disease, particularly in a subset with mild synovitis and a disproportionately high erythrocyte sedimentation rate (ESR). In some patients, the lymphocytes in muscle synthesize IgM RF, emphasizing the systemic nature of RA. The patchy “nodules of myositis” contain plasma cells and lymphocytes.

Skin

The most frequently recognized skin lesion in RA is the rheumatoid nodule but several other manifestations may be observed as well. “Senile” purpura resulting from skin atrophy and capillary fragility is especially common in patients treated with glucocorticoids. Palmar erythema is common, but Raynaud’s syndrome is rare. Manifestations of vasculitis range from occasional nail fold infarcts to a deep, erosive, scarring pyoderma gangrenosum. Palpable purpura in rheumatoid patients often occurs as a reaction to a drug that the patient is taking but can be primary and a direct function of the severity of articular disease. Livedo reticularis, the lacy, dusky purple, asymptomatic discoloration seen on the extremities, is believed to signify a deep dermal vasculopathy. It can be present in any or all diffuse connective tissue diseases and often is associated with antiphospholipid antibodies.

Eye

Virtually all ocular manifestations of RA can be considered complications of the disease (see Chapter 44). Keratoconjunctivitis sicca, a component of Sjögren’s syndrome, is discussed in Chapter 73. Scleritis and episcleritis are associated with RA. Highly differentiated connective tissues in the eye make rheumatoid manifestations particularly interesting and, when they occur in aggressive form, very serious. The episclera of the eye is highly vascular compared with the dense sclera. Scleritis, episcleritis, or both occur in less than 1% of rheumatoid patients. In episcleritis, the eye becomes red and, in contrast to conjunctivitis, causes tearing but no discharge. Loss of vision does not occur as a direct result of episcleritis, but a keratitis or a cataract developing secondarily can cause visual loss. Scleritis causes severe ocular pain and dark red discoloration (Figure 70-12C). No discharge is present. Depending on the intensity of the process, scleritis can be localized and superficial or generalized, with or without granulomatous resolution of the sclera down to the uveal layer; when this complication occurs, it is termed scleromalacia perforans. In contrast to superficial eye disease, which usually can be treated conservatively with topical steroids, scleritis usually requires systemic or intraocular corticosteroid treatment. In some cases, the sclera can become thin even in the absence of overt inflammation, leading to scleromalacia (Figure 70-12D). Rarely, perilimnic ischemic ulcers can be caused by cryoproteins (RF-IgG complexes) and if untreated can result in perforation of the anterior chamber. Patients with RA who have an associated keratoconjunctivitis sicca secondary to Sjögren’s syndrome have pruritic and painful eyes, sometimes leading to chronic blepharitis.

Host Defense and Infection

The incidence of infection as a complication of RA has paralleled the use of glucocorticoids, biologics, and immunosuppressive agents. TNF blockers are especially noteworthy because they have been associated with reactivation of tuberculosis and other opportunistic infections such as histoplasmosis. Pulmonary infections, skin infections, and septic arthritis are the most common infections in RA. Difficulty in diagnosis is accentuated by the similarity of aggressive RA to infection, particularly in joints; a “pseudoseptic” arthritis in rheumatoid patients, associated with fever, chills, and grossly purulent synovial fluid, can be part of a severe exacerbation of RA and must be distinguished from infection. Mortality attributable to respiratory infections such as pneumonia and bronchitis is increased in RA patients compared with the general population. A retrospective longitudinal cohort study compared the frequency of infection in a population-based incidence cohort
of RA patients versus that in a group of individuals without RA from the same population; this study looked at 7900 to 9100 person-years. A total of 609 RA patients and 609 non-RA patients were studied; 73% were women, and mean patient age was 58 years. Hazard ratios for RA patients versus controls after adjustment for age, sex, smoking status, leukopenia, corticosteroid use, and diabetes mellitus were nearly twofold increased for confirmed infection and for infection requiring hospitalization. Bone, joints, skin, respiratory tract, and soft tissues were the organs with highest hazard ratios. In a subsequent study in this cohort, predictors of infection were identified as increasing age, extra-articular manifestations of RA, leukopenia, and comorbidities such as chronic lung disease, alcoholism, diabetes mellitus, and the use of glucocorticoids.

Traditional DMARD use generally is not associated with a major increased incidence of infection. However, vigilance in using biologic agents is essential because mortality is increased from infection due to immunosuppression. Physicians should always have a low threshold of concern for infection in rheumatoid patients.

**Hematologic Abnormalities**

Most patients with active RA have a mild normocytic normochromic anemia that correlates with ESR elevation and the activity of the disease. Anemia has mixed causes in RA. One deficiency may mask evidence of others. A useful guide is that three-quarters of rheumatoid patients with anemia have the anemia of chronic disease, whereas one-quarter respond to iron therapy. Patients in both groups may have superimposed vitamin B₁₂ or folate deficiency. Hemoglobin of less than 10 is rarely due only to RA. The following guidelines may prove helpful to the clinician in diagnosing the cause of anemia in rheumatoid patients:

- Anemia of chronic disease is associated with a significantly higher serum ferritin concentration than is found in isolated iron deficiency.
- Folate or vitamin B₁₂ deficiency or the use of methotrexate can mask iron deficiency, especially in patients taking nonsteroidal anti-inflammatory drugs (NSAIDs) with chronic gastrointestinal blood loss, by increasing the mean cell volume and the mean cell hemoglobin level of erythrocytes.
- The ESR correlates inversely with hemoglobin levels in RA.
- Erythropoietin levels are higher in patients with iron deficiency anemia compared with those with anemia of chronic disease; rheumatoid patients have a diminished response to erythropoietin.

In patients with anemia of chronic disease, the total erythroid heme turnover is slightly reduced, and ineffective erythropoiesis accounts for a much higher than normal percentage of total heme turnover. In contrast to anemia associated with blood loss, ineffective erythropoiesis returns to normal in RA if remission can be induced. Red blood cell aplasia, immunologically mediated, is a rare finding in RA.

Thrombocytosis is often associated with RA. A significant relationship has been noted between thrombocytosis and extra-articular manifestations of rheumatoid disease and disease activity. Eosinophilia (5% of total white blood cell count) has been observed in some patients.

Felty’s syndrome comprises an uncommon but severe subset of seropositive, erosive RA symptoms that occur in patients with neutropenia and splenomegaly. Patients with Felty’s syndrome have more frequent and robust extra-articular RA manifestations and are at increased risk for infection, skin ulceration, and other complications. Another subset of patients with RA includes those with increased numbers of large granular lymphocytes in the peripheral blood, bone marrow, and liver. These lymphocytes contain many azurophilic granules in the cytoplasm and may account for more than 90% of mononuclear cells in blood. They are increased in certain viral infections. The cells are Fc receptor positive, do not produce IL-2, respond poorly to mitogens, and involve antibody-dependent cell-mediated cytotoxicity activity (expressing CD3, CD8, and CD57) or natural killer cells (expressing CD16 and CD56). Among previously described patients with large granular lymphocyte proliferation, almost one-third have had RA. Because the large granular lymphocyte syndrome in patients with RA has the same HLA-DR4 association seen in Felty’s syndrome, the proposal has been made that Felty’s syndrome and large granular lymphocyte syndrome represent different variants of a broader syndrome comprising RA, neutropenia, large granular lymphocyte expansions, HLA-DR4 positivity, and variable splenomegaly. Neutropenia has been described as an adverse event in more recent drug trials in RA patients, most notably with the humanized anti–IL-6 receptor monoclonal antibody tocilizumab. Paraproteinemia, typified by monoclonal gammopathies, has a poor prognostic significance when it appears in rheumatoid patients. This evidence for monoclonal B cell proliferation carries with it a high frequency of malignant transformation to lymphoma or myeloma.

**Vasculitis**

The initial pathologic change in RA is often seen as inflammatory changes in medium and small blood vessels. It is useful to use the term vasculitis to group extra-articular complications related not to proliferative granulomas, but rather to inflammatory vascular disease. Systemic rheumatoid vasculitis, one of the most feared complications of RA, has become increasingly uncommon in recent years. This decline in rheumatoid vasculitis, similar to many other extra-articular manifestations, is likely related to marked improvement in therapy resulting from widespread use of methotrexate and new biologic agents. Male gender, long-standing disease, high-titer RF in serum, hypocomplementemia, erosive disease, circulating cryoglobulins, deposition of immune complexes and complement in blood vessels, and extra-articular features such as subcutaneous nodules are variables associated with the development of rheumatoid vasculitis.

Rheumatoid vasculitis affects a very small subset of patients with established, often severe, RA. It can present as the following:

- Distal arteritis (including from splinter hemorrhage, nail fold infarcts, and gangrene) (Figure 70-13)
- Cutaneous ulceration (including pyoderma gangrenosum)
- Peripheral neuropathy (mononeuritis multiplex or sensory stocking-glove neuropathy)
CHAPTER 70  CLINICAL FEATURES OF RHEUMATOID ARTHRITIS

Other entities in the differential diagnosis for rheumatoid vasculitis include diabetes mellitus, infection, atherosclerosis, and drug reactions. Current practice is to treat organ-specific vasculitis aggressively when it occurs in RA patients, similar to treatment for patients with polyarteritis. This therapeutic approach may be responsible for the small excess mortality in rheumatoid vasculitis patients compared with “controls” with RA alone. In 61 patients with rheumatoid vasculitis, after allowance for general risk factors such as age and sex, the mortality risk was only 1.26 times that of rheumatoid patients without vasculitis.

Renal Disease

The kidney is rarely involved directly in RA but often is compromised indirectly by therapy. Amyloidosis is an unusual complication of chronic RA. AA amyloidosis is one of the most important life-threatening complications of RA. Phenacetin abuse causes renal papillary necrosis; salicylates and other NSAIDs may cause abnormalities as well. Membranous nephropathy is related to therapy with gold salts and penicillamine and was seen when these agents were commonly used to treat RA. Rarely, a focal necrotizing glomerulitis is seen in patients dying with RA and disseminated vasculitis.

Pulmonary Disease

Pleural Disease

Pleuritis is commonly found on autopsy of patients with RA, but clinical disease during life is seen less frequently. In about 20% of patients, pleuritis develops concurrently with the onset of arthritis. Pleuritic pain is not usually a major complaint. Effusions are exudative and can be large enough to cause dyspnea. Characteristics of exudative rheumatoid effusions include a very low glucose concentration in the range of 10 to 50 mg/dL, protein greater than 4 g/dL, mononuclear cells 100 to 3500/mm³, elevated lactate dehydrogenase, low pH, and depressed total complement activity.

- Palpable purpura
- Arteritis of viscera, including heart, lungs, bowel, kidney, liver, spleen, pancreas, lymph nodes, or testis

The pathologic finding in rheumatoid vasculitis is a panarteritis. All layers of the vessel wall are infiltrated with mononuclear cells. Fibrinoid necrosis is seen in active lesions. Intimal proliferation may predispose to thrombosis. Obliterative endarteritis of the finger is a common manifestation of vasculitis, and immune complex deposits have been shown in affected vessels. In patients with hypocomplementemia, the cellular infiltrate around the vessels contains neutrophils. In normocomplementemic patients, lymphocytes predominate.

Neurovascular disease may be the only manifestation of vasculitis. The two common clinical patterns include a mild distal sensory neuropathy and a severe sensorimotor neuropathy (mononeuritis multiplex). The latter form is characterized by severe arterial damage on nerve biopsy specimens. Symptoms of the milder form may consist of paresthesias or “burning feet” in association with decreased touch and pin sensation distally. Patients with mononeuritis multiplex have weakness (e.g., footdrop), in addition to sensory abnormalities. Symptoms and signs are identical to those found in polyarteritis nodosa. Rheumatoid pachymeningitis is a rare complication of RA; confined to the dura and pia mater, this process may be limited to certain areas (e.g., lumbar cord, cisternae). Elevated levels of IgG (including IgM and IgG RFs and low-molecular-weight IgM) and immune complexes are found in the cerebrospinal fluid.

Visceral lesions occur generally as claudication or infarction of the organ supplied by the involved arteries. Intestinal involvement with vasculitis manifests as abdominal pain, at first intermittent and progressing often to continuous pain and a tender, quiet abdomen on examination. If infarction develops, resection must be accomplished promptly. Gangrene of digits and extremities, intestinal lesions with bleeding or perforation, cardiac or renal involvement, and mononeuritis multiplex indicate extensive vasculitis and are associated with a poor prognosis.

Figure 70-13  A, Digital vasculitis in a 65-year-old man with seropositive rheumatoid arthritis. B, Nail fold infarcts can occur in patients with rheumatoid arthritis and typically are associated with rheumatoid factor positivity and active joint disease. (A, Courtesy Eileen Moynihan, MD.)
(CH$_3$). The low glucose concentrations are of importance because infection (particularly tuberculosis) is the only other condition in which such a low pleural fluid glucose level is seen. Impaired transport of glucose into the pleural space seems to be the cause of this.$^6$

**Interstitial Pneumonitis and Fibrosis**

Pulmonary fibrosis can occur with some regularity in RA and is associated with increased mortality. Similar to findings in scleroderma, physical findings consist of fine, diffuse, dry rales. Radiographs show a diffuse reticular (interstitial) or reticulonodular pattern in both lung fields; this can progress to a honeycomb appearance on plain radiographs and a characteristic lattice network on high-resolution CT scans. Pathologic findings include diffuse fibrosis in the midst of a mononuclear cell infiltrate. The principal functional defect is impairment of alveolocapillary gas exchange with decreased diffusion capacity, best measured using single-breath carbon monoxide diffusion capacities.$^9$ Interstitial lung disease and RF and ACPA positivity can occur in smokers before articular symptoms develop.$^8$ This finding has led to the novel hypothesis that RA-specific autoimmunity might be generated in the lung in response to environmental insults such as smoking. Bronchoalveolar lavage may reveal increased numbers of lymphocytes, even in patients with only mildly abnormal chest radiographs and normal pulmonary function test results.$^9$ In more aggressive disease, a higher proportion of neutrophils can be found in bronchoalveolar lavage. Lymphoid interstitial pneumonitis has been described in patients with RA and Sjögren’s syndrome. This relatively indolent disorder is associated with elevated serum globulin levels. Bronchoalveolar lavage shows a primarily lymphocytic response.$^{10}$

**Nodular Lung Disease**

Pulmonary nodules can appear singly or in clusters that coalesce. Single nodules appear as coin lesions and, when significant peripheral arthritis and nodules are present, can be evaluated by needle biopsy. Caplan’s syndrome,$^9$ in which pneumoconiosis and RA are synergistic, producing a violent fibroblastic reaction with obliterative granulomatous fibrosis, has become a rare occurrence as the respiratory environment in mining operations has improved. Nodules may cavitate, creating a bronchopleural fistula. In several cases, solitary pulmonary nodules in RA patients have proved to be a rheumatoid nodule and a coexistent bronchogenic carcinoma.$^{10}$ If the index of suspicion is high for malignancy, the workup should be more aggressive.

**Bronchiolitis**

An uncommon finding is interstitial pneumonitis that progresses to alveolar involvement and bronchiolitis, respiratory insufficiency, and death. Pathologic studies show a cellular loose fibrosis and a proteinaceous exudate in bronchioles and alveoli; interstitial infiltrations of lymphocytes attest to the immunogenic aspects of the disease. The course and prognosis are similar to those of idiopathic bronchiolitis obliterans with organizing pneumonia.

**Pulmonary Hypertension**

Pulmonary hypertension is more common in RA than was previously appreciated. Noninvasive echocardiograms have suggested that mild pulmonary hypertension can be detected in more than 30% of patients with RA.$^9$ Most of these patients are asymptomatic and do not have significant progression.

**Small Airways Disease**

Defined by a reduced maximal midexpiratory flow rate and maximal expiratory flow rate at 50% of functional vital capacity, small airways disease was observed in 50% of 30 RA patients compared with 22% of a control population.$^{94}$ The study was adjusted for pulmonary infection, α1-antitrypsin deficiency, penicillamine treatment, environmental pollution, and smoking. Other investigations have not found small airways dysfunction in RA and have suggested that, if present, it probably is related to factors other than RA.$^{95}$ If real, this phenomenon may be part of a generalized exocrinopathic process in the disease, expressed most commonly in Sjögren’s syndrome.

**Pulmonary Disease due to Treatment of Rheumatoid Arthritis**

Many of the drugs used to treat RA are associated with pulmonary toxicity. For instance, methotrexate and leflunomide can cause pulmonary fibrosis in a pattern that resembles RA. Differentiating rheumatoid pulmonary fibrosis from drug-induced fibrosis can be difficult, and this uncertainty might require discontinuing methotrexate therapy in some cases. An idiosyncratic form of methotrexate pulmonary toxicity is less common and usually is accompanied by fever, lung infiltrates, and eosinophilia. Treatment with TNF inhibitors can lead to reactivation of pulmonary or extrapulmonary tuberculosis.

**Cardiovascular System**

Cardiac disease in RA can take many forms. It has become apparent that increased risk of premature death in RA is due largely to an increased incidence of cardiovascular disease, primarily myocardial infarction and congestive heart failure. Advances in echocardiography have made the diagnosis of pericarditis with endocardial inflammation easier and more specific. Myocardial biopsy through vascular catheters has facilitated diagnosis and classification of myocarditis. In a detailed study of rheumatoid patients using echocardiography, Holter monitors, and electrocardiography, it was reported that 70% of patients with nodular disease and 40% of those with non-nodular RA have some cardiac involvement, including valve thickening or incompetence.$^{95}$

**Atherosclerosis**

Multiple risk factors for coronary artery disease in RA patients are known, in addition to the risk factors that are relevant in the general population. Patients with prolonged RA have a greater extent of atherosclerosis than patients of
the same risk factors noted in RA patients have been implicated in patients without rheumatic diseases, including molecules involved in the immune response, markers of inflammation, and therapeutic agents. It is apparent that, with all else being equal, tobacco smoking is an important factor in augmenting early atherosclerosis in RA patients. In a large, well-studied population of rheumatoid patients at the Mayo Clinic, patients were followed until death, migration from Olmsted County, or 2001. Data show that congestive heart failure was more important than ischemic heart disease as a cause of death. Even in RA patients without clinically evident cardiovascular disease, left ventricular diastolic function and right ventricular diastolic function are reduced.

**Pericarditis**

Infrequently diagnosed on the basis of history and physical examination in RA, pericarditis is present in 50% of patients at autopsy. In one study, 31% of patients with RA had echocardiographic evidence of pericardial effusion. The same study revealed only rare evidence of impaired left ventricular function in prospectively studied outpatients with RA. Although unusual, cardiac tamponade and/or the more common constrictive pericarditis can develop in RA and may require pericardiectomy. Almost all patients have a positive test for RF; and half have nodules. Preservation of good ventricular function on echocardiography in the face of deteriorating clinical myocardial function should raise a high index of suspicion of constrictive pericarditis.

**Myocarditis**

Myocarditis is rare but may be due to granulomatous disease or interstitial myocarditis. The granulomatous process resembles subcutaneous nodules and could be considered specific for the disease. Diffuse infiltration of the myocardium by mononuclear cells may involve the entire myocardium and yet may have no clinical manifestations, but it could be suggested by echocardiography.

**Endocardial Inflammation**

Echocardiographic studies have reported evidence of previously unrecognized mitral valve disease of the anterior leaflet of the mitral valve. Although aortic valve disease and arthritis are more commonly associated with ankylosing spondylitis, numerous patients with granulomatous nodules on the valve have been reported.

**Conduction Defects**

Atrioventricular block is unusual in RA but is probably related to direct granulomatous involvement. Pathologic examination may reveal proliferative lesions or healed scars. Complete heart block has been described in more than 30 patients with RA. It generally occurs in patients with established erosive nodular disease. Complete heart block usually is permanent and is caused by rheumatoid granulomas in or near the atrioventricular node or bundle of His. Rarely, amyloidosis is responsible for heart block.

**Granulomatous Aortitis or Valvular Disease**

In severe rheumatoid heart disease, granulomatous disease can spread to involve even the base of the aorta. Occasionally, granulomatous disease associated with RA necessitates urgent valve replacement for aortic regurgitation. With improved therapy for synovitis, these complications (similar to many others) have become rare.

**DIAGNOSIS**

Criteria to establish the diagnosis of RA are based on an effective clinical history and physical examination, laboratory tests, and exclusion of other diagnoses. No single feature or laboratory test is sufficient for a definitive diagnosis. The 1987 American College of Rheumatology (ACR) criteria for classification usually are not used for diagnosis in individual cases; however, the requirement that objective evidence of synovitis must be present for at least 6 weeks is important, especially because many transient forms of synovitis are observed in primary care settings (Table 70-3). A physician would be less likely to make a premature diagnosis of RA in a patient who might have a self-limited synovitis. To attempt preventing irreversible damage to joints, the diagnosis of RA should be confirmed or ruled out within 2 months after the onset of synovitis.

Classification criteria were revised in 2010 by the ACR and the European League Against Rheumatism (EULAR). The new criteria place greater emphasis on serology and early diagnosis in patients with very few or even no swollen or tender joints (Table 70-4). Symmetric joint disease is not a feature of the new criteria. In addition, classic radiographic changes such as marginal erosions automatically qualify patients for RA classification if they have a single swollen joint. Imaging evidence of synovitis, including ultrasound or MRI, can be used to classify patients even in the absence of symptoms if patients have high titers of RF or ACPA and elevated acute phase reactants. The long-term usefulness and the specificity of the classification of the revised criteria are uncertain, but their use will likely result in earlier diagnosis and treatment. It is important to recognize that these are not diagnostic criteria, and they are used mainly in classifying patients for clinical research.

The characteristic patient with RA reports pain and stiffness in multiple joints, with prominent and prolonged morning stiffness. Joint swelling is boggy and includes soft tissue and synovial fluid. Joints are tender, especially the small joints of the hands and feet, but usually are not painful when the patient is at rest. Palmar erythema and prominent veins on the dorsum of the hand and wrist indicate increased blood flow. DIP joints are rarely involved. The temperature over the involved joints (except the hip) can be elevated, but the joints usually are not red. The range of motion is limited, and muscle strength and function around inflamed joints are diminished. Soft, poorly delineated subcutaneous nodules are often found in the olecranon bursa. Findings on general physical examination are normal except for a possible low-grade fever in a few patients. Soft, small lymph nodes are found occasionally in epitrochlear, axillary, and cervical areas. The history and physical examination are the
Table 70-3  1987 Revised American Rheumatism Association Criteria for Classification of Rheumatoid Arthritis

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning stiffness</td>
<td>Morning stiffness in and around the joints lasting at least 1 hour before maximal improvement</td>
</tr>
<tr>
<td>Arthritis of ≥3 joint areas</td>
<td>At least 3 joint areas simultaneously having soft tissue swelling or fluid (not bony overgrowth alone) observed by a physician (the 14 possible joint areas are [right or left] PIP, MCP, wrist, elbow, knee, ankle, and MTP joints)</td>
</tr>
<tr>
<td>Arthritis of hand joints</td>
<td>Simultaneous involvement of the same joint areas (as in criterion 2) on both sides of the body (bilateral involvement of PIP, MCP, or MTP joints is acceptable without absolute symmetry)</td>
</tr>
<tr>
<td>Symmetric arthritis</td>
<td>Subcutaneous nodules over bony prominences or extensor surfaces, or in juxta-articular regions, as observed by a physician</td>
</tr>
<tr>
<td>Rheumatoid nodules</td>
<td>Changes typical of RA on posteroanterior hand and wrist radiographs, which must include erosions or unequivocal bony decalcification localized to or most marked adjacent to involved joints (osteoarthritis changes alone do not qualify)</td>
</tr>
</tbody>
</table>

**American College of Rheumatology Criteria**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning stiffness</td>
<td>68</td>
<td>65</td>
</tr>
<tr>
<td>Arthritis in ≥3 areas</td>
<td>80</td>
<td>43</td>
</tr>
<tr>
<td>Arthritis of the hand joints</td>
<td>81</td>
<td>46</td>
</tr>
<tr>
<td>Symmetric arthritis</td>
<td>77</td>
<td>37</td>
</tr>
<tr>
<td>Rheumatoid nodules</td>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td>Rheumatoid factor</td>
<td>59</td>
<td>93</td>
</tr>
<tr>
<td>Radiographic change</td>
<td>22</td>
<td>98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical or Laboratory Variable</th>
<th>Persistent Nonerosive Versus Self-Limiting</th>
<th>Persistent Erosive Versus Persistent Nonerosive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds Ratio</td>
<td>Score</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>Symptom duration at first visit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;6 weeks, &lt;6 months</td>
<td>2.49</td>
<td>2</td>
</tr>
<tr>
<td>&gt;6 months</td>
<td>5.49</td>
<td>3</td>
</tr>
<tr>
<td>Morning stiffness &gt;1 hour</td>
<td>1.96</td>
<td>1</td>
</tr>
<tr>
<td>Arthritis in ≥3 joints</td>
<td>1.73</td>
<td>1</td>
</tr>
<tr>
<td>Bilateral MTP compression pain</td>
<td>1.65</td>
<td>1</td>
</tr>
<tr>
<td>Rheumatoid factor positivity</td>
<td>2.99</td>
<td>2</td>
</tr>
<tr>
<td>Anticitrullinated protein antibody positivity</td>
<td>4.58</td>
<td>3</td>
</tr>
<tr>
<td>Radiographic erosions (hands or feet)</td>
<td>2.75</td>
<td>2</td>
</tr>
</tbody>
</table>

*For classification purposes, a patient is said to have RA if he or she has satisfied at least four of the seven criteria. Criteria 1 through 4 must be present for at least 6 weeks. Patients with two clinical diagnoses are not excluded. Designation as classic, definite, or probable RA is not to be made.

MCP, metacarpophalangeal; MTP, metatarsophalangeal; PIP, proximal interphalangeal; RA, rheumatoid arthritis.

Most sensitive and specific tools for diagnosis of RA. Initial laboratory tests often show the results in the following list (essential tests are indicated with an asterisk [*]):

- Normal white blood cell count and differential*
- Thrombocytosis*
- Mild anemia, normochromic and either normocytic or microcytic*
- Normal urinalysis*
- ESR 30 mm/hr or greater and C-reactive protein level greater than 0.7 pg/mL*
- Normal renal, hepatic, and metabolic tests*
- Normal serum uric acid level
- Positive RF test (about 70% to 80% of patients; occurs in many normal people, in patients with other rheumatic diseases, and in those with chronic infection)*
- ACPA (about 80% to 90% of patients; can be seen in other diseases, including active tuberculosis) (especially useful as diagnostic and prognostic indicator of early synovitis)*
- Other autoantibodies (commonly found but with limited differential diagnosis utility, including antinuclear antibody, SS-A, and SS-B). Although they usually are not clinically indicated, tests for anti-double-stranded DNA and antineutrophil cytoplasmic antibody are usually negative.
- Polyclonal gammapathy as determined by serum protein electrophoresis
- Normal or elevated serum complement level
- “Typical” arthrocentesis, when obvious fluid is present, in RA reveals the following:
  - Joint fluid is straw-colored, is slightly cloudy, and contains many flecks of fibrin and 5000 to 25,000 white blood cells/mm³; at least 50% of these are polymorphonuclear leukocytes.
  - No crystals
  - Complement C4 and C2 levels are depressed, but C3 level can be normal.
  - Normal synovial fluid glucose level
  - Negative cultures and Gram stain
Table 70-4  2010 ACR/EULAR Classification Criteria for Rheumatoid Arthritis (Score-Based Algorithm for Classification in an Eligible Patient [Cutpoint for RA: ≥6/10])

<table>
<thead>
<tr>
<th>Joint Involvement*</th>
<th>(0-5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 medium to large(^1) joint</td>
<td>0</td>
</tr>
<tr>
<td>2-10 medium to large joints</td>
<td>1</td>
</tr>
<tr>
<td>1-3 small(^1) joints (with or without involvement of large joints)</td>
<td>2</td>
</tr>
<tr>
<td>4-10 small joints (with or without involvement of large joints)</td>
<td>3</td>
</tr>
<tr>
<td>&gt;10 joints(^4) (at least one small joint)</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Serology*</th>
<th>(0-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative RF AND negative ACPA</td>
<td>0</td>
</tr>
<tr>
<td>Low-positive RF OR low-positive ACPA</td>
<td>2</td>
</tr>
<tr>
<td>High-positive RF OR high-positive ACPA</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Acute Phase Reactants(^6)*</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal CRP AND normal ESR</td>
<td>0</td>
</tr>
<tr>
<td>Abnormal CRP OR abnormal ESR</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Duration of Symptoms**</th>
<th>(0-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;6 weeks</td>
<td>0</td>
</tr>
<tr>
<td>26 weeks</td>
<td>1</td>
</tr>
</tbody>
</table>

\(^*\)Joint involvement refers to any swollen or tender joint on examination, or evidence of synovitis on magnetic resonance imaging or ultrasonography.

\(^1\)Distal interphalangeal joints, first carpometacarpal joint, and first metatarsophalangeal joint are excluded from assessment. Categories of joint distribution are classified according to the locations and numbers of involved joints, with placement into the highest category possible based on the pattern of joint involvement.

\(^4\)Medium to large joints refer to shoulders, elbows, hips, knees, and ankles.

\(^3\)Small joints refer to the metacarpophalangeal joints, proximal interphalangeal joints, metatarsophalangeal joints 2 through 5, thumb interphalangeal joints, and wrists.

\(^\dagger\)In this category, at least one of the involved joints must be a small joint; the other joints can include any combination of large and additional small joints, as well as other joints not specifically listed elsewhere (e.g., temporomandibular, acromioclavicular, sternoclavicular).

\(^\ddagger\)Individuals should be scored by these criteria only if at least one serologic test and at least one acute phase reactant test result are available. Where a value for a serologic test of acute phase reactant is not available, that test should be considered as negative/normal.

\(^$\)Negative refers to international unit (IU) values that are less than or equal to the upper limit of normal (ULN) for the laboratory and assay. Low-positive refers to IU values that are greater than the ULN but less than or equal to 3 × ULN for the laboratory. High-positive refers to IU values that are greater than 3 ULN for the laboratory.

\(^\circ\)Duration of symptoms refers to patient self-report of the duration or signs of symptoms (e.g., pain, swelling, tenderness) of joints that are clinically involved at the time of assessment.

\(^\text{ACPA, anticitrullinated protein/peptide antibody; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; RF, rheumatoid factor; ULN, upper limit of normal.}\


Differential Diagnosis

Other diseases must be excluded before the diagnosis of RA is established.\(^{106}\) The differential diagnosis of adult polyarthritis is outlined in Table 70-5.\(^{107}\) The following sections lists various diseases; the relative frequency of each illness is specified as common, uncommon, or rare.

### Table 70-5  Discriminating Features in Patients Presenting with Polyarthritis and Fever

<table>
<thead>
<tr>
<th>Symptom or Sign</th>
<th>Possible Diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature &gt;40° C</td>
<td>Still’s disease, bacterial arthritis, SLE.</td>
</tr>
<tr>
<td>Fever preceding arthritis</td>
<td>Viral arthritis, Lyme disease, reactive arthritis, Still’s disease, bacterial endocarditis.</td>
</tr>
<tr>
<td>Effusion disproportionately greater than pain</td>
<td>Tuberculous arthritis, bacterial endocarditis, inflammatory bowel disease, giant cell arteritis, Lyme disease.</td>
</tr>
<tr>
<td>Pain disproportionately greater than effusion</td>
<td>Rheumatic fever, familial Mediterranean fever, acute leukemia, acquired immunodeficiency syndrome.</td>
</tr>
<tr>
<td>Positive test for rheumatoid factor</td>
<td>Rheumatoid arthritis, viral arthritis, tuberculous arthritis, bacterial endocarditis, SLE.</td>
</tr>
<tr>
<td>Symmetric small joint synovitis</td>
<td>Rheumatoid arthritis, SLE.</td>
</tr>
<tr>
<td>Leukocytosis (&gt;15,000/mm(^3))</td>
<td>Viral arthritis, bacterial arthritis, bacterial endocarditis, Still’s disease, systemic vasculitis.</td>
</tr>
<tr>
<td>Leukopenia</td>
<td>Acute leukemia</td>
</tr>
<tr>
<td>Episodic recurrences</td>
<td>Lyme disease, crystal-induced arthritis, inflammatory bowel disease, Whipple’s disease, Mediterranean fever, Still’s disease, SLE.</td>
</tr>
</tbody>
</table>

Spondyloarthropathies (Common)

Ankylosing spondylitis, psoriatic arthritis, inflammatory bowel disease-associated arthritis, and reactive arthritis are often referred to as spondyloarthropathies. These diseases generally are marked by their respective nonarticular features and the following pattern of joint disease: asymmetric, oligoarticular, lower extremities more than upper extremities, and large joints more than small joints (note that there are...
many exceptions to these general guidelines). The problem in differentiating these diseases from RA arises with a patient (particularly a woman) who has minimal back pain and definite peripheral joint involvement. The presence of low back pain and lumbar involvement helps distinguish these diseases from RA.

In distinguishing patients with reactive arthritis from those with RA, a careful search for heel pain or tenderness and ocular or urethral symptoms is of great importance. Polyarthritis persists chronically in more than 80% of patients with reactive arthritis. The characteristics of enthesopathy in patients with reactive arthritis (i.e., “sausage” digits indicating periarticular soft tissue inflammation, insertion tendinitis, periostitis, and peri-insertional osteoporosis or erosion) may point to the diagnosis.

The differential diagnosis between RA with psoriasis and some forms of psoriatic arthritis may be artificial (see Chapter 43). Some patients with DIP joint involvement and severe skin involvement have a disease that is not RA. Others have a seroserositis symmetric polyarthritis that appears to be RA, yet they also have psoriasis. These patients can be treated with the same disease-modifying drugs that are given to patients with progressive RA, including TNF inhibitors.

As a syndrome described extensively in the French literature, synovitis, acne, pustulosis, hyperostosis, and calcium pyrophosphate deposition disease (SAPHO) may resemble psoriatic arthritis and, occasionally, when peripheral arthritis is present, RA. As the name implies, patients variably have severe acne, palmar and plantar pustules, hyperostotic reactions (particularly in the clavicles and sternum), sacroilitis, and peripheral inflammatory arthritis.

Inflammatory bowel disease (ulcerative colitis and Crohn’s disease) is associated with arthritis in 20% of cases (see Chapter 78). Peripheral arthritis occurs more commonly than spondylitis in many series. Ankle, knees, and elbows are the most typically involved peripheral joints, with PIP joints and wrists next in frequency. Simultaneous attacks of arthritis and the development of erythema nodosum are common. Only two or three joints are affected at once. Involvement is usually asymmetric, and erosions are uncommon. The occurrence of peripheral arthritis in inflammatory bowel disease is not related to HLA-B27.

Behçet’s syndrome is marked by asymmetric polyarthritis in 50% to 60% of cases (see Chapter 93). It is rare, with a prevalence of less than 1 in 25,000 in the United States. In more than half of cases, the attacks of arthritis are monoarticular. Knees, ankles, and wrists are affected most often; synovial fluid usually contains more than 5000 but less than 30,000 white blood cells/mm³. Joint deformity is unusual. Painful oral and genital ulcers and central nervous system involvement are characteristic. Uveal tract involvement in Behçet’s syndrome must be differentiated from the scleritis characteristic of RA in patients with ocular and joint disease.

Enteric infections are complicated occasionally by inflammatory joint disease resembling RA. The joint disease associated with Yersinia enterocolitica infection occurs several weeks after the gastrointestinal illness. Knees and ankles are the joints most commonly involved, and most patients (even patients with peripheral arthritis and no spondylitis) have HLA-B27. Reactive arthritis also has been reported after Salmonella, Shigella, and Campylobacter (Helicobacter) jejuni infection.

Calcium Pyrophosphate Dihydrate Deposition Disease (Common)

Calcium pyrophosphate dihydrate deposition disease is a crystal-induced synovitis that takes many forms, ranging from a syndrome of indolent osteoarthrosis to that of an acute, hot joint. About 5% of patients have chronic polyarthritis (sometimes referred to as pseudo-RA) associated with proliferative erosion of subchondral bone. Although radiographs are helpful when chondrocalcinosis is present, calcium pyrophosphate dihydrate deposition disease may be present in the absence of calcification on radiographs. Diagnosis then can be made only by arthrocentesis. A radiographic sign of calcium pyrophosphate dihydrate deposition disease that helps to differentiate it from RA is the presence of unicompartamental disease in the wrists (see Chapter 96). On physical examination, the MCPs in calcium pyrophosphate dihydrate deposition disease generally have bony enlargement rather than soft tissue swelling owing to synovial hyperplasia.

Fibromyalgia (Common)

In fibromyalgia, no evidence of synovitis is found (see Chapter 52). Although no specific diagnostic tests define this entity, nonarticular pain and tenderness are common. In an analysis contrasting the pain properties of fibromyalgia versus those of RA, fibromyalgia patients used diverse adjectives to describe their pain, the most common being “pressing,” “shooting,” “gnawing,” “cramping,” and “crushing.” Most patients in both groups defined the pain as aching and exhausting. Some patients with diffuse connective tissue diseases, including RA, can develop a superimposed fibromyalgia, adding to the difficulty of treating the arthritis.

Gout (Common)

Before a diagnosis of chronic erosive RA is made, chronic tophaceous gout must be ruled out. The reverse applies as well. Features of gouty arthritis that can mimic the features of RA include synovitis, symmetric involvement, fusiform swelling of joints, subcutaneous nodules, and a subacute presentation of attacks. Conversely, certain aspects of RA that suggest gouty arthritis include periarticular nodules and seronegative disease (particularly in men). Radiographic findings may be similar, with the appearance of subcortical erosions of RA resembling small osseous tophi in gout. Although large asymmetric erosions with ballooning of the cortex and overhanging edges are more likely to be caused by gout than by RA, this is not always the case. Serologic test results may be misleading as well; RF has been reported in 30% of patients with chronic tophaceous gout who have no clinical or radiographic signs of RA.
**Human Immunodeficiency Virus Infection (Common)**

Several types of arthropathy have been described in association with human immunodeficiency virus (HIV) infection, including the following:

- Brief, acute arthralgias concurrent with initial HIV viremia
- HIV-associated arthritis, lower extremity noninflammatory oligoarthritis, or a persistent polyarthritides
- Seronegative spondyloarthropathy resembling psoriatic arthritis, or reactive arthritis, often more severe than in patients without HIV infection

It is crucial to rule out HIV in any patient with an acute polyarthritis and fever: HIV-positive patients are at greater risk for toxicity or opportunistic infection when using immunosuppressive drugs (see Chapter 114). HIV-positive patients also can present with syndromes of vasculitis.

**Infectious Diseases (Including Viral Causes Such As Hepatitis C) (Common)**

Arthritis is commonly associated with viral infection and sometimes can mimic RA. For example, symmetric inflammatory arthritis complicates rubella or rubella vaccination more often in adults than in children and may affect the small joints of the hands. Lymphocytes predominate in synovial effusions of rubella.

Rheumatoid-like arthritis often precedes jaundice in viral hepatitis and is associated with the presence of circulating hepatitis B surface antigen and hypocomplementemia. The surface antigen has been found in synovial tissues with the use of direct immunofluorescence; this supports the concept that this synovitis is mediated by immune complexes. Acute onset of diffuse polyarthritis with small joint effusions and minimal synovial swelling, often accompanied by urticaria, should prompt the physician to request liver function tests in the patient with a history of exposure to hepatitis. With the onset of icterus, arthritis usually resolves without a trace.

Increasing recognition of hepatitis C as a cause of joint problems is related to the availability of specific serologic tests for this virus. About one-third of people infected with hepatitis C virus have arthralgias or arthritis, and in a Korean series, the prevalence of cryoglobulins (mean concentration of 9.8 g/L) was 59%. Patients can present with palmar tenosynovitis, small joint synovitis, carpal tunnel syndrome, and positive tests for RF. The presence of ACPAs can be a useful feature in distinguishing RA from hepatitis C–associated arthritis. Other findings, including mixed cryoglobulinemia syndrome, glomerulonephritis, and cutaneous vasculitis, round out the clinical spectrum of rheumatic complaints associated with this viral infection. Because exacerbation of hepatitis can be associated with the use and the cessation of methotrexate therapy, a good case has been made for testing for hepatitis C every patient with RA scheduled to be started on therapy with this drug.

Fever, sore throat, and cervical adenopathy followed by symmetric polyarthritis are compatible with infection resulting from hepatitis B, rubella, adenovirus type 7, echovirus type 9, *Mycoplasma pneumoniae*, or Epstein-Barr virus and acute rheumatic fever or adult-onset Still’s disease. In Japan, many more patients with RA have circulating antibodies against human T-lymphotropic virus type I (HTLV-I), and epidemiologic evidence indicates that HTLV might be associated with an RA-like condition. Multiple nodules within tendon sheaths associated with inflammation resembling rheumatoid tenosynovitis have been described in a patient with HTLV-I arthropathy.

A chronic polyarthritis resembling RA has been described after serologic proof of parvovirus infection. Usually the process is self-limited and does not progress to a destructive synovitis (see Chapter 114). Adults, often those involved in child care, present with a history of a viral-type illness, sometimes with desquamating finger involvement and a diffuse, red facial rash (“slapped cheeks”), followed by arthralgias and synovitis.

Chikungunya is an insect-borne virus found in Africa and Asia that manifests infection as an acute febrile phase with rash followed by prolonged arthralgia or arthritis of multiple joints. Arthritis usually persists for weeks to months but in some cases can last for years. Polyarthritis after bacterial infection such as poststreptococcal arthritis also can resemble RA. Typically, patients have an antecedent skin or oropharynx group A streptococcal infection in the weeks preceding symptoms. Antistreptolysin O antibody titers are usually elevated. Although the same bacteria can cause glomerulonephritis, it is uncommon to see concomitant arthritis and renal disease.

**Lyme Disease (Common in Endemic Areas)**

Lyme disease can closely simulate RA in adults or children because of its intermittent course with the development of chronic synovitis. A proliferative, erosive synovitis necessitating synovectomy has evolved in several cases. The histopathologic appearance of the proliferative synovium is not different from that of RA (see Chapter 110), and the Lyme synovial cells produce a similar excess of metalloproteinases. Lyme serologic tests can help distinguish this disease from RA, as can a history of tick bites, characteristic skin rash, or neurologic involvement.

**Osteoarthritis (Common)**

Although osteoarthritis can begin as a degenerative cartilage disease, and RA begins as synovial inflammation, these diseases can overlap as they progress (Table 70-6). In osteoarthritis, as cartilage deteriorates and joint congruence is altered and stressed, a reactive synovitis often develops. Conversely, because the rheumatoid pannus erodes cartilage, secondary osteoarthritic changes in bone and cartilage occur. In osteoarthritis, involvement of the DIP and first carpometacarpal (CMC) joints is common, but it is rare in RA. Alternatively, MCP and wrist involvement are uncommon in OA. During end stages of degenerative joint disease and RA, involved joints appear the same. In some cases, to differentiate clearly between the two, the physician must delve into the early history and functional abnormalities associated with the disease.

Erosive osteoarthritis occurs frequently in middle-aged women (more frequently than in men) and is characterized
bony osteophytes, minimal soft tissue swelling early.

- Pain increases through the day and with use.
- Subchondral sclerosis, osteophytes.

Congenital abnormalities (e.g., shallow acetabulum)

- Increases with age.
- Pain increases through the day and with use.
- Distal interphalangeal joints (Heberden's nodes),
  weight-bearing joints (hips, knees).

Soft tissue swelling, warmth.

- Bony osteophytes, minimal soft tissue swelling early.
- Subchondral sclerosis, osteophytes.
- Normal.

Table 70-6  Factors Useful for Differentiating Early Rheumatoid Arthritis from Osteoarthritis

<table>
<thead>
<tr>
<th></th>
<th>Rheumatoid Arthritis</th>
<th>Osteoarthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at onset</strong></td>
<td>Childhood and adults, peak incidence in 50s</td>
<td>Increases with age</td>
</tr>
<tr>
<td><strong>Predisposing factors</strong></td>
<td>Susceptibility epitopes (HLA-DR4, HLA-DR1)</td>
<td>Trauma</td>
</tr>
<tr>
<td></td>
<td>PTPN22, PADI4 polymorphisms, and others</td>
<td>Congenital abnormalities (e.g., shallow acetabulum)</td>
</tr>
<tr>
<td></td>
<td>Smoking</td>
<td></td>
</tr>
<tr>
<td><strong>Early symptoms</strong></td>
<td>Morning stiffness</td>
<td></td>
</tr>
<tr>
<td><strong>Joints involved</strong></td>
<td>Metacarpophalangeal joints, wrists, proximal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>interphalangeal joints most often; distal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>interphalangeal joints almost never</td>
<td></td>
</tr>
<tr>
<td><strong>Physical findings</strong></td>
<td>Soft tissue swelling, warmth</td>
<td></td>
</tr>
<tr>
<td><strong>Radiologic findings</strong></td>
<td>Periarticular osteopenia, marginal erosions</td>
<td></td>
</tr>
<tr>
<td><strong>Laboratory findings</strong></td>
<td>Increased C-reactive protein, rheumatoid factor,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>anticitrullinated protein antibody, anemia, leukocytosis</td>
<td></td>
</tr>
</tbody>
</table>

by inflammatory changes in PIP joints with destruction and functional ankylosis of the joints. The PIP joints can be red and hot, yet almost no synovial proliferation or effusion occurs; joint swelling involves hard, bony tissue, not synovium. The ESR may be slightly elevated, but RF is not found (see Chapter 99).

**Polymyalgia Rheumatica and Giant Cell Arteritis (Common)**

Joint radionuclide imaging studies demonstrate increased vascular flow in the synovium of patients with classic polymyalgia rheumatica. Clinical synovitis commonly occurs, but markedly symptomatic arthritis is uncommon. A careful history usually can differentiate shoulder or hip-girdle muscle pain from shoulder or hip joint pain. Examination of synovial biopsy specimens from patients with polymyalgia rheumatica indicates that the synovitis is usually milder than that found in RA.

Several patients have been described whose initial symptom of giant cell arteritis was a peripheral polyarthritis clinically indistinguishable from RA. Among 19 such patients in a group of 522 with biopsy-proven giant cell arteritis, however, only 3 were RF positive. The interval between onset of each set of symptoms was 3 years or less in 15 of the 19 patients; this suggests a relationship between the two (see Chapter 88). It is known that patients with giant cell arteritis often have HLA-DR4 alleles.

**Systemic Lupus Erythematosus (Common)**

The distribution of involved joints and deformities in systemic lupus erythematosus (SLE) can be identical to that seen in RA. In contrast to RA, SLE arthritis usually does not cause cartilage destruction or bone erosion. The deformities are often reducible, sometimes leading to normal radiographs, owing to the effect of placing the hand firmly on the film cassette. Serologies (antinuclear antibody, anti–double-stranded DNA) and major organ system involvement usually can distinguish RA from SLE. However, RA and SLE can overlap, suggesting a mixed picture of both diseases, sometimes referred to as “rups.” The presence of erosive arthritis in SLE patients is associated with ACPAs.

The frequency of erosive arthritis in SLE patients is approximately 10% to 15%, and the subset of SLE patients that are anti-CCP antibody positive are more likely to have erosive arthritis, suggesting a pathogenic role for the antibody in the formation of erosions.

**Musculoskeletal Pain of Thyroid Disease (Common)**

In hypothyroidism (see Chapter 121), synovial effusions and synovial thickening can simulate RA. The ESR may be elevated because of hypergammaglobulinemia, but C-reactive protein is normal. The joint fluid is noninflammatory and may have increased viscosity. Knees, wrists, hands, and feet are involved most often, and coexisting calcium pyrophosphate dihydrate deposition disease is frequently found. This syndrome should be distinguished from arthralgias and other nonspecific musculoskeletal problems that often accompany hyperthyroidism and hypothyroidism.

The syndrome of thyroid acropachy complicates less than 1% of cases of hyperthyroidism. This syndrome comprises periosteal new bone formation, which may be associated with a low-grade synovitis similar to hypertrophic osteoarthropathy. Patients with coexisting RA and hyperthyroidism have pain from the arthritis that, although impossible to quantify, seems to exceed the pain expected from the degree of inflammation.

**Vasculitis (Uncommon)**

Patients with a variety of vasculitides can present with inflammatory arthritis; these syndromes are readily distinguished from RA. Many small vessel vasculitides show palpable purpura and are associated with hepatitis C and cryoglobulinemia. Medium vessel forms of vasculitis, such as granulomatosis with polyangiitis, Churg-Strauss syndrome, or microscopic polyangiitis, include major organ system involvement (e.g., reactive airways disease, glomerulonephritis) and are usually antineutrophil cytoplasmic antibody positive. Polyarteritis nodosa, especially in arthritis associated with hepatitis B, usually is distinguished by renovascular hypertension and other systemic symptoms.

**Adult-Onset Still’s Disease (Uncommon)**

Significant fever at the onset of RA is unusual in adults. Later in the course, if vasculitis or serositis is present, or if intense exacerbations of disease occur, low-grade fever is more common. Adult Still’s disease, in contrast, usually
manifests with high spiking fevers. Serologic studies (RF and antinuclear antibody) are negative, and patients do not have subcutaneous nodules. Fever patterns in these patients are often quotidian (i.e., reaching normal levels at least once each day). Occasionally, evanescent salmon-colored or pink macules on the trunk and extremities become more prominent when patients are febrile. The cervical spine may be involved, and loss of neck motion may be striking. Approximately 20% of patients with adult-onset Still's disease showed significant functional deterioration from erosive joint disease, similar to RA.124

Abnormal liver function tests consistent with hepatitis and severe abdominal pain can be performed and may confirm findings at diagnosis. Liver involvement is observed in most cases and was noted in more than two-thirds of patients in one series;125 hypergammaglobulinemia is present in more than 60%. Pericarditis and pleural effusions are observed in less than 25% of cases. In contrast to active SLE with nephritis, the serum complement level is normal or high. Serum ferritin levels can be enormously elevated to well beyond levels expected compared with other acute phase reactants in the same individual.126 When levels are greater than 10,000 ng/mL, physicians should strongly consider adult Still's disease as the diagnosis rather than RA. The glycosylated form of serum ferritin, usually greater than 50% of the total, is reportedly low (mean, 16%) during active phases and in remission.127

Yamaguchi and associates128 developed criteria for establishing the diagnosis of adult Still's disease, which, in numerous series, have greater than 90% sensitivity (Table 70-7). After other diseases have been excluded, adult Still's disease should be considered if five criteria (with more than two being major ones) are met. It is unknown yet whether adding hyperferritinemia would increase the specificity of the diagnosis.

**Bacterial Endocarditis (Uncommon)**

Arthralgias, arthritis, back pain, and myalgias occur in approximately 30% of patients with subacute bacterial endocarditis.129 Symptoms typically occur in one or several joints, usually large, proximal ones. This synovitis probably is caused by the deposition of circulating immune complexes. Confusion with RA can occasionally arise because more than half of patients with endocarditis are seropositive for RF. Fever out of proportion to joint findings in the setting of leukocytosis should lead to consideration of infective endocarditis as a diagnostic possibility, even in the absence of a significant heart murmur. Peripheral emboli with digital infarctions may be found, simulating palpable purpura when they occur on the lower legs. Blood cultures should be obtained in all patients with polyarthritis and significant fever. Embolic phenomena with constitutional symptoms, including arthralgias, can be presenting symptoms of atrial myxoma, but this process usually mimics systemic vasculitis or subacute bacterial endocarditis more than RA.

**Hemochromatosis (Uncommon)**

The characteristic articular feature of hemochromatosis that is almost diagnostic is firm bony enlargement of the MCP joints, particularly the second and third joints, with associated cystic degenerative disease and large hook-like osteophytes on radiographs and, frequently, chondrocalcinosis. Marginal erosions, juxta-articular osteoporosis, synovial proliferation, and ulnar deviation are not seen in the arthropathy of hemochromatosis but are common in RA. Wrists, shoulders, elbows, hips, and knees are involved less often than the MCP joints. Arthritis leads the list of diagnoses provided to patients to explain their symptoms before the diagnosis of hemochromatosis is decided.131 In the series by McDonnell and colleagues,131 patients with symptoms received a diagnosis of hemochromatosis only after the symptoms had been present, on average, for an extended period (10 years) and after they had visited an average of 3.5 physicians (see Chapter 118). Other distinguishing features include the fact that hemochromatosis is more common in males, and that patients generally will be RF and ACPA negative and may have other features of iron overload, including CHF, liver abnormalities, adult-onset diabetes, and hyperpigmentation.

**Hemophilic Arthropathy (Uncommon)**

A deficiency of factor VIII or, less frequently, factor IX, sufficient to produce clinical bleeding, frequently results in hemarthroses. Iron overload at the joint generates a proliferative synovitis that often leads to joint destruction. Because iron stimulates metalloproteinase production by synovial cells, when feasible, large hemarthroses should be aspirated, and the joint should be immobilized and wrapped well. The clotting abnormality is rarely overlooked, however, and it is unlikely that a diagnosis of RA would be made in the setting of hemophilia A or B (see Chapter 119).

**Hyperlipoproteinemia (Uncommon)**

Achilles tendinitis and tenosynovitis can be presenting symptoms in familial type II hyperlipoproteinemia and may be accompanied by arthritis. Synovial fluid findings may resemble the findings of mild RA, and tendon xanthomas may be mistaken for rheumatoid nodules or gouty tophi. Conversely, bilateral pseudoxanthomatous rheumatoid nodules have been described. Treatment of hyperlipoproteinemia with statins may cause an acute or subacute musculoskeletal syndrome that resembles myositis or polymyalgia rheumatica more than it resembles RA (see Chapter 85).

**Hypertrophic Osteoarthropathy (Uncommon)**

Hypertrophic osteoarthropathy may present as oligoarthritis involving the knees, ankles, or wrists. However, patients

---

**Table 70-7 Criteria for Diagnosis of Still’s Disease**

<table>
<thead>
<tr>
<th>Major Criteria</th>
<th>Minor Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature &gt;39°C for &gt;1 week</td>
<td>Sore throat</td>
</tr>
<tr>
<td>Leukocytosis &gt;10,000/mm³ with &gt;80% PMNs</td>
<td>Lymph node enlargement</td>
</tr>
<tr>
<td>Typical rash</td>
<td>Splenomegaly</td>
</tr>
<tr>
<td>Arthralgias &gt;2 weeks</td>
<td>Liver dysfunction (high AST/ALT)</td>
</tr>
</tbody>
</table>

ALT, alanine aminotransferase; AST, aspartate aminotransferase; PMNs, polymorphonuclear leukocytes.
generally give a history characteristic of bone pain rather than joint pain. Synovial inflammation accompanies periostal new bone formation, which is visible on radiographs. Correction of the inciting factor (e.g., cure of pneumonia in a child with cystic fibrosis) is likely to alleviate the synovitis. The synovium is characterized primarily by an increased blood supply and synovial cell proliferation. Little infiltration by mononuclear cells is seen. Pain in the bones that increases when extremities are dependent is characteristic, although it is not always present. If clubbing is not present or is not noticed, this entity may rarely be confused with RA.

**Relapsing Seronegative Symmetric Synovitis with Pitting Edema (Uncommon)**

Relapsing seronegative symmetric synovitis with pitting edema (RS3PE) is an uncommon syndrome marked by significant pitting edema of the hands with synovial thickening and joint tenderness. It occurs predominantly in elderly men and is characterized by rapid onset of symmetric synovitis and pitting edema over the involved joints. Edema usually occurs in the hands and ankles overlying involved distal joints. Symptoms rapidly respond to short courses of corticosteroids and can lead to residual abnormalities, including flexion contractures of the wrists and fingers. Patients are RF negative, and increased risk of neoplastic disease has been suggested. A poor response to corticosteroid treatment and the presence of constitutional symptoms have been observed in patients with an underlying malignancy. Patients can evolve into other autoimmune diseases, including seronegative RA. In addition, the syndrome might represent a variant of another disease, such as polymyalgia rheumatica or reflex sympathetic dystrophy, rather than a distinct entity.

**Rheumatic Fever (Uncommon)**

Rheumatic fever is much less common in the developed world than it once was, but it still must be considered in adults with polyarthritis. In adults, arthritis is the most prominent clinical finding of rheumatic fever; carditis is less common than in children; and erythema marginatum, subcutaneous nodules, and chorea are rare. The presentation is often that of an additive, symmetric, large joint polyarthritis (involving lower extremities in 85% of patients), developing within 1 week and associated with a severe tenosynovitis. This extremely painful process is often dramatically responsive to salicylates. In contrast to Still's disease in adults, rheumatic fever generally has no remittent or quiescent periods and shows evidence of antecedent streptococcal infection. It also has a less protracted course than Still's disease. Many similarities have been noted between rheumatic fever in adults and "reactive" postinfectious synovitis developing from *Shigella*, *Salmonella*, *Brucella*, *Neisseria*, or *Yersinia* infection. As rheumatic fever becomes less common, and because penicillin prophylaxis effectively prevents recurrence of the disease, Jaccoud's arthritis (chronic post-rheumatic fever arthritis) is becoming more rare. This entity, described by Bywaters in 1950, results from severe and repeated bouts of rheumatic fever and synovitis, which stretches joint capsules and produces ulnar deformity of the hands without erosions. The same deformity can develop in SLE characterized by recurrent synovitis and soft tissue inflammation and in Parkinson's disease. Differentiating rheumatic fever from RA is particularly difficult when subcutaneous nodules are present with rheumatic fever.

**Sarcoidosis (Uncommon)**

The two most common forms of sarcoid arthritis often can be easily distinguished from RA. In the acute form with erythema nodosum and hilar adenopathy (Löfgren's syndrome), articular problems usually are related to periartritis affecting large joints of the lower extremities, classically the ankles. Differential diagnosis may be complicated because many of these patients have RF in serum. Joint erosions and proliferative synovitis do not occur in this form of sarcoidosis.

In chronic granulomatous sarcoidosis, cyst-like areas of bone destruction, mottled rarefaction of bone, and a reticular pattern of bone destruction with a lace-like appearance on radiographs may simulate destructive RA. This form of sarcoidosis is often polyarticular, and biopsy of bone or synovium for diagnosis may be essential because often no correlation is noted between joint disease and clinical evidence of sarcoid involvement in other organ systems. Poncelet's disease (tuberculous rheumatism) might actually represent granulomatous "idiopathic" arthritis (i.e., sarcoidosis) (see Chapter 117).

**Amyloidosis (Uncommon)**

Deposits of amyloid can be found in synovial and periarticular tissues and are presumably responsible for the joint problems of some patients. The synovial fluid in amyloid arthropathy is noninflammatory, and particulate material with apple-green fluorescence after Congo red staining may be found in the fluid. Amyloid may cause punched-out bone lesions that rarely can mimic RA erosions. Amyloid formed of β2-microglobulin is found in the joints of patients with chronic renal failure, usually patients who are on dialysis (see Chapter 116).

**Malignancy (Uncommon)**

Direct involvement by cancer of the synovium usually manifests as a monoarthritis. Non-Hodgkin's lymphoma can manifest as seronegative polyarthritis, without hepatomegaly or lymphadenopathy. Lymphoma can manifest as a symmetric polyarthritis. In children, acute lymphocytic leukemia can manifest as a polyarticular arthritis. Paraneoplastic syndromes and others related to direct involvement with cancer can mimic RA and are described in detail later (see Chapter 122).

**Multicentric Reticulohistiocytosis (Rare)**

Multicentric reticulohistiocytosis causes severe arthritis mutilans with an opera-glass hand (main en lorgnette). Other causes of arthritis mutilans include RA, psoriatic arthritis, erosive osteoarthritis treated with glucocorticoids, and gout (after tophi are resorbed by treatment with allopurinol). The cell that causes damage to tissues is the
multinucleate lipid-laden histiocye, which apparently releases degradative enzymes sufficient to destroy connective tissue. These cells in aggregate produce multiple small nodules around joints of the hands.

**Pigmented Villonodular Synovitis (Rare)**

Pigmented villonodular synovitis is a nonmalignant but proliferative disease of synovial tissue that has many functional characteristics similar to those of RA but usually involves only one joint. The histopathologic appearance is characterized by proliferation of histiocytes, multinucleate giant cells, and hemosiderin- and lipid-laden macrophages. Clinically, this is a painless chronic synovitis (most often of the knee) with joint effusions and greatly thickened synovium. Subchondral bone cysts and cartilage erosion may be associated with bulky tissue. It is unclear whether this condition should be classified as an inflammation or as a neoplasm of the synovium (see Chapter 123).

**OUTCOMES**

Many difficulties are associated with establishing a change in patterns of RA in different time periods or different communities. The best data suggest that clinical manifestations of the disease and the extent of disability are declining. Epidemiologic studies suggest that the disease is not changing, but that earlier, more effective treatment has diminished morbidity.

Criteria for clinical remission were proposed by ACR/EULAR in 2011. Remission can be defined as absence of disease, but its application to RA patients has changed over time. With more effective treatment available, stricter criteria have become more important. Definitions vary widely and can mean absence of clinical and radiologic signs of disease while on treatment, or a disease state with minimal or no activity after therapy is withdrawn. A previous composite system using the Disease Activity Score (DAS) is a mathematical method that includes swollen and tender joints, ESR, and patient assessments of global health. Notably, this criterion does not mean that the patient truly has complete remission with no evidence of synovitis; however, it can be associated with several active joints. ACR criteria from 1981 require absence of joint tenderness, fatigue, joint pain, joint swelling, and morning stiffness, along with a normal ESR. The goal of the ACR/EULAR revised remission criteria was to develop more stringent but achievable criteria for clinical trials. Remission was defined as tender joint count of 1 or less, swollen joint count of 1 or less, C-reactive protein (CRP) of 1 or less, and patient global assessment of 1 or less. In addition, an index-based definition was included as an alternative. Both definitions performed similarly in validation studies. Some controversy still exists regarding whether the goal should be “clinical” or “imaging” remission when ultrasound or MRI evidence of synovitis is considered. The clinical importance of synovitis diagnosed only by imaging in an asymptomatic patient is uncertain.

With increased numbers of effective therapies available, it becomes increasingly important for physicians to be able to determine which patients would be at greatest risk for progressive destructive disease, and which patients would have a more benign illness that is not erosive and is responsive to moderate intervention. In addition to predicting which patients may or may not develop erosion, it is equally important to identify which patients who already have erosions are more likely to progress rapidly to joint destruction. One study of an inception cohort of patients newly presenting with inflammatory polyarthritis confirmed the fact that although the initial radiographic score is, as expected, a powerful predictor of subsequent radiographic damage, high titers of RF and ACPAs continue to be powerful predictors of deteriorating radiographic damage in subjects receiving conventional therapy. ACPAs (e.g., to fibrinogen, to vimentin, to fibronectin) are more specific for RA than for RF, can appear before disease onset, and are predictive of more aggressive and destructive disease.

Antimutated citrullinated vimentin antibodies recognize peptides with amino acid substitutions that most likely occur as the result of somatic mutations in the joint and might be more predictive of radiographic progression compared with standard ACPAs. Joint erosions and deformity may not be the most important aspects of disease for the patients. In several studies, the Health Assessment Questionnaire has been shown to be an excellent predictor of work disability and mortality, and its results can be discrepant from damage measured by radiographs.

**Mortality**

Infection, renal disease, and respiratory failure traditionally have been the primary factors contributing to excess mortality in RA patients, although congestive heart failure, ischemic heart disease, and peripheral atherosclerosis are more correctly identified as the most common cause of excess mortality in RA (see Chapter 36). In addition, disability develops most rapidly during the first 2 years of RA, supporting the current practice of early aggressive therapy. Persistent premature death and potential widening of the mortality gap in RA patients compared with the general population have been attributed to three potential contributing factors. RA patients are at greater risk for multiple comorbid conditions and poorer outcomes after development of these illnesses. In addition, RA patients often do not receive the best preventive care. Last, systemic inflammation and altered immune responses associated with RA tend to increase and hasten comorbidity and mortality.

Careful epidemiologic studies have indicated that cardiovascular disease is the main cause of increased mortality in RA patients. In the Norfolk Arthritis Register, a primary care–based inception cohort, patients who were seropositive for RF died at an excessive rate within the first 7 years of disease resulting from cardiovascular causes (men, 1.34; women, 2.02) compared with controls. This increased incidence of cardiovascular events in RA patients is independent of traditional risk factors, such as age, sex, smoking status, diabetes mellitus, hypercholesterolemia, systolic blood pressure, and body mass index. The generally accepted explanation is that inflammatory cytokines that are produced in excess in RA (e.g., TNF, platelet-derived growth factor) have the capacity to activate endothelial and subendothelial myofibroblasts, and numerous
inflammatory cells are found in atheromatous plaques. Given that nonrheumatoid patients who have higher levels of C-reactive protein than control groups have higher incidences of coronary disease, these data are consistent with hypotheses. Ultrasonography has shown that RA patients have greater thickness of the common carotid and femoral arteries than healthy controls—a finding that was independent of glucocorticoid therapy but was related to the duration and severity of RA. Platelet-derived microparticles, the small vesicles that are released from the plasma membrane when these cells are activated, are elevated in RA in proportion to disease activity. As was noted earlier, the following factors and pathobiologic mechanisms could contribute to atherosclerosis in RA:

- Immune complex–mediated endothelial damage
- Acute phase reactants (C-reactive protein and serum amyloid A, both of which have proinflammatory activity)
- Inflammatory cytokines
- High expression of endothelial cell leukocyte adhesion molecules
- Medications (e.g., steroids)
- Prothrombotic factors (e.g., increased platelets, fibrinogen, and thromboxane)
- Endothelial cell dysfunction induced by inflammation

Therapies considered for rheumatoid patients must consider the effects on atherogenesis. In patients with an unfavorable vascular profile, these considerations might include supplementation with omega-3 fatty acids in the diet, early use of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors (statins that, in addition to providing lipid-lowering effects, reduce C-reactive protein), attempts to reduce elevated levels of homocysteine induced by methotrexate, avoidance of cyclosporine, and aggressive weight loss disciplines and smoking cessation programs. The IL-6 receptor antibody tocilizumab suppresses several risk factors for mortality, including inflammation, elevated ESR, and elevated C-reactive protein. It also alters the lipid profile by increasing low-density lipoproteins, albeit with a concomitant increase in high-density lipoproteins to maintain a similar ratio. The ultimate effects on cardiovascular risk factors and mortality are still uncertain under these circumstances.

In addition to cardiovascular causes of death associated with RA are causes of death due to complications (articual and extra-articular) of RA and to side effects of therapy. The probability of death varies directly with the severity of complications. Potentially morbid articular complications include the various forms of atlantoaxial subluxation, cricoarytenoid synovitis, and sepsis of involved joints. Unfortunately, mortality from infection in RA has actually increased approximately fourfold to fivefold on average from reports from 1982 to 2004. Extra-articular complications directly causing higher mortality include Felty’s syndrome, Sjögren’s syndrome, pulmonary complications, and diffuse vasculitis.

One of the largest and best documented studies of survival, prognosis, and causes of death in RA was published by Mitchell and associates. In this prospective trial of 805 patients, which included 12 years of observation, 233 patients died during the course of the study, and survivorship was only 50% of that in population controls. Increased mortality associated with RA is impressive and equals that seen in all patients with Hodgkin’s disease, diabetes mellitus, or stroke (age adjusted).

One prediction suggests that tight control of inflammation might decrease cardiovascular and cerebrovascular events and improve survival. Aggressive use of methotrexate can decrease mortality in patients with cardiovascular or noncardiovascular mortality. Biologic agents such as TNF inhibitors seem to improve survival, especially in women. TNF blockade can potentially exacerbate heart disease, however, and increases mortality in individuals with preexisting congestive heart failure. Longer-term studies are required to assess the impact of new agents such as rituximab and abatacept on mortality (see Chapter 64).

Increased risk for malignancy has been observed in RA, especially with lymphoma. RA patients have a two to three times higher risk of developing Hodgkin’s disease, non-Hodgkin’s lymphoma, or leukemia compared with the normal population; this is independent of immunosuppressive therapy. Of lymphomas arising in RA, about half are low grade and half are high grade. Most are B cell lymphomas, although it is not clear whether these originate from clonally proliferated lymphocytes associated with RA. Although the relative risk for total cancer in patients with Felty’s syndrome is only twofold, the relative risk for non-Hodgkin’s lymphoma in this complication of RA is nearly 13-fold—similar to that associated with Sjögren’s syndrome.

A meta-analysis of malignancy in adult RA patients indicated that the risk of lung cancer is increased 1.5- to 3.5-fold compared with the general population. Interstitial fibrosis might serve as a risk factor for lung carcinoma, particularly the bronchoalveolar variety. In contrast, risks for colorectal and breast cancer were slightly decreased in RA patients. RA patients have demonstrated a consistently reduced risk for cancer of the gastrointestinal tract. NSAIDs could be responsible for lowering the risk for this form of cancer, as supported by evidence that these drugs can diminish the occurrence and numbers of colonic polyps.

Early clinical trials suggested that solid tumors might be enhanced in patients with RA treated with TNF blockers. Similar data in patients with Wegener’s granulomatosis suggest that the combination of etanercept and cyclophosphamide can increase the risk for cancer. However, several large registry studies evaluating the effects of TNF inhibitors on cancer rates in RA patients suggest that the oncogenic effect, if it exists, is small.

**Variables Related to Outcome**

In attempting to sort out the relative roles of disease manifestations, compared with nondisease factors, in generating disability in RA, investigators have proposed hypothetical models to predict disability in RA using demographic, sociocultural, and clinical characteristics of a consecutive cohort of RA patients. Although their methods could not be used to explain the dynamics of disability in 41% of cases, disease-related factors explained 33% and nondisease factors (e.g., depression and psychologic status, education) accounted for 26% of the disability. Other studies have emphasized the following disease factors, which correlate with a poorer prognosis and a greater likelihood of joint destruction:

**Table 70-8**  Activities of Daily Living and Visual Analog Questionnaire (see Supplemental Figures 70-1 to 70-12 on www.expertconsult.com)

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Sometimes</th>
<th>Most of the Time</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dress yourself?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Get into and out of bed?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lift a cup or glass to your lips?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walk outdoors on flat ground?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wash and dry your entire body?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bend down to pick up clothing from the floor?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turn faucets on or off?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Get into and out of a car?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. How much pain have you had in the PAST WEEK? (mark the scale)

No pain
Pain as bad as it could be
0 100


- Positive RF in serum
- Positive ACPA in serum
- Rheumatoid nodules
- Elevated Health Assessment Questionnaire level of disability
- Depression
- Persistent ESR elevation (serving as a surrogate for disease control)
- Presence of the shared epitope (QKRAA) on class II major histocompatibility genes

**Assessment of the Individual Patient**

Assessment of disease activity and its progression is different from prognosis. Prognosis extrapolates from a known set of indices (as noted earlier) and the degree of measured activity of disease to prediction of the outcome. Assessment is the accurate evaluation of disease progression over time. Although the indices listed in the previous section are useful as a way to predict outcomes from one-time measurements, use of three or more assessment measures provides the physician with a graph of progression in an individual patient that he or she can try to flatten out by therapy. Whatever assessment index is used, it should be used early in the patient's disease, so that values are recorded before significant loss of function.

For most patients, a self-report questionnaire based on degrees of difficulty in performing activities of daily living correlates well with joint count, radiographic score, acute phase reactants, grip strength, walking time, functional class estimates, and global self-assessment. One useful self-report includes only eight items from the much longer Stanford Health Assessment Questionnaire (Table 70-8). The limitation of this form—failure to detect clinical improvement in patients with few impairments in activities of daily living—may be offset by its acceptability to patients within busy office practices.

In some situations, more comprehensive joint counts are needed. These include points when large changes in drug therapy are about to be instituted, and when patients are to undergo joint reconstruction by orthopedic or hand surgeons. The Thompson index uses a few joints and weights data from each joint to reflect the joint surface area, giving a better measure of the “burden of synovitis.”

The choice of imaging techniques and measures is important in assessment of the destructive lesions of RA. The inflammatory lesion in RA is reflected reasonably well by heat, pain, swelling, and tenderness. Joint destruction can occur with minimal inflammation, however. MRI and ultrasound provide ways to visualize pannus development and loss of cartilage (see Chapter 58). In each patient, when the diagnosis of RA is reasonably certain, these measures of assessment and estimates of prognosis should be recorded. They should be major determinants of what therapies are instituted (see Chapter 71).

**Selected References**

Supplemental Figure 70-1  Synovitis of metacarpophalangeal joint with volar subluxation of digits.

Supplemental Figure 70-2  Synovitis of metacarpophalangeal joints.

Supplemental Figure 70-3  Erosion and rupture of fourth and fifth extensor tendons.

Supplemental Figure 70-4  Rupture of Baker’s cyst posteriorly at left knee.
Supplemental Figure 70-5  Carpal-metacarpal subluxation.

Supplemental Figure 70-6  Rheumatoid nodules in olecranon bursa and along proximal ulna.

Supplemental Figure 70-7  Livedo reticularis and incipient digital gangrene in rheumatoid vasculitis.

Supplemental Figure 70-8  Exuberant synovitis of metacarpophalangeal and proximal interphalangeal joints.
Supplemental Figure 70-9  Subluxation of metatarsal heads and ulcer is developing over the second head.

Supplemental Figure 70-10  Multiple swan neck deformities in rheumatoid arthritis.

Supplemental Figure 70-11  Boutonniere deformities in rheumatoid arthritis.

Supplemental Figure 70-12  Multiple infarcts, including classic nail fold infarcts, due to small vessel vasculitis in rheumatoid arthritis.
The treatment of rheumatoid arthritis (RA) has evolved dramatically over the past 30 years, perhaps more so than any of the rheumatic diseases. The majority of patients newly diagnosed with RA in 2013 can expect to have their disease in remission if treated early by a rheumatologist. This remarkable fact has come about because of a tremendous expansion of the number of disease-modifying antirheumatic drugs (DMARDs) available (Table 71-1), the realization that these drugs can and should be used in combinations,1,3 and the acceptance that all patients should be treated to a target or goal of remission or low disease activity.4,14 To put the current situation into perspective and to celebrate how far we have come, a look back on the most immediate history of RA treatment as chronicled in the 30 years of the Textbook of Rheumatology (TOR) seems appropriate.

When the first edition of TOR was published in 1981, Dr. Ruddy, the author of the RA treatment chapter, discussed intramuscular gold and penicillamine as the mainstay of RA treatment, hydroxychloroquine (HCQ) was mentioned as a possible option, and cyclophosphamide and azathioprine were discussed as experimental therapies. By 1985 and the second edition of TOR, azathioprine was approved for treatment of RA and cyclosporine and methotrexate (MTX) appeared in experimental sections along with total lymphoid irradiation. By the fourth edition in 1993, Ted Harris, always ahead of his time, championed early DMARD treatment but suggested starting with the least toxic DMARDs—HCQ and sulfasalazine (SSZ). MTX was recommended only after the failure of HCQ, gold, and penicillamine. By 1997 and the fifth edition, early DMARD therapy was firmly established, MTX had moved to the front of the line of DMARD therapy, and combinations of conventional DMARDs were first prominently mentioned. The sixth edition (2001), authored by Mark Genovese, reflected sweeping changes in the landscape: Combination DMARD therapy was firmly established; leflunomide had become an alternative to MTX; and, most importantly, the biologics etanercept (ETAN) and infliximab had forever changed the landscape. In the immediate predecessor to this edition of TOR (2009), the concept of treating to a target was discussed and our armamentarium of biologics had expanded to three anti–tumor necrosis factor (TNF) agents and two biologics with new mechanisms of action: abatacept and rituximab. Critical features of contemporary therapy of RA as discussed in the ninth edition of TOR include the strategy of treating all patients to the goal of remission or low disease activity and emphasizing treating to target and not the drugs used; the rediscovery of conventional combination DMARDs; the introduction of two new TNF inhibitors and a biologic with yet another mechanism of action (anti-IL-6 receptor), tocilizumab; and highlighting the critical role of comorbidities, particularly cardiovascular disease in RA. In this edition, for the first time a section on tapering or discontinuing therapy in patients who are doing well has been added, which illustrates just how far we have come in the past 30 years.

Despite how far we have come, many challenges remain and include first and foremost identifying markers that predict in a differential fashion who will respond to or have side effects from which treatment and developing methods to allow us to measure the amount of immunosuppression that agents are producing. If the current pace of advance continues, the next few editions of TOR may see discussion of RA remissions off all treatments and even treatment of and perhaps cures for “patients” before they develop symptomatic RA.

This chapter attempts to discuss the broad principles of treatment, the goals of RA therapy, the timing of different therapies, and the strategies of employing the plethora of options now available to achieve the best control of each patient. Most of the drugs are not covered in detail here; please refer to other excellent chapters for specifics on the NSAIDs (Chapter 59), glucocorticoids (Chapter 60), traditional DMARDs (Chapter 61), immunoregulatory drugs
GOAL OF RHEUMATOID ARTHRITIS TREATMENT

It is remarkable that rheumatologists now have more than 19 approved conventional or biologic DMARDs to choose from. However, despite all these terrific DMARD options, the most important paradigm shift for the treatment of RA has been the realization that patients should be treated early and to a target of low disease activity or remission. For those outside the rheumatologic community, this seems like stating the obvious—if you have hypertension, hyperlipidemia, or diabetes, patients are of course treated to get the blood pressure, low-density lipoprotein, or HbA1C levels down to a defined and easily measured goal. The problem in RA has been having valid reproducible measures of disease activity and remission and then routinely measuring and following those in a clinic. Unfortunately, in RA, there is no single examination finding or laboratory test that satisfactorily measures disease activity.

Many measures have been proposed, and all are composite measures that include information derived from some combination of joint examinations, patient and physician assessment of disease activity, patient function, and laboratory measures of inflammation (erythrocyte sedimentation rate [ESR] or C-reactive protein [CRP]). Recently, the American College of Rheumatology (ACR) has endorsed a list of disease activity measures that have been shown to correlate with outcomes. Table 71-2 is a partial list of some of the better known of these measures. Each of these measures have strengths and weaknesses; some rely on only data from the patient, some require complete joint counts by clinicians, and some require laboratory tests to measure.

Table 71-1 Disease-Modifying Antirheumatic Drugs*

<table>
<thead>
<tr>
<th>Conventional</th>
<th>Biologics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methotrexate</td>
<td>Etanercept (Enbrel)</td>
</tr>
<tr>
<td>Hydroxychloroquine</td>
<td>Infliximab (Remicade)</td>
</tr>
<tr>
<td>Sulfasalazine</td>
<td>Anakinra (Kineret)</td>
</tr>
<tr>
<td>Leflunomide</td>
<td>Adalimumab (Humira)</td>
</tr>
<tr>
<td>Gold (intramuscular and oral)</td>
<td>Abatacept (Orencia)</td>
</tr>
<tr>
<td>Azathioprine</td>
<td>Rituximab (Rituxan)</td>
</tr>
<tr>
<td>Minocycline</td>
<td>Certolizumab (Cimzia)</td>
</tr>
<tr>
<td>Cyclosporine</td>
<td>Golimumab (Simponi)</td>
</tr>
<tr>
<td>Penicillamine</td>
<td>Tocilizumab (Actemra)</td>
</tr>
<tr>
<td>Glucocorticoids</td>
<td></td>
</tr>
</tbody>
</table>

*Currently available drugs that have the ability to slow or halt progression of rheumatoid arthritis including radiographic progression.

(Chapter 62), and anticytokine therapies or biologics (Chapter 63).

Because none of our therapies cure RA, it seems obvious that the next best goal should be remission. The concept of remission as a goal for RA patients is problematic, however. First, a remission definition that is both relevant and practical has been elusive. To be relevant, remission should be highly predictive of the absence of disease progression over time. To be practical, for clinicians, a remission definition should be easy to apply in real time to patients seen in a clinic as discussed earlier with regard to measures of disease activity. Recently, a new definition of remission for use in clinical trials has been developed by ACR and the European League against Rheumatism (EULAR) (Table 71-3). This definition has been rigorously tested against short-term radiologic outcomes in 1- to 2-year randomized controlled trials (RCTs) follow-up. This definition standardizes remission and is therefore a huge step forward for reporting and comparing results across clinical trials.

However, this definition was designed for clinical trials, not clinical care, where the need to have results of CRPs in real time becomes a problem. Versions of this that do not require laboratory values have been suggested but not fully accepted (e.g., CDAI, Patient Activity Scale). Perhaps more problematic, many believe that remission defined by clinical data alone will always underestimate the amount of low-level disease activity that could be found if synovial biopsies or advanced imaging techniques such as ultrasound (US) or magnetic resonance imaging (MRI) were employed (see Chapter 58). Significant data exists that many and perhaps most RA patients who meet definitions of “remission” have active disease if assessed by US or MRI. Indeed, the newly accepted ACR/EULAR definition allows for a swollen joint, which many would argue is not actually remission. Another major problem with “remission” is that from currently available data it is not at all clear that remission, regardless of how it is defined, should be the treatment goal for all RA patients. Many patients do well despite low levels of disease activity. This situation may be analogous to the recent studies that show pushing HbA1C levels below 6.5, which seemed appropriate for diabetic control, was inflammation. The busy clinician rarely has time to document more than 60 tender and swollen joints or wait for laboratory test results to make decisions on patients during their visit. Therefore measures that simplify this process as much as possible are being embraced including those that limit the joints counted to 28 (Disease Activity Score 28 [DAS28]), do not require laboratory tests (Clinical Disease Activity Index [CDAI]), or are entirely dependent on patient data Routine Assessment Patient Index (RAPID). There is high correlation among these measures, so currently in the clinic it is very important that disease activity is measured and less important which measure is used.

Table 71-2 Instruments Used to Measure Rheumatoid Arthritis Disease Activity

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Score Range</th>
<th>Remission</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease Activity Score in 28 joints (DAS28)</td>
<td>0-9.4</td>
<td>≤2.6</td>
<td>≤3.2</td>
<td>≥3.2 and ≤5.1</td>
<td>&gt;5.1</td>
</tr>
<tr>
<td>Simplified Disease Activity Index (SDAI)</td>
<td>0.1-86.0</td>
<td>≤3.3</td>
<td>≤11</td>
<td>&gt;11 and ≤26</td>
<td>&gt;26</td>
</tr>
<tr>
<td>Clinical Disease Activity Index (CDAI)</td>
<td>0-76.0</td>
<td>≤2.8</td>
<td>≤10</td>
<td>&gt;10 and ≤22</td>
<td>&gt;22</td>
</tr>
<tr>
<td>Rheumatoid Arthritis Disease Activity Index (RADAI)</td>
<td>0-10</td>
<td>≤1.4</td>
<td>≤2.2</td>
<td>2.2 and ≤4.9</td>
<td>&gt;4.9</td>
</tr>
<tr>
<td>Patient Activity Scale (PAS or PASII)</td>
<td>0-10</td>
<td>≤1.25</td>
<td>&lt;1.9</td>
<td>≥1.9 and ≤5.3</td>
<td>&gt;5.3</td>
</tr>
<tr>
<td>Routine Assessment Patient Index Data (RAPID)</td>
<td>0-30</td>
<td>≤1</td>
<td>&lt;6</td>
<td>≥6 and ≤12</td>
<td>&gt;12</td>
</tr>
</tbody>
</table>
associated with increased cardiovascular mortality mainly due to hypoglycemia in patients with prior cardiovascular histories.47

- When do the risks and considerable expense of some of our RA therapies outweigh the benefits of escalating therapy further?
- Which patient who has improved dramatically but still has two tender or swollen joints needs a third biologic?
- With regard to the previously cited diabetic patients with cardiovascular disease, which RA patients are most at risk if we push too hard for remission?
- Finally, with current therapies, the vast majority of remissions in RA require ongoing treatment with DMARDs, so the concept of a true remission, meaning one where no therapy is required, remains beyond our current reach for the majority of patients.

Despite the problems defining remission or low-disease activity, it is clear that patients do better if clinicians have a goal. The Tight Intensive Control of RA (TICORA) study,18 was the first to convincingly demonstrate this in a randomized fashion. TICORA was a Scottish study in which patients with less than 5 years of disease were randomized to receive either routine care or to receive intensive care. Both groups were treated with an algorithm of conventional DMARDs (Figure 71-1A). The routine care group had regular follow-up and monitoring, while the intensive group was seen monthly and had prescribed escalation of therapy (protocolized) if they had not achieved the goal of low disease activity (defined in this study as DAS ≤ 2.4). Both groups improved significantly, but the group that was treated to a target (intensive group) did significantly better with mean DAS scores (<1.6) in the remission range at 18 months (Figure 71-1B). In the intensive group, 71% achieved an American College of Rheumatology 70% improvement criteria response (ACR70) compared with 18% in the routine care group (P < .0001). Further, this clinical improvement translated to significantly less radiographic progression of erosions compared with the routine group (0.5 vs. 3.0; P = .002). Importantly, this improved disease control was not associated with an increase in treatment-associated adverse events. Finally, despite more frequent visits, intensive therapy resulted in cost savings even in the short term. These results were particularly remarkable considering they were obtained using conventional DMARDs alone (see Figure 71-1A) without the use of biologics. Findings from other studies have corroborated these findings.19-21 Further, a meta-analysis of tight control suggested that tight control strategies work best if protocolized, as was done in TICORA.

Although the TICORA investigators selected low disease activity as the target, the target could have been remission as discussed earlier. Most of the previously endorsed measures of disease activity have defined levels that signify “remission” (see Table 71-2). Predictably, the harder we push for remission with increasing numbers and doses of DMARDs, both conventional and biologic, the more toxicity and the expense (Table 71-4) of our treatments become a concern. Both the ACR and EULAR guidelines, as well as recent reviews,14 currently state that low disease activity or remission is the goal, and they leave the decision on which one is most appropriate for each unique patient’s situation to the clinician. Therefore until further data elucidate this question, clinicians will need to continue to practice the art and science of medicine when selecting the most appropriate target for each patient.

Table 71-3  ACR/EULAR Definitions of Remission in Rheumatoid Arthritis Clinical Trials

<table>
<thead>
<tr>
<th>Boolean-Based Definition</th>
<th>Index-Based Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>At any time point, patient must satisfy all of the following:</td>
<td>At any time point, patient must have a Simplified Disease Activity Index score of ≤3.3</td>
</tr>
<tr>
<td>Tender joint count ≤1*</td>
<td>*Include 28 joints plus feet and ankles</td>
</tr>
<tr>
<td>Swollen joint count ≤1*</td>
<td>ACR/EULAR, American College of Rheumatology/European League against Rheumatism</td>
</tr>
<tr>
<td>Patient global assessment ≤1 (on a 0-10 scale)</td>
<td></td>
</tr>
</tbody>
</table>

*Other DMARD MTX + CSA
Sulfasalazine + MTX and HCQ
Dose escalation
*Other DMARD MTX + CSA
Oral prednisolone

A RANDOMIZED CONTROLLED TRIAL OF TIGHT CONTROL OF RHEUMATOID ARTHRITIS (TICORA)

Intensive therapy: Goal of DAS < 2.4
Sulfasalazine + MTX and HCQ
Dose escalation
*Other DMARD MTX + CSA
Oral prednisolone

A *Intra-articular steroids were used, but biologics were not.

Figure 71-1  A, Schematic of treatment escalation used in TICORA trial. B, Comparison of intensive and routine treatment groups. CSA, cyclosporine A; DAS, disease activity score (44 joints); DMARD, disease-modifying antirheumatic drug; HCQ, hydroxychloroquine; MTX, methotrexate. (Modified with permission from Grigor C, Capell H, Stirling A, et al: Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): a single-blind randomised controlled trial, Lancet 364:263–269, 2004.)
Table 71-4  Average Medication Expense

<table>
<thead>
<tr>
<th>Medication</th>
<th>Dosage</th>
<th>Monthly Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methotrexate</td>
<td>20 mg/wk</td>
<td>$26</td>
</tr>
<tr>
<td>Hydroxychloroquine</td>
<td>200 mg twice daily</td>
<td>$35</td>
</tr>
<tr>
<td>Sulfasalazine</td>
<td>1 g twice daily</td>
<td>$30</td>
</tr>
<tr>
<td>Prednisone</td>
<td>10 mg/day</td>
<td>$10</td>
</tr>
<tr>
<td>Triple therapy</td>
<td>Above doses</td>
<td>$91</td>
</tr>
<tr>
<td>Etanercept</td>
<td>50 mg/wk</td>
<td>$1974</td>
</tr>
<tr>
<td>Adalimumab</td>
<td>40 mg every other wk</td>
<td>$1915</td>
</tr>
<tr>
<td>Infliximab</td>
<td>300 mg/4 wk</td>
<td>$2264†</td>
</tr>
<tr>
<td>Rituximab</td>
<td>1500 mg every 6 mo</td>
<td>$1597*</td>
</tr>
<tr>
<td>Abatacept</td>
<td>750 mg monthly</td>
<td>$1690*</td>
</tr>
<tr>
<td>Tocilizumab</td>
<td>400 mg monthly</td>
<td>$1555*</td>
</tr>
</tbody>
</table>

*U.S. $/month cost to consumer from large national chain pharmacy (Walgreens) in 2011.  † Based on average wholesale price as listed by Redbook and does not include infusion costs. *Does not include infusion costs.

CLASSES OF DRUGS

DMARDs: Methotrexate, Sulfasalazine, Hydroxychloroquine, and Leflunomide

The definition of a DMARD is one that has the ability to change (for the better) the course of RA. The most rigorous application of this definition requires RCTs that show not only the ability to change the clinical course of the disease but also the ability to decrease or halt the radiographic progression. By this definition, all of the 10 conventional DMARDs and the 9 biologic DMARDs listed in Table 71-1 qualify with the possible exception of minocycline and HCQ, where only weak evidence exists for radiographic benefits. With the DMARDs listed in Table 71-1 and using these drugs individually or in combinations of two, three, or four as is often done, there are 2569 possible combinations for each individual patient, assuming that biologics are not used in combinations with each other. Obviously, this huge number of choices is both good and bad news for the clinician; it is great to have all the options but impossible to keep them all straight. Therefore to employ these effectively, the clinician must have goals, strategies, and an up-to-date knowledge of the drugs and their interactions and toxicities.

The most widely used conventional DMARDs—MTX, SSZ, HCQ, and leflunomide (LEF)—are discussed in detail in Chapter 61. Together, these four DMARDs along with glucocorticoids (see Chapter 60) currently account for the vast majority of conventional DMARD use. Although less commonly used, gold (both intramuscular [IM] and oral), azathioprine, cyclosporine, and the tetracyclines (minocycline and doxycycline), which are not covered elsewhere in the book, are therefore discussed as follows. Penicillamine is of historical interest but is rarely used and is not discussed in this chapter.

Biologic DMARDs

The biologic DMARDs are discussed in detail in Chapter 63. Within this class of agents we now have the ability to inhibit multiple inflammatory cytokines including TNF (with three monoclonal antibodies (infliximab, adalimumab, and golimumab); the TNF receptor protein (ETAN); and the pegylated Fab fragment certolizumab, interleukin-1 (IL-1) with the IL-1 receptor antagonist (anakinra), IL-6 (monoclonal antibody to the IL-6 receptor tocilizumab) or to kill or inhibit cell lines important in inflammation including B cells (rituximab) and T cells (abatacept). It is an understatement to say biologics have changed the landscape of RA therapy forever both in terms of therapeutic expectations and understanding of RA pathogenesis. Because of their often quick onset of action, particularly with the TNF inhibitors, and their ability to retard radiographic progress, they are increasingly used earlier and more often in RA. The challenge for clinicians is to appropriately integrate conventional and biologic therapies and to use biologics when necessary but to make sure the much less expensive conventional therapies have been maximized.

Glucocorticoids

Glucocorticoids, discussed in detail in Chapter 60, have a long and storied history in the treatment of RA. RA was selected as the first disease to be treated with “compound E” at the Mayo Clinic in 1948. Responses were both rapid and dramatic; an analysis of the first 14 patients treated revealed that 100% improved their ESRs by more than 50% within 1 to 3 months, with 80% improving ESRs by at least 70% (an ESR70—a convenient ACR70). Several landmark studies have proven not only clinical efficacy but also the significant radiographic efficacy of glucocorticoids. The recent COBRA trial and the COBRA arm of the BeSt study (discussed later) again demonstrated the significant clinical and radiographic benefit that glucocorticoids can provide. Despite the rapid onset and all the other positive benefits of glucocorticoids in RA, their toxicities are also legend. Currently glucocorticoids are most often and most appropriately used along with DMARDs as part of initial “induction” therapy to get RA patients under control rapidly and then aggressively tapered as the slower-acting DMARDs start to kick in. Historically, the belief has been that once an RA patient was on glucocorticoids, he or she would never get off. This is clearly not true with the effective DMARDs now available—successful tapering is the rule. If a patient cannot be successfully tapered off or at least tapered to an “acceptable” low dose, it is a strong indication that the current DMARD program is not working. Long-term use of doses equivalent to prednisone of greater than 7.5 to 10 mg per day is a clear indication that DMARD therapy needs to be escalated. Importantly, glucocorticoids should almost never be used in RA without DMARDs.

Other Conventional DMARDs

Gold Salts

Gold injections have been used in the treatment of RA for close to a century—initially intramuscularly and, more recently, orally. With the advent of newer agents, gold is rarely used in most parts of the world. IM gold is a difficult and cumbersome therapy. It is initiated with weekly IM injections, usually starting with 10 mg the first week, 25 mg the second week, and then 50 mg thereafter, until a response is seen, generally between 3 and 6 months. Once the desired
efficacy is seen, the IM injections can eventually be monthly. Frequent monitoring with complete blood counts (CBCs) and urine for protein is necessary. Significant toxicities can occur and include skin rashes, bone marrow depression, and nephrotic syndrome. As problematic as gold treatment is, there is a wealth of evidence that IM gold therapy is beneficial for RA including retarding radiographic progression. Some patients, perhaps 10% to 20% of those started on IM gold, will have essentially complete long-term remissions if maintained on injections every 2 to 4 weeks. Of recent interest was a 48-week RCT in patients with active disease despite MTX; the addition of IM gold to MTX resulted in 61% of patients achieving an ACR20 compared with 30% for the MTX + placebo group (Figure 71-2). Despite the known benefits, there is also ample evidence that the two IM compounds, gold sodium thiomalate and gold sodium thioglucose, are being used less by rheumatologists because of the need for meticulous monitoring for serious toxicity and the inconvenience of administration and monitoring.

Auranofin, the gold oral preparation, has been available for more than 20 years but is rarely used. Auranofin has different and less severe toxicity than the IM gold and reportedly less efficacy. Cytopenia and proteinuria are rare, but an enterocolitis with diarrhea leads to intolerance in many. Auranofin is an effective DMARD, but RCT data show that it is less effective than MTX, IM gold, penicillamine, or SSZ. Until or unless factors can be found that reliably predict who will get the almost magical clinical response that can be seen with gold, this cumbersome and often difficult-to-manage therapy will continue to disappear. To this end, HLA-DR3 is found in more patients who develop either thrombocytopenia or nephropathy while taking gold injections. These data must be balanced against the evidence that human leukocyte antigen (HLA)-DR3 may be associated with a better response to gold therapy, which corroborates a long-time belief shared by many clinicians that those patients who get rashes on gold are often destined to have excellent clinical responses.

### IMMUNOSUPPRESSIVE AGENTS

#### Azathioprine

Azathioprine (AZA), 50 to 200 mg/day, has been used to treat RA for almost 50 years. Because it has been generic for many years, little recent research has been done. Although clearly not a first-line DMARD in contemporary RA treatment, AZA is most commonly used as a substitute for MTX when there are contraindications or intolerance to MTX. This most commonly arises in patients with the
so-called “MTX flu,” but other situations including pregnancy, liver disease, and renal disease may be indications for AZA in RA. Azathioprine is usually used in combination with other conventional or biologic DMARDs. McCarty and co-workers, who was one of the pioneers of combination therapy in RA, has reported on the combination of MTX, AZA, and HCQ in 69 patients treated in an open-label fashion. With this combination, 45% of patients reached remission by the old ACR criteria and this combination was well-tolerated.

Neutropenia is the most common complication of AZA treatment. Neutropenia can be predicted with a genetic test for polymorphisms of the enzyme thiopurine methyltransferase (TMPT). Patients who are homozygous for the mutant polymorphism that is nonfunctional (1 in 300 or 0.3% of patients) are sensitive to bone marrow and other toxicities of AZA. Patients who are heterozygotes (perhaps 10% of the population) may have milder neutropenia. Unfortunately, this test is expensive. In some centers it may cost up to $1000 and is not always reimbursed. Some clinicians elect to start with low doses of 50 mg/day and check CBCs at 2 weeks and then increase the dose as needed if the white blood cell (WBC) count is normal. It has been speculated that the subset of patients with the nonfunctional polymorphisms could be the patients who, when AZA was added to a stable MTX regimen, developed an acute febrile toxic reaction characterized by fever, leukocytosis, and a cutaneous leukocytoclastic vasculitis.

Cyclosporine

In the 1990s, cyclosporine (CSA) gained a foothold in the treatment of RA. Cyclosporine, mostly used in transplantation to prevent allograft rejection, inhibits the activation of CD4+ helper-inducer T lymphocytes by blocking IL-2 and other T helper type 1 cytokine production and by inhibiting CD40 ligand expression in T lymphocytes. The latter effect prevents T cells from delivering CD40 ligand-dependent signals to B cells. Interest in CSA peaked in the mid-1990s when Tugwell and colleagues showed that the addition of CSA (2.5 to 5 mg/kg/day) to a stable dose of MTX provided substantial additive benefit over MTX alone. In the CSA + MTX group, ACR20 responses were achieved by 48% compared with 16% for placebo (see Figure 71-2). Additionally, this therapy seemed to slow radiographic progression of erosions. In this trial, the dose of CSA was decreased if the patient’s creatinine level increased to more than 30% of initial values. Unfortunately, follow-up reports on this regimen have revealed that only 22% of patients continued on this combination at 18 months with the most common reasons for discontinuation being hypertension or increasing creatinines.

MINOCYCLINE AND DOXYCYCLINE

Tetracycline and derivatives have a long and somewhat checkered history with regard to the treatment of RA and other arthritides. The mechanism of action of tetracyclines in RA is poorly understood. Tetracyclines are, of course, antibiotics, but additionally they inhibit metalloproteinases, modulate immune responses, and have anti-inflammatory effects. No evidence indicates that tetracyclines treat the “infection that causes RA” as was touted by some of the original supporters. However, it is entirely possible that inhibition of nonspecific infections that upregulate the immune response (IL-1, TNF, IL-6) such as periodontitis, bronchitis, and gastritis, to name a few, may be helpful in controlling disease in RA patients. Tetracyclines also have the ability to inhibit biosynthesis and activity of matrix metalloproteinases that have a principal role in degrading articular cartilage in RA. This has been effective in animal models of osteoarthritis (OA) treatment. The presumed mechanism is through chelation of calcium and zinc molecules, which subsequently leads to altered molecular conformations of proenzymes sufficiently to inactive them. Minocycline has mild but definite inhibitory effects on synovial T cell proliferation and cytokine production and has been shown to upregulate IL-10 production. Further evidence of its ability to modulate the immune system is the fact that it is known to induce anti-DNA-antibody positive lupus in some patients, especially when used to treat acne.

Given in a dose of 100 mg twice daily, moderate statistically significant improvement in clinical parameters of disease activity was found in patients with established RA treated with minocycline compared with placebo. Findings in the treatment of early RA have been more impressive. A study of 46 patients with early rheumatoid factor (RF)-positive RA who had not received previous treatment reported 65% of patients meeting 50% improvement in tender and swollen joints, duration of morning stiffness, and ESR (Paulus criteria), whereas only 13% of the placebo recipients improved similarly over a 6-month period. In 2001 the results of a 2-year trial comparing minocycline with HCQ were published. In this small study of patients with early RF-positive RA, the patients treated with minocycline were more likely to achieve an ACR50 (the primary end point) than the patients treated with HCQ (60% vs. 33%) and were more successful in tapering glucocorticoids. This study reconfirms the potential utility of minocycline, particularly in early RF-positive patients.

Although less studied, there is evidence supporting the use of doxycycline in the treatment of RA. In a trial of patients with early RA, doxycycline plus MTX was compared with the use of MTX alone. Investigators studied low-dose doxycycline (20 mg twice a day) and high-dose doxycycline (100 mg twice a day) in combination with MTX and found that both approaches were superior to MTX alone. Despite the positive results of this study, replication is necessary.

Potential side effects of tetracyclines include light-headedness, vertigo, rare liver toxicity, drug-induced lupus, and, with longer-term use, cutaneous hyperpigmentation. Elderly patients appear to be at an increased risk of vertigo. Patients on minocycline have developed lupus-like syndromes, complete with autoantibodies including anti-DNA and occasionally perinuclear antineutrophil cytoplasmic antigen. Drug-induced lupus has not been reported to develop with doxycycline or with the use of minocycline in patients with RA. The hyperpigmentation can be impressive and may limit treatment in some. Hyperpigmentation resolves with discontinuation of minocycline and occurs much less commonly if at all with doxycycline.
NONSTERoidal ANTi-INFLAMMATORY DRUGS

NSAIDs including salicylates, covered in detail in Chapter 59, have been a ubiquitous part of RA treatment for more than a century. Over the past several decades as toxicities, particularly gastrointestinal and cardiovascular, have become apparent and as DMARDs have become better, the use of NSAIDs has fortunately declined. The somewhat surprising cardiovascular toxicities, now known to be strongly associated with not only the cyclooxygenase-2 (COX-2) specific NSAIDs but essentially all NSAIDs, have been particularly concerning with regard to the RA patient populations in which the main excess mortality is largely due to accelerated cardiovascular disease. Like glucocorticoids, NSAIDs should rarely be used without DMARDs. Also like glucocorticoids, the goal should be to taper off as soon as possible to avoid gastrointestinal and cardiovascular toxicities.

Treatment Approaches and Strategies

As detailed earlier, the goal of treatment for all RA patients is remission or at least low disease activity; other than toxicity concerns and affordability issues, clinicians should not care which drug or combination of drugs are used in an individual patient but should focus on getting patients to target with whatever it takes. There is no magic or correct DMARD or combination of DMARDs that is right for all patients. Each patient presents a unique challenge and comes with unique expectations, biases, disease activity level, damage burden, comorbidities, and insurance coverage issues. One of the most important areas of investigation in RA is to identify parameters that will predict in a differential fashion which patient will respond to which therapy—so far no clear-cut answers that are applicable to the clinical care of the vast majority of patients have emerged. Some have suggested treatment should be different for RA patients with good prognosis versus poor prognosis. This concept is problematic; separating patients into good versus poor prognosis is difficult. Although data suggest certain features are associated with worse prognosis (Table 71-5), unfortunately no data suggest that stratifying our therapies on the basis of prognosis at the individual patient level yields better outcomes. For example, if we could perfectly score prognosis on a scale of 1 to 10, those patients with intermediate scores would conceivably benefit the most from our most aggressive therapies while those with low scores do not need them and those with the highest scores may have an unacceptable benefit-to-toxicity ratio. Most patients who meet criteria for classification of RA in clinic and essentially all of those included in clinical trials have poor-prognosis RA with multiple factors listed in Table 71-5. Regardless of the prognostic factors, the goal for each patient is to achieve at least low disease activity. Until or unless parameters are identified, rheumatologists will of necessity continue to use their clinical judgment at the individual patient level.

Without parameters that predict in a differential way response to medications in terms of efficacy or toxicity, one approach to treatment decisions is illustrated in Figure 71-3. Note that several iterations may be required. This figure indicates that we do not have these much-needed parameters and emphasizes that if we hope to treat RA in a rational, scientific way, it is critical to find these parameters. Figure 71-3 also highlights and reinforces the need to include bio-banks with all of our clinical trials. With regard to individual patients, Figure 71-4 (ACR recommendations) is perhaps a more practical approach. Specific patient populations are discussed as follows.

<table>
<thead>
<tr>
<th>Table 71-5 Factors Associated with Poor Prognosis in Rheumatoid Arthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of rheumatoid factor and titer</td>
</tr>
<tr>
<td>Presence of antibodies to anticitrullinated protein antibody and titer</td>
</tr>
<tr>
<td>Presence of shared epitope (HLA-DR alleles) and number of copies</td>
</tr>
<tr>
<td>Presence of erosive disease at presentation</td>
</tr>
<tr>
<td>Disease activity at presentation</td>
</tr>
<tr>
<td>Magnitude of erythrocyte sedimentation rate or C-reactive protein elevations</td>
</tr>
<tr>
<td>Presence of nodules or other extra-articular features</td>
</tr>
<tr>
<td>Female gender</td>
</tr>
<tr>
<td>Current and past smoking</td>
</tr>
<tr>
<td>Obesity</td>
</tr>
</tbody>
</table>

Treatment of the DMARD-Naive Patient

The most critically important principle of treating RA effectively is to initiate and rapidly advance DMARD therapy as early as possible. Although this is a universally accepted principle, there is a paucity of rigorous data that directly addresses this point. Few randomized double-blind trials in which patients are randomized to early treatment versus late treatment have been conducted, and it is unlikely, not to mention unethical, that any more will be forthcoming. Rather, we accept this foundation principle on the basis of the common and firm belief that treating patients early prevents damage and deformity and preserves...
function. Many trials and case series provide strong, credible evidence to support this central tenet. Evidence to support this commonsense belief comes from cohort studies of early versus delayed therapy,\textsuperscript{66} randomized studies of intensive therapy versus usual care,\textsuperscript{18} RCTs of combination versus monotherapy,\textsuperscript{19,20,29,67-71} and finally studies of what was previously defined as preclinical RA.\textsuperscript{65}

One cohort study of note reported on early RA patients. The first cohort received DMARD therapy early (mean, 123 days after diagnosis), whereas the second cohort received DMARDs very early (mean, 15 days after diagnosis\textsuperscript{66}). The first cohort had significantly more radiographic damage at 2 years and, importantly, continued to have radiographic progression while the second did not. The findings of the TICORA trial (detailed earlier\textsuperscript{18}) clearly show the advantages of better control of disease earlier. Multiple studies in early RA have demonstrated that when groups of patients are compared, those that get combinations of therapy fare better than those that get monotherapy.\textsuperscript{19,20,29,67-71} COBRA with the combination of MTX, SSZ, and prednisolone versus SSZ alone; FINRACO\textsuperscript{67} with the combination of MTX, SSZ, HCQ, and prednisone versus SSZ; BeSt with the multiple combinations compared with step-ups or switches\textsuperscript{19,20}; ATTRACT\textsuperscript{69} with the combination of infliximab and MTX versus MTX alone and PREMIER with the combination of adalimumab and MTX versus each alone are some examples of these.\textsuperscript{70} In all these trials, the combinations outperformed monotherapy.
The PROMPT trial addressed a somewhat different question—patients with inflammatory arthritis who could not yet be classified as RA were randomized to MTX treatment or placebo, and the end point was the development of clinical RA. The MTX-treated group had significantly delayed progression to full-blown RA. Taken together, all these data make a compelling argument for early DMARD therapy in RA.

The new ACR/EULAR RA classification criteria should allow us to classify patients with RA earlier and therefore treat RA patients earlier. The criteria, discussed in detail in Chapter 70, were designed to allow rheumatologists to classify patients in clinical trials with RA as early as possible. Gone is the previous absolute requirement that all patients must have certain features for a minimum of 6 weeks before classification. The 6-week threshold is still acknowledged as important but no longer required, and many patients, particularly those with poor prognostic features, will fulfill criteria before 6 weeks. Importantly, the presence of anticitrullinated protein antibodies (ACPAs), particularly higher-titer antibodies, is weighed heavily in the new criteria (two points). Of note in the previously mentioned PROMPT trial, the benefit of early MTX was seen only in patients who were ACPA positive. Most, if not all, of these patients in PROMPT would fulfill new criteria for classification as RA, so PROMPT can be thought of as a trial to test the very early treatment of RA versus the delayed approach.

The First DMARD

Accepting that DMARD therapy should be started as early as possible, which DMARD should be started? And should we begin with mono or combination DMARD therapies? Although many of the previously cited studies have shown that combinations outperform monotherapy in randomized controlled trials, this does not mean that initial combination therapy should be the standard approach for all patients in the clinic. Most clinicians initially start most patients on mono-DMARD therapy; the evidence to support this is discussed later and is validated by the recent findings of the Treatment of Early Aggressive Rheumatoid (TEAR) trial. The decision of which DMARD to initiate at the individual patient level is complex, and at this time there is clearly no one right answer for all patients and all clinical situations (see Figure 71-3 or 71-4A)—one size, or in this case one drug, clearly does not fit all. Many factors need to be considered including, but not limited to, the patient’s disease activity, comorbidities and preference, the relative expense to the patient and to the health care system (weighing the benefits and including direct and indirect costs in our thinking) and importantly where relevant, the patient’s desire (both female and male) to conceive. Until or unless parameters that allow selection on the basis of data become available, this complicated decision will still require the best of the clinician’s judgment.

With all this said, MTX should be the initial DMARD for the majority of patients. MTX is inexpensive (see Table 71-4), effective, well-tolerated, and, importantly, is the cornerstone of most successful combination therapies (see Figure 71-2). In particular, excellent data exist that anti-TNF therapy with all the currently available agents is much more effective with MTX on board both in terms of clinical and radiographic outcomes (see Chapter 61 for specific details on MTX). MTX is usually administered orally at first, although subcutaneous dosing has more predictable bioavailability. In most cases, the dose should be pushed to a minimum of 20 to 25 mg/week if necessary to control disease unless there are contraindications or tolerance problems. Most studies have shown that maximum efficacy may take up to 6 months to achieve but that in most situations, the response at 3 months predicts ultimate success. If MTX is used in this way, approximately 50% of patients will have a good response, and according to consistent data from several trials, approximately 30% will achieve low disease activity status.

Initiating Treatment with a Single DMARD versus Combinations of DMARDs

Although most clinicians still favor starting with monotherapy, combination DMARD therapy has clearly changed the treatment of RA forever. In the early 1990s, the treatment model called for individual DMARDs and then switching to a different DMARD when necessary. At that time, combinations of DMARDs were not used. Studies published in the mid-1990s showing the impressive efficacy and tolerability of various combinations of DMARDs dramatically changed the treatment paradigm. Now the majority of RA patients are treated with combinations of two, three, or more DMARDs. The first study that put combination DMARD therapy on the map was the so-called triple therapy study by the RAIN group of investigators. This study clearly demonstrated that the triple combination of MTX, HCQ, and SSZ was significantly more effective than MTX alone or the combination of HCQ and SSZ (Figure 71-5). Importantly, this combination of DMARDs did not lead to any increase in toxicity. Multiple publica-
tions on other successful combinations of conventional DMARDs soon followed.\textsuperscript{20,31,44,67,68,76}

Today, the important question of whether all patients should be started on combinations of DMARDs and then stepped-down or whether started on monotherapy initially and then stepped-up only if patients are not at target is an ongoing debate. Both approaches have their supporters, and data can be presented on both sides of this question. On the one hand, there is no doubt that if short-term responses are looked at in groups of patients, combinations outperform monotherapy. This is true for combinations of conventional DMARDs,\textsuperscript{71-6,67,68} as well as combinations of conventional DMARDs with a biologic.\textsuperscript{70} On the other hand, the strategy of initial monotherapy with step-ups only in patients who need it was clearly effective in TICORA (as discussed earlier\textsuperscript{18}) and importantly in the BeSt trial\textsuperscript{19,20} and TEAR trials (discussed later\textsuperscript{21}), and although initial responses were better in both of the latter trials in the clinical parameters of patients initially treated with combinations, all groups in both BeSt and TEAR had identical DAS or DAS28 scores at the end of 2 years. Clinicians do not treat groups of patients but individual patients, and if the results are similar at 2 years or beyond, effective treatment approaches that minimize the number of DMARDs and therefore their potential toxicities and certain expense will be desirable for patients and health care systems alike.

**BeSt (Dutch Acronym for Behandel-Strategieen, “Treatment Strategies”) Study**

BeSt continues to be an important randomized, multicenter trial in which 508 early RA patients were randomized to receive one of four treatment “strategies” in an open-label study.\textsuperscript{19,20,77} The four arms of the study were as follows:

- **Group 1**: Sequential DMARD monotherapy; initial MTX 15 mg/week \(\rightarrow\) MTX 25 to 30 mg/week \(\rightarrow\) SSZ \(\rightarrow\) LeF \(\rightarrow\) MTX + infliximab \(\rightarrow\) etc.
- **Group 2**: Step-up combination therapy; initial MTX 15 mg/week \(\rightarrow\) MTX 25 to 30 mg/week \(\rightarrow\) MTX + SSZ \(\rightarrow\) MTX + SSZ + HCQ \(\rightarrow\) etc. \(\rightarrow\) MTX + prednisone \(\rightarrow\) MTX + infliximab \(\rightarrow\) etc.
- **Group 3**: Initial combination therapy; initial MTX 7.5 mg/week + SSZ 2000 mg/day \(\rightarrow\) prednisone 60 mg/day \(\rightarrow\) tapered to 7.5 mg/day by 7 weeks \(\rightarrow\) MTX 25 to 30 mg/week \(\rightarrow\) MTX + prednisone \(\rightarrow\) MTX + CSA + prednisone \(\rightarrow\) MTX + infliximab \(\rightarrow\) etc.
- **Group 4**: Initial MTX and infliximab; initial MTX 25 to 30 mg/week + infliximab 3 mg/kg \(\rightarrow\) 6 mg/kg every 8 weeks \(\rightarrow\) infliximab 7.5 \(\rightarrow\) infliximab 10 mg/kg every 8 weeks \(\rightarrow\) etc.

Treatment adjustments were made every 3 months with the target of a DAS less than or equal to 2.4 (low disease activity). The DAS is calculated using four variables: Richie Articular Index (66 tender joint count [RAI]), swollen joint count (44 joints [SJ]), the ESR (mm/hour), and the global health assessment score (0 to 100; [GH]); DAS = .53938 \(\times\) \([\text{RAI} + .06465 \times (\text{SJ}) + .33 \times \ln(\text{ESR}) + .00722 \times (\text{GH})]\).

The major results at 1 and 2 years are shown in Figure 71-6—both combination groups (Groups 3 and 4) improved quicker than Groups 1 and 2, which was expected with high-dose prednisone in Group 3 and high-initial-dose MTX and also infliximab in Group 4. At 1 year, DAS and other clinical outcomes were similar in all four groups and, importantly, at 2 years they were identical. Health assessment questionnaire (HAQ) scores improved more in the early combination groups (Groups 3 and 4) at 1 year but were not different at 2 years, and radiographic progression at 2 years was greater in Group 1 than in Group 2 and greater in Groups 1 and 2 than in the combination groups (mean modified Sharp van der Heijde Score progression of 9, 5.2, 2.6, and 2.5, respectively). Progression of the joint space narrowing score (importance discussed later), although numerically higher in Group 1, was not statistically different among the four groups (4.3, 2.1, 1.5, 1.2, respectively). Interpretation of these differences is complicated by the fact that despite only 6 months of disease, Group 2 had more radiographic progression at baseline.

**Conclusions from BeSt.** Treat all patients to a target, and they will all do well, although on many different therapies. It is not so important what patients are on, only that they are at low disease activity or below. Several important caveats apply:

1. The strategy of adding DMARDs as in Group 2 was more effective than the strategy of switching from one DMARD to another, at least for the DMARDs and their order of use in BeSt. Although the clinical outcomes were similar at 2 years for these groups, radiographic progression was greater in Group 1 (mean \(9 \times 5.2\)), more Group 1 patients required infliximab (26\% vs. 6\%), and many Group 1 patients ended up on combinations anyway.

2. Initial therapy with combinations of conventional DMARDs (Group 3) or combinations including biologics (Group 4) work more quickly than step-up therapy (Group 2). However, at 2 years, the clinical results were identical. There was a statistical advantage of the combination groups over the step-up group in terms of total radiographic progression (\(\Delta 2.5\) points over 2 years); there was no difference in terms of joint space narrowing. This difference in radiographic progression is of no clinical significance unless it continues to grow at similar rates for years. More strategies need to be employed to further change therapies on the basis of radiographic progress in the small group (perhaps 10\%) of patients on only conventional therapy where it is relevant.

3. A subset of patients was able to discontinue drug therapy for a period of time. In BeSt, if a DAS score of less than 1.6 was achieved for 6 months, patients were tapered off all medications—this occurred in 115 (23\%) of patients. Although many patients relapsed, 59 patients (11.6\%) were in remission off all treatment (median follow-up 23 months\textsuperscript{77}).

**Treatment of Early Aggressive Rheumatoid (TEAR) Trial**

The TEAR trial was a landmark study of initial DMARD therapy in patients with early (mean disease duration \(= 3.6\) months), poor-prognosis (all RF positive, CCP positive, or erosive) RA.\textsuperscript{71} It is the largest (\(n = 755\)) investigator-initiated, randomized, double-blind trial in RA to date. TEAR was a 2-year study and sought to address two critical questions in this early RA population:
1. Should patients be initiated on combination therapy or stepped-up to combinations only after a trial of MTX monotherapy?
2. Is combination therapy with MTX-ETAN superior to therapy with MTX-SSZ-HCQ (triple) therapy?

The major clinical findings are nicely illustrated in Figure 71-8. All four groups had excellent improvement—mean change in DAS28 equals 2.8. For the primary end point of the study (mean DAS28 between weeks 48 and 102), all four groups were identical. Patients started on combination therapy (either triple or MTX-ETAN) were better than those started on MTX alone at the 6-month point. However, this difference disappeared within 12 weeks of the step-up point and the step-up group’s DAS28s were identical to the other groups for the remainder of the study (see Figure 71-8). Data of functional assessments and quality of life measures closely mirrored the DAS28 results. Radiographics

Figure 71-6 A-E. Two-year results of the BeSt study. The error bars indicate 95% confidence intervals. ACR20 indicates American College of Rheumatology 20% composite improvement. DAS, disease activity score; HAQ, health assessment questionnaire; SHS, modified Sharp van der Heijde Score. (Modified from Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, et al: Comparison of treatment strategies in early rheumatoid arthritis: a randomized trial, Ann Intern Med 146:406–415, 2007.)
were done at initiation, at 48 weeks, and at 102 weeks. Radiographic data are presented in Table 71-6, and the cumulative probability plot is shown in Figure 71-9. As shown nicely, the radiographic outcomes of all four groups were superimposable. Table 71-6 shows that there was no difference between the four groups, but if you combine all the ETAN groups and compare them with all the triple therapy groups, the ETAN group had less total modified Sharp score (TSS) progression than the triple therapy group—ΔTSS of 0.51/year (P = .047). This small statistical difference was not clinically significant in the 2 years of the trial (see later discussion about interpreting radiographic changes in clinical trials). Importantly, extensive data on side effects revealed no differences between the two groups. Many had speculated that serious adverse advents may be more common in the ETAN-treated group, whereas minor toxicities such as gastrointestinal upset and others would be more common in the triple-treated patients—neither was seen in this large, blind, clinical trial.

**Conclusions from the TEAR Trial**

1. Strategies that initially treat with MTX and step-up to combination therapy only if patients have not achieved target at 6 months are as effective as initial combination therapy in disease control at 1 and 2 years.
with no difference in clinical or radiographic outcomes. At the core of the debate about initial therapy, single DMARD versus combination, is the question of how soon is soon enough when considering control of RA? Is it important to control synovitis and radiographic progression in days or weeks or is control by 3 to 6 months adequate? No clear long-term data exist to answer this important question. In both the TEAR trial and the BeSt study as discussed earlier, the groups treated with combination therapy early did better quicker, but in both trials the step-up groups looked identical to the early combination groups at 2 years in terms of clinical parameters. However, in both studies the early combination groups had small but statistically significantly less radiographic progression at 2 years. In TEAR, this difference was .5 TSS per year and in BeSt it was 1.3 per year.90 Therefore central to this issue is the clinical relevance of small degrees of radiographic progression, which is discussed in detail later.

2. The strategy of using triple therapy was as effective as the combination of MTX-ETAN. This is clearly an important observation and should allow clinicians to be comfortable with this conventional therapy combination. Importantly, and somewhat sobering, the mean DAS28s of the four groups between 1 and 2 years of the trial was around 3.0. This means that almost half the patients were not at target (DAS28 < 3.2) regardless of treatment assignment. Therefore in clinical situations, usually after 6 months of therapy these patients should be switched to other therapies to obtain better control of disease. For example, patients on triple therapy are switched to or have ETAN added or MTX-ETAN patients have HCQ and SSZ added. No trial data exist on patients who have failed triple therapy and then are treated with biologics or vice versa.

Figure 71-9  Percentage of patients with radiographic improvement or progression during the 102 weeks of the TEAR trial71 by treatment group. Measured by total Sharp score. Marks above the line indicate the amount of radiographic progression, while marks below the line indicate improvement. Thus, in this study, less than 10% of patients in any of the four groups had radiographic progression.

**Treatment of the Patient with Active Disease despite Methotrexate**

As discussed earlier, excellent data from clinical trials where patients in general have high disease activity and poor prognostic features show that approximately 30% of patients will achieve a state of low disease activity or better if treated with MTX alone. This figure is consistent across three major trials that documented this response.19,20,71,74 Of the 70% that have not achieved this level of control, experience suggests that perhaps a third have improved significantly and either the patient or the physician is not willing to escalate therapy. This still leaves approximately 50% of patients who will need something different than MTX monotherapy. The first question to ask is, “Should patients switch to another DMARD or should other DMARDs be added to MTX?” Data from multiple sources including BeSt, as discussed earlier, strongly support adding conventional DMARDs or biologics to MTX over switching to another DMARD.19,20,78 A recent trial has nicely addressed the question of stopping MTX versus continuing MTX in patients with active disease who are started on ETAN. This trial randomized 151 patients to either stopping MTX or continuing MTX while starting ETAN. The group that continued MTX fared much better with ACR20s of 86% versus 64% and experienced significantly less radiographic progression as well.78 Therefore central to this issue is the clinical relevance of small degrees of radiographic progression, which is discussed in detail later.

Excellent data now support the efficacy of 15 different clinically available DMARDs or DMARD combinations when added to MTX in this group of patients (active disease despite MTX). This is true for all of the commonly used DMARDs, conventional and biologics alike. Figure 71-2 highlights the blind clinical trials that have studied currently approved therapies in this patient population.25,36,69,79-88
paucity of data to compare active therapies with each other in this critically important patient population. Knowing that 15 different treatments are better than a placebo is of limited value because most clinicians do not commonly use placebos to treat patients. Therefore trials that compare active therapies are urgently needed.

Two studies shown in Figure 71-2 provide some insight by comparing active therapies to each other in a limited way. In the study by the RAIN group, the combination of SSZ and HCQ appears to be better than either drug alone when added to MTX.62 In the study by Schiff and colleagues,63 abatacept was numerically better than infliximab for efficacy but this difference did not reach statistical significance. This study has been faulted by some because of the low and fixed dose of infliximab (3 mg/kg every 8 weeks). This is a fair criticism with regard to the efficacy comparison, but importantly, even at this “low dose,” serious and opportunistic infections were increased in the infliximab group compared with the abatacept group. Serious infections occurred in 1.9% of the abatacept and 8.5% of the infliximab group, and all five opportunistic infections in the trial including both cases of tuberculosis occurred in the infliximab group.

The really critical decision for clinicians is whether to add conventional therapies to MTX or take the leap and add biologics. To date, only two studies have addressed this and, unfortunately, both in an indirect way and with different conclusions. An open-label study (SWEFOT) in this population of patients with active disease despite MTX compared adding SSZ and HCQ (triple) to adding infliximab.64 At 6 months, both approaches appeared identical with 25% of triple patients and 26% of infliximab and MTX patients responding (EULAR good response); by 1 year, the results were surprisingly different (26% for triple vs. 39% infliximab). It is unclear how these data should be reconciled. One explanation is that infliximab had somehow delayed efficacy with increasing effectiveness between 6 and 12 months. This explanation is problematic because essentially all blind trials have shown that TNF inhibition works quickly and maximum results are achieved by 3 to 6 months.

An alternate explanation would be that the open nature of the trial and/or the ability to switch to other therapies affected the perceived responses. Regardless, at 2 years there again was no difference in the number of patients with EULAR good response or the number in remission between these two different approaches.

The second study to address this question, also in an indirect way, was the TEAR trial by Moreland and colleagues as discussed earlier.13 Although the TEAR trial was designed to look at initial therapy, half of patients initially were treated with MTX alone and are relevant to this discussion. Of these MTX-only patients, 72% or 271 patients did not achieve a DAS28 of less than or equal to 3.2 and were treated with ETAN or the addition of HCQ/SSZ (triple) in a blind fashion. As shown in Figure 71-8, both of these groups looked identical to each other 12 weeks after stepping-up and continued to look identical out to week 102, which was the end of the study. Importantly, with regard to radiographic progression, there was no difference in the radiographic progression of these MTX “failures” in the HCQ/SSZ group versus those in the ETAN group (mean progression of 1.7 vs. 1.1; \( P = .57 \)). Reconciling the findings of the larger blind TEAR trial with the open Swefot trial is difficult, and further data from blind RCTs designed specifically to address this important question are eagerly awaited.

Until further data are available and despite the problems of comparing data across trials, the clinician will have to base the decision of what to add to MTX in this group of patients on the efficacy data presented in Figure 71-2, the economic information presented in Table 71-4, concerns for toxicity in individual patients, and, importantly, patients’ wishes. This question of course begs for research, and appropriate clinical trials to address this question are urgently needed. However, addressing this question will take several trials because so many therapies with multiple different mechanisms of action (MOA) have been shown to work in this clinical situation (see Figure 71-2). It would be terrific to be able to select patients on the basis of parameters that suggested which MOA would be most likely to work. Until other data exist, largely because of their long track record, most clinicians favor a TNF inhibitor as the first biologic in this group of patients.

### Treatment of “Refractory” Patients or Those with Active Disease despite TNF Inhibition

Despite the effectiveness of conventional DMARD combinations and the TNF inhibitors, a subset of patients will continue to have “unacceptable” levels of disease activity. Estimates of the size of this subset range from 10% to 40% of RA patients. The lack of a universal definition of “refractory” and of “unacceptable” levels of disease activity hampers this discussion. With regard to levels of disease activity, it is clear and probably appropriate that clinicians are willing to tolerate more active disease in this “refractory” group before changing therapy, compared with patients who are naïve to treatment or those treated with active disease despite MTX only. Although clear data are necessary, this approach currently seems prudent as the risk of toxicity and expense of treatment become significant factors.

Increasingly, as our expectations for patients increase, we push to control disease sooner and use biologics earlier and patients are being labeled “refractory” much earlier than ever before. When faced with a “refractory” patient, a close inspection of previous treatments “failed” is in order. It is common for patients to be treated initially with low doses of MTX (≤15 mg orally) for 3 months and then started on a TNF inhibitor with or without MTX. If they still have active disease after a few months, patients get labeled inappropriately as “refractory.” It is critical to assess whether MTX therapy has been maximized (see Chapter 61); unless contraindications are present, most patients should have MTX pushed to 25 mg/week and subcutaneous administration should be considered. Data show that triple therapy is as effective as MTX- ETAN,71 so this combination should be considered when resources are limited before progressing too far down the biologic road.

For patients who are truly refractory, as with the previous section on the patient with active disease despite MTX, we urgently need markers or factors that inform us about the best of many possible approaches. The good news is that we still have multiple options with supportive data even in this difficult patient population, which include switching to...
another TNF agent,89-92 starting rituximab,93 starting abatacept94 or starting tocilizumab.95 Currently there are limited data (discussed later) that direct this choice, so until markers or parameters are available, we will make this decision largely empirically (see Figure 71-3).

TNF inhibition has been around the longest, and because of its efficacy and clinicians’ comfort level, many would switch to a second TNF inhibitor before switching to a biologic with a different MOA. A number of observational studies and one RCT support this approach.89-92 From the observational studies, a consistent theme has been that if the first TNF inhibitor was discontinued for toxicity or loss of efficacy, patients have a better chance of efficacy of the second TNF than if the first was discontinued because of primary inefficacy (hazard ratio [HR], 2.7; confidence interval [CI], 2.1 to 3.493). Discontinuation rates, particularly as measured in observational trials, are not the same as efficacy and reflect many things including other treatment options that are available at the time of the study. In the previously cited study published in 2007, golimumab, certolizumab, and tocilizumab were not options.

A single randomized trial has addressed this question of switching to a second TNF inhibitor96; golimumab was given after failure of at least one anti-TNF agent, and 43% responded compared with 17% in the placebo-treated group. A special case in this patient population may be patients who have developed antidrug antibodies. Recently, data have been published with regard to antidrug antibodies in patients treated with adalimumab. After 3 years of treatment, 28% of patients developed antidrug antibodies; in 67% of cases, these developed in the first 6 months.97 Development of antidrug antibodies was associated with lack of or loss of efficacy (HR, 3.0; 95% CI, 1.6 to 5.5), and patients were less likely to achieve remission (HR, 7.1; 95% CI, 2.1 to 23.4). Importantly, 38% of patients with antidrug antibodies compared with 14% of those without discontinued therapy for lack of efficacy (P = .001). Perhaps of more concern was a recent finding that the development of anti-adalimumab antibodies was associated with thromboembolic events (HR, 7.6; 95% CI, 1.3 to 45.1; P = .2597). These interesting findings raise the possibility that monitoring patients for development of antidrug antibodies not only to adalimumab but also other biologics may be an important strategy to predict not only lack of efficacy but also to prevent toxicities such as thromboembolic events.

Once a clinician decides to go with a biologic with a different MOA, there are currently three choices (although this will likely increase in coming years): rituximab, abatacept, or tocilizumab. Data to help differentiate among them are scarce. On the basis of strong suggestions in the literature that patients who are seronegative for CCP and RF do less well than seropositive patients when treated with rituximab,98,99 most clinicians would opt for either abatacept or tocilizumab in the seronegative patient. Randomized trials have shown the efficacy of all three MOAs in this patient population.93-95 These studies have been similar in design and also in the results of intervention. Patients on MTX with active disease or intolerance to at least one TNF inhibitor have had the TNF inhibitor stopped and are randomized to receive the intervention + MTX or MTX + placebo. Response rates in the three different trials for ACR20 for the active drug versus placebo were rituximab 51% versus 18%, abatacept 50% versus 20%, and tocilizumab 50% versus 10%.101 All these responses, however, are less robust than in biologic-naïve patients. On the basis of the strikingly similar response rates of 50% across these three trials, there is little to choose from until further predictors of responses or comparison trial data are available. Some have raised concern about treatment with a second biologic after treatment with rituximab, which depresses B cell numbers significantly for at least 6 to 12 months. Limited data from observational studies provide some reassurance as a significant increase in toxicities has not been reported.102

What to Do with the Patient in Remission (on DMARDs)

It is a powerful testament of how far therapy for patients with RA has come that it is appropriate to include a new section in this edition on what to do with the patient in remission. It is now common practice for clinicians to ask how to taper DMARDs in patients who are “in remission.” Indeed, a recent ACR expert panel acknowledged the importance of this question and included it as one of the top priorities for future investigation.103 Putting aside the obvious difficulty in defining “remission,” be it on clinical grounds, with radiographic evidence or by using advanced imaging, clinicians see an increasing number of patients in “remission” or with low levels of disease activity and are increasingly seeking direction on whom should be tapered off of what medications and when.

Unfortunately, again, little data exist. We do not have laboratory tests, inflammatory parameters, or cytokine profiles that help predict who can be safely tapered. Because of their well-chronicled toxicities, the highest priority should be to get all patients off glucocorticoids and NSAIDs. A number of trials19,20,31 have shown that with effective DMARD therapy, glucocorticoid tapering is not only possible but should be expected. Perhaps the best example of this is from the BeSt trial in which 92% of Group 3 (the group that got high-dose prednisone up front) was off all prednisone at 2 years.20 In this trial, prednisone was tapered only if the DAS was at or below 2.4 (low level of disease activity).

Once the patient is off glucocorticoids and perhaps on NSAIDs only as needed but still in remission, it is more difficult to know which drug should be tapered next. A common situation is the patient on combination therapy with MTX and a TNF inhibitor. Because of the expense and concerns for long-term toxicities, it would make sense to taper the TNF inhibitor down to the lowest dose possible. Many of these patients in practice may have been started on this combination initially or had the TNF inhibitor added after only a short course of MTX. These patients may be similar to those in the TEAR trial who were started on the combination of MTX-ETAN.71 For this group of patients we can infer that approximately 30% of them did not need the ETAN and could have done extremely well on MTX alone just as the 28% of patients who started on MTX alone and did not need to step-up. The MTX-only patients who did not step-up not only had the lowest DAS28 at 2 years (mean, 2.7) but also had the least amount of radiographic progression.71,102 Unfortunately, we do not know how to select these patients.
If patients have been in remission for 6 months to a year and are being treated with a biologic, it seems prudent to decrease the dose or lengthen the interval between injections. Although most biologics, certainly the subcutaneously administered ones, are dosed in a fixed, one-size-fits-all approach, we know from some of the earlier trials that lower doses than usually clinically prescribed can be effective. As an example, the ERA trial compared doses of ETAN of 10 mg and 25 mg given twice weekly\(^1\); although the data were slightly better for the higher dose, the low dose clearly had substantial efficacy nearly equal to “full-dose” ETAN, both as measured by clinical (ACR20) and lack of radiographic progression. Therefore a significant percentage of patients may be overdosed, particularly those in remission. In our clinic it is not uncommon to have patients who are in remission taking their 50 mg of ETAN every 2 to 4 weeks or their 40 mg of adalimumab every 3 to 6 weeks.

Although wondering what to do with patients in remission on DMARDs is clearly a wonderful problem to have, further direction is certainly necessary for this emerging and fortunately increasing patient group. Trials to address this question with the collection of biomarkers will be increasingly important as patients continue to do better with current therapies. It is clear that both US and MRI (discussed later) can detect levels of synovitis that are not apparent on clinical examination. As this question is examined further, perhaps we will learn that patients with inflammation by US are not good candidates for discontinuing therapy or, alternatively, US or MRI may provide early warning signs of flares. An interesting recent report indicated that US was able to predict which patients in clinical remission would flare.\(^2\)

**Use of Combinations of Biologics**

Combinations of conventional DMARDs and combinations of conventional DMARDs (especially MTX) with biologics have played a critical role in the vast improvement in responses to treatments we currently see. However, so far efforts to combine biologic products have not met with success. To date, studies to combine ETAN with anakinra have shown no improvement in disease control with ACR50 of 41% in the ETAN-only arm compared with 31% in the combination arm and have shown an increase in serious infections in the combination arm (7.4% vs. 0%).\(^3\)

Similarly, in patients with active disease despite ETAN (\(n = 80\)), patients were randomized to add abatacept or placebo; efficacy showed a nonstatistically significantly trend toward improvement (ACR20 for combination, 48% vs. 31%; \(P = .07\)), but again an increased risk of serious adverse events in the combination group (17% vs. 3%) and serious infections (4% vs. 0%).\(^4\) Finally, a recent small trial (\(n = 51\)) in which patients with active disease despite MTX and either ETAN or adalimumab were treated with the addition of rituximab or placebo showed modest improvement in efficacy (ACR20 of 30% vs. 17%), and again serious adverse events including infections were numerically greater in the combination-treated group.\(^5\)

Despite these less than encouraging results from early biologic combination trials, it is likely that ultimately patients will be treated with combinations of biologics. What is critically needed are good ways to measure the degree of immune suppression we create with our interventions—a thermostat for TNF, IL-1, or IL-6, if you will.\(^6\) Currently, we are likely inadequately suppressing some patients, therefore not achieving optimal disease control and suppressing others to dangerous levels leading to increased toxicity. When better immune system monitoring techniques are developed, we will be able to better use the biologics we currently have. In addition, it is likely that we will be able to safely and effectively combine biologics to improve outcomes both in terms of efficacy and safety for patients.

**Interpreting Radiographic Progression and the Use of Other Imaging Modalities**

What is the clinical relevance of radiographic progression? How much progression is significant and over what time period? Are erosions important or should we be concerned only with joint space narrowing? Finally, what role, if any, should US or MRI play in the management of RA?

In clinical trials, treatments of RA are evaluated by clinical parameters (ACR20 responses, DAS, etc.) and radiographic progression (TSS or SHS). This is problematic for a number of reasons. Clinical and radiographic progression is not always parallel. Maybe the most dramatic example of this from an RCT is the PREMIER trial. In the monotherapy arms, MTX alone was significantly better statistically than adalimumab alone, while at the same time, adalimumab alone was better than MTX alone with regard to radiographic progression (Figure 71-10A and B).\(^7\) So which treatment was better? Some would say it is moot because the combination of the two outperformed either monotherapy arm in both measures. However, that misses the point about how to balance radiographic progression versus clinic parameters. Because patients do not come to a clinic complaining of radiographic progression or demanding that it be stopped, the following question is highly relevant: “How is radiographic progression related to things patients do care about—most importantly, physical function?”

Some data relate total Sharp score (TSS) or, similarly, total (SHS) progression to changes on health assessment questionnaires or HAQs (the gold standard for physical function in RA). These data suggest that a change in TSS of 1 equals a change in HAQ of approximately .01.\(^8\) It is well accepted that a clinically significant change in HAQ is approximately .22, so it follows that a change in TSS of 22 is required for a radiographic change to be clinically relevant. This large change is never seen in clinical trials in which active therapies are compared with each other. So although many RCTs have shown statistically significant radiographic progression differences, few if any have shown clinically significant difference within 1 to 2 years of the clinical trial. Additionally, recent data suggest that joint space narrowing correlates well with clinical progression, whereas erosions do not,\(^9\) so perhaps we should not be concerned with TSS but only the joint space narrowing component. The next critically important point is over what duration should we be concerned about radiographic progression? If a trial shows Therapy A results in two less TSS points progression per year compared with Therapy B, it follows that if those same therapies are used for 11 years,
then patients on Therapy A will have a clinically significant clinical benefit, HAQ .22 less than Therapy B. This is the magnitude of difference (ΔTSS of 1 to 2/year) that has been seen in trials and has led to claims in superiority of one therapy over another. Two problems are readily apparent with this type of extrapolation:

• Within clinical trials patients are often randomized to a therapy and left on it for the duration of the trial regardless of how they are doing. This is not how patients are taken care of in a clinic; if patients are not doing well, therapy is or should be adjusted. Again, the PREMIER trial\textsuperscript{70} is an excellent example of this issue—patients were in their assigned groups for 2 years (see Figure 71-10B). Many were not having optimal clinical response, and therapy would have been changed if similar things were happening in a real-life clinical setting. These patients would not still be in the trial demonstrating their predictable radiographic progression, so in any RCT, radiographic progression that occurs after the time the therapy should have had maximum clinical benefit, and therefore patients would have been switched to a new therapy, is not clinically relevant. In the PREMIER trial, the therapies should have had maximal benefit by 6 months, so if the data after that time point include patients who were not at target (at least low disease activity), then it is not relevant to patients seen in a clinic.

• The second point is similar—to extrapolate forward in the example of a change in TSS of 2/year to 11 years of therapy would only be relevant if all patients were doing well clinically. Patients who are not would have their therapies changed before the two TSS progression per year had a chance to become clinically relevant.

Another major reason why radiographic progression is problematic for the clinician is that formal evaluation of radiographic progression is almost never done outside of clinical trials; few, if any, patients in many countries have formal Sharp score assessment of radiographs. In this regard, routine yearly radiographs in RA are not indicated in many RA patients and only add to the expense of care. This is true for patients who are not doing well clinically—those patients need a change of care regardless of what the radiographs show, and it is also true for patients who are unwilling or unable to change therapies. The small minority of patients who are at target but who are progressing radiographically are the ones we hope to discover important information with serial radiographs. Finally, the way radiographic information is presented in clinical trials is problematic. Most commonly, the mean TSS progression of Group A is compared with Group B and statistics are done. We are told that Therapy A is superior to Therapy B because there is one or two less TSS progression in that group. A much more informative way to look at radiographic outcomes is to look at cumulative probability plots for radiographic changes (Figure 71-11 for the TEMPO trial\textsuperscript{72,73} and Figure 71-9 earlier for the TEAR trial\textsuperscript{71}). When data are looked at this way, it is readily appreciated that a small number of patients, perhaps 15% to 20% in the TEMPO trial and in many studies even fewer, are driving

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure71-10}
\caption{Results from the PREMIER trial.\textsuperscript{70} A, Clinical improvement measured by American College of Rheumatology 20\%, 50\%, and 70\% composite improvement. B, Radiographic progression. ADA, adalimumab; MTX, methotrexate. (Modified from Breedveld FC, Weisman MH, Kavanaugh AF, et al: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment, Arthritis Rheum 54:26–37, 2006.)}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure71-11}
\caption{Radiographic probability of progression plot from TEMPO.\textsuperscript{72,73} TEMPO: PROBABILITY OF RADIOGRAPHIC DAMAGE}
\end{figure}
the radiographic outcome. In the TEAR trial, there are no differences across the four groups in the probability plot. With this epiphany, the important questions are who are these patients and how do we focus on this small group that would benefit from a different therapy and not subject the other 80% to 95% to additional risks and expense? Recently, as mentioned earlier, Aletaha and colleagues have called into question the relevance of erosions with regard to clinical correlations and have suggested that joint space narrowing correlates much more strongly with clinical progression. Therefore going forward we need to take this into account.

Finally, what is the place, if any, of “advanced” imaging techniques in evaluating RA patients? Clearly, both US and MRI are powerful techniques to evaluate early erosions and to detect synovitis (see Chapter 58). One potential use of these obviously sensitive techniques, as suggested earlier, is to verify the true absence of synovitis in patients who are felt to be in remission on clinical grounds. Until further research is available, these techniques will remain largely research tools with regard to routine care of RA patients.

**ADJUNCTS TO MEDICATIONS**

**Patient Education**

RA is a lifelong disease, so it only makes sense that patients who are educated about their disease do better, and in fact good data exist to show that education leads to better outcomes. Use of the Arthritis Self-Management Program has been shown to lead to less pain, decreased visits to physicians, and economic savings. Chronic diseases such as RA affect the whole family, and participation of spouses in educational group sessions leads to additional beneficial effects. Training in stress management is particularly effective in improving measures of helplessness, coping, pain, self-efficacy, and health status. Unfortunately, formal education level, a factor that we cannot modify, is associated with morbidity and mortality rates. Formal educational level is inversely proportional to these outcomes, and this cannot be explained by age, duration of disease, joint count, functional measures, or medications.

Chapter 67 provides an in-depth discussion of many of these important aspects. It is clearly important for patients to take an active role in the management of their chronic disease. The more patients understand their disease and medications, the more control they feel they have over the whole situation. The bond that develops between the patient and physician working together to control this lifelong disease is an important factor in outcome, as well as the satisfaction level of both parties. The stronger the bond is, the less likely that patients will become frustrated and turn to alternative therapies that are not only expensive but may do more harm than good.

**Pain Control**

If patients with RA are treated early and effectively with DMARDs and therapy is escalated to achieve excellent control of the active components of disease, the need for specific pain medications, particularly narcotics, can be minimized. If pain is a major problem, the clinician should first review the DMARD program and modify it to achieve maximum control of any active synovitis. Unfortunately, some patients will present in the latter course of disease after substantial joint damage has already occurred and will need pain relief. Pain can be the factor that limits effectiveness of physical and occupational therapy, and as pointed out during a special workshop sponsored by the National Advisory Board for Arthritis and Musculoskeletal and Skin Diseases, it is frequently undertreated in patients with arthritis. In addition to inhibiting function, pain is a major cause of depression in patients with polyarthritis. To maximize therapy in patients with early RA or undifferentiated polyarthritis, pain must be controlled without altering consciousness or generating addiction. Treatment strategies favoring education, rest, exercise, and disease-modifying therapies are generally favored as an approach to pain control in arthritis, and strategies that rely on narcotic derivatives may not address control of active RA. In most medical centers, experts in pain management are available for consultation by rheumatologists and primary care physicians (PCPs). There is an excellent discussion of analgesics in rheumatic disease in Chapter 66.

**Rest, Exercise, and Activities of Daily Living**

Education and supervision of a patient by trained professionals on the importance of finding the best balance of rest and exercise for inflamed joints are essential. This component of therapy can be started well before a definitive diagnosis is made. No matter what the cause, finding this balance should ensure that a patient develops or retains sufficient strength to support joint function without exacerbating inflammation.

Details of physical and occupational therapy are outlined in Chapter 38. A patient with acutely and severely inflamed joints may not be the ideal exercise candidate and may need application of resting splints to immobilize the joint until anti-inflammatory medications, specifically DMARDs, take effect. Even the most painful joints, when splinted, must be moved passively through a full range of motion each day to prevent flexion contractures, particularly in children. For moderately inflamed joints, isometric exercise with muscles contracted in a fixed position (the resting length of the muscle) provides adequate muscle tone without exacerbating joint inflammation and pain. Maximal contractions, held for 6 seconds and repeated 5 to 10 times, performed several times each day, can prevent further loss of muscle mass around arthritic joints.

Patients with well-controlled arthritis will benefit from variable-resistance programs or high-intensity strength training, which has been shown to provide significant improvements in strength, pain, and fatigue levels. Older patients with RA benefit from progressive resistance exercises, similar to younger patients. In a study of older patients given closely regulated workouts on pneumatic resistance equipment, maximal strength of all major exercised muscle groups was increased 75% without exacerbation of clinical disease activity. Prescribed sustained exercise not only increases muscle strength but also helps the ability of patients to perform daily routines, improves global assessments and moods, and can decrease pain.
Most patients with RA should have one or more sessions with a licensed occupational therapist (OT) to learn how to preserve joint function and alignment while carrying out the necessary and enjoyable activities of daily living and to be exposed to the assistive devices that are available. The basic concept is to avoid excessive force applied across non-weight-bearing joints and to avoid unnecessary impact loading on weight-bearing joints. A prospective and controlled Canadian trial demonstrated that home therapy by OTs produced a statistically significant and clinically important improvement in function in RA patients.121

Treatment of Comorbidities and Interaction of Rheumatologist with Primary Care Physician

The best possible outcomes for patients with RA can be achieved only with a carefully orchestrated collaboration between PCPs and rheumatologists. On the one hand, the ever-increasing complexity of RA management options, combination therapies, and possible toxicities of therapy have all made it essential that RA patients are seen by rheumatologists. Good evidence indicates that RA patients are more likely to be on DMARDs, are more likely to be on combination DMARDs, and are happier with their care when seen by rheumatologists.122,123 On the other hand, the realization of the critical nature of the comorbidities associated with RA, especially cardiovascular disease, makes ongoing engagement of a PCP essential124 to produce optimal outcomes.

Significant data indicate that patients with RA have better outcomes when rheumatologists are the primary manager of the RA. In one series of 561 patients with definite RA followed over 20 years, the patients seen by a rheumatologist during the first 2 years of disease fared significantly better than those who did not.125 The favorable outcome could be related to an early start of DMARD therapy. Further, data indicate that regular ongoing care by a rheumatologist (mean of 8.6 visits per year) results in less functional disability than for patients who received only intermittent, sporadic care by a rheumatologist.126 Analysis of these data supported the interpretation that worsening disability was not the reason for intermittent care, but rather a consequence of it. Additionally, evidence supporting care by a rheumatologist showed that patients who had access to a rheumatologist had higher performance scores than patients who saw only a PCP. Access to specialist care resulted in significant improvements in arthritis care, comorbid illness, and health care maintenance overall beyond seeing a PCP alone.128

As important as the rheumatologist is, the PCP needs to be closely engaged (Table 71-7). Recently, the comorbidities of RA have been highlighted and include premature atherosclerosis and congestive heart failure, increased risks of osteoporosis and fractures, and increased risks of infections. A recent review by a panel of cardiologists and rheumatologists highlighted RA as a significant risk factor for cardiovascular morbidity and mortality on par with diabetes.129 In this regard, the PCP needs to aggressively address traditional cardiovascular risk factors such as hypertension and especially hyperlipidemia in RA patients. Statins should be used aggressively because the primary reason for the excess mortality seen in RA is cardiovascular. As an added bonus, statins, probably secondary to their anti-inflammatory effects, have been shown in animal studies of RA models130 and in at least one RCT in humans131 to decrease RA activity. Further, there are data that statin use may protect against the development of RA.132

Because RA and particularly glucocorticoids are risk factors for osteoporosis, the PCP, in concert with the rheumatologist, needs to address this. Most patients should be receiving calcium and adequate vitamin D3. Bisphosphonates have been shown to prevent steroid-induced osteoporosis,133 and their use should be strongly considered in patients on long-term steroids unless there are contraindications (e.g., women of childbearing age).

With the increased infection risk that RA patients have, both because of their disease and therapies, up-to-date immunizations including yearly influenza shots, every 5-year pneumococcal vaccinations, and appropriate zoster vaccinations are critical.1 The latter is a live-virus vaccination and should be avoided in patients currently receiving biologics. Current recommendations are to give the zoster vaccination 2 weeks before starting any of the biologics. Influenza and pneumococcal vaccinations can be safely given to patients on biologics with reasonable immunologic response,134 except in patients on rituximab, in whom response is severely impaired.135,136 The rheumatologist and PCP should push hard for smoking cessation for all the usual reasons plus concern for cardiovascular and pulmonary disease already over-represented in RA; cessation has the added benefit of potentially making patients more responsive to therapies.137 Finally, because the PCP often sees the patient first during an illness, the PCP needs to be familiar with the common toxicities of RA medications including concerns about MTX in the patient with decreasing renal function, MTX pneumonitis, and the need for heightened concern and aggressive workup of all potential infections including opportunistic infections in immunosuppressed patients, particularly those on biologics. Often, the difference between life and death or at least between excellent outcomes and less than optimal results is the rapidity of response to the early warning signs for things as simple as a cellulitis or a pulmonary infiltrate.

Clearly the best care is given by a team: a rheumatologist, a PCP, and an educated patient all working closely together. Timely consultation with physical and occupational therapists and orthopedic surgeons also plays a role in optimal outcomes for RA patients. Although the need for joint replacement surgery is thankfully decreasing, few interventions have been as successful in improving patient mobility and quality of life as having properly timed hip and knee replacements. With the potential of huge changes in health care delivery on the horizon, at least in the United States,

Table 71-7  Role of Primary Care Physicians in Rheumatoid Arthritis

| Monitor and aggressively treat cardiovascular risk factors |
| Monitor and treat/prevent osteoporosis* |
| Recognize toxicities of rheumatoid arthritis medications and initiate appropriate and timely workup* |
| Recognize the risks for infections and ensure immunization status is current |

*In partnership with a rheumatologist.
it may be an increasing challenge to assure that the right team is always engaged to achieve the best outcomes for each patient.

Evidence That Patients with Rheumatoid Arthritis Are Doing Better

The prognosis for patients with RA has improved dramatically. Every rheumatologist who has been fortunate enough to see the changes over the past quarter of a century is only too happy to witness to this. The days when wheelchairs were commonplace in clinics and many RA patients had symptomatic C1 subluxations, chronic leg ulcers, constrictive pericarditis, and corneal melts are hopefully gone forever. Despite this firmly held perception by clinicians, strong data to support this change have been slow to accumulate.

Recent data from Olmsted County, Minnesota, show that RA patients diagnosed after 1995 lived significantly longer, almost 9 years, than those diagnosed before 1995. In the same cohort of patients in Olmsted County, knee surgery was decreased by 46% and hand surgery by 55%. Data from more than 35,000 U.S. veterans with RA showed a greater than 30% reduction in extra-articular manifestations of RA since the year 2000. Ward, using data from California (years 1983 to 2001), showed that hospitalizations for RA vasculitis or splenectomy for Felty’s syndrome decreased by 33% and 71%, respectively. Further, primary knee arthroplasty for RA decreased by 10%. Early unpublished reports from many centers indicate that joint replacement surgery for RA may be reduced by up to 50% to 80%. Data from Sweden and Spain both show that disease activity scores and health assessment scores have significantly improved over the past decade.

Investigators postulate it may take up to 20 years for the results of a change in therapy or approach to fully translate into improvement in long-term outcomes such as joint replacements and mortality. Therefore the encouraging data presented earlier likely reflect our therapies of the early to mid-1990s, and therefore we expect that these early reports were commonplace in clinics and many RA patients had symptomatic C1 subluxations, chronic leg ulcers, constrictive pericarditis, and corneal melts are hopefully gone forever. Despite this firmly held perception by clinicians, strong data to support this change have been slow to accumulate.

Horizon

It is always difficult to predict the future, particularly in an area where things are changing as rapidly as they are in RA, and certainly the authors of the early editions of TOR could not have predicted we would make the giant strides in RA treatment that we have. As Bill Gates remarked, “We always overestimate the change that will occur in the next 2 years and underestimate the change that will occur in the next 10.” By the tenth edition of TOR we will almost certainly have several new biologics and perhaps several new small molecule DMARDs available for treatment. Leading biologic candidates include rilonacept (IL-1 Trap), already approved for cryopyrin-associated periodic syndromes, and potential targets IL-12, IL-17A, IL-23 and IL-33. Prime small molecule DMARDs that inhibit signaling molecules, particular Janus kinases (Jak), and spleen tyrosine kinase (Syk), might become available in the next few years and could dramatically change the treatment paradigm yet again.

Regardless of new therapies that expand our portfolio of choices, the two areas where we will need to see the most progress are profiling patients on the basis of parameters that predict differential responses to different therapies and in monitoring the kinds and intensity of immunologic modification that we are producing. If substantial progress can be made in these areas, even without new therapies, we should be able to obtain vastly superior and more expeditious disease control. On a somewhat more pragmatic note, the treatment of RA has become an expensive endeavor and with all the changes in health care and the continuing increases in the cost, it is imperative that we have the appropriate studies to allow us to take the best care of
patients for the least cost. If we do not have the appropriate research to justify expensive medications, we may have difficulty justifying their use. Therefore it is critical that trials are designed with this in mind and that long-term costs, both direct and indirect, of suboptimally controlled RA be part of the equation.

It is clear from many observations that the immune response that becomes RA starts before most patients present with the classic symptom complex. Studies to elucidate this transition are under way, and studies to treat presymptomatic RA are not far behind. Even further, these studies on preclinical disease hold the promise of elucidating the triggers of RA and, if so, the ultimate goal of RA prevention cannot be far behind.

Selected References


5. van der Heijde DM, van ’t Hof M, van Riel PL, et al: Validity of 28-joint count, Disease Activity Score (DAS28), Health Assessment Questionnaire (HAQ), Multidimensional HAQ (MDHAQ), and Routine Assessment of Patient Index Data (RAPID) scores, J Rheumatology 35:603–609, 2008.


8. Yazici Y, Bergman M, Pincus T: Time to score quantitative rheumatoid arthritis measures: 28-Joint Count, Disease Activity Score, Health Assessment Questionnaire (HAQ), Multidimensional HAQ (MDHAQ), and Routine Assessment of Patient Index Data (RAPID) scores, J Rheumatology 35:603–609, 2008.


11. Yazici Y, Bergman M, Pincus T: Time to score quantitative rheumatoid arthritis measures: 28-Joint Count, Disease Activity Score, Health Assessment Questionnaire (HAQ), Multidimensional HAQ (MDHAQ), and Routine Assessment of Patient Index Data (RAPID) scores, J Rheumatology 35:603–609, 2008.


Full references for this chapter can be found on [www.expertconsult.com](http://www.expertconsult.com).
References


Early Synovitis and Early Undifferentiated Arthritis

ANNETTE H.M. VAN DER HELM-VAN MIL • THOMAS W.J. HUIZINGA

Early synovitis refers to synovitis that is detected by physical examination; the symptom duration defining early has changed over time.

Intervention in arthritis with symptoms less than 12 weeks can be beneficial and can potentially prevent damage later.

Undifferentiated arthritis (UA) is a diagnosis per exclusion. The phenotypic characteristics and outcome change depending on the classification criteria used for rheumatoid arthritis.

Few clinical trials have been performed in UA, and in none of these studies treatment strategies were adjusted to individual patients’ chance on spontaneous remission versus progression to rheumatoid arthritis.

WHAT IS EARLY SYNOVITIS?

Recent data suggest that early therapy and achieving low disease activity improve long-term outcomes in inflammatory arthritis. These observations have increased awareness among rheumatologists of the importance of early diagnosis and treatment. As a result, considerable pathophysiologic research and clinical trials have focused on identifying early synovitis as soon as possible. Stratifying patients into well-characterized diseases, such as rheumatoid arthritis (RA), is essential so that appropriate therapy can be initiated. Equally important, identifying patients with a high chance for spontaneous remission is needed to avoid overtreating individuals with transient undifferentiated synovitis.

Uniform descriptions of patients and accurate diagnostic criteria for early synovitis are essential for proper interpretation of clinical research. At present, no standardized definition for early synovitis is widely accepted. In fact, the definition of early has evolved over time. Previously, studies used a cutoff of less than 5 years to define early disease for RA. By the 1990s, duration of symptoms for less than 12 to 24 months was considered early. This duration was chosen because most RA patients incur significant damage when treated conventionally.

The window for modifying disease outcomes might be narrow, and maximum benefit might require interventions even earlier; some evidence suggests that treatment within months might be beneficial. For this reason, several early arthritis cohorts in recent years strictly restrict the inclusion criterion of symptom duration and include only patients with a symptom onset of less than 12 weeks; this is referred to as very early synovitis. Even this definition probably does not capture the earliest phase of disease, as circulating antibodies might appear years before the onset of symptoms in RA, and biomarkers reflecting bone destruction are elevated before arthritis is present. Animal models show that histologic evidence of inflammation and activation of signaling pathways can occur well before clinically apparent disease.

Although the term synovitis generally refers to a swollen joint detected by physical examination, alternative definitions are also applied. Physical examination might not be sensitive enough, and clinically undetectable synovitis may be present and relevant to identify. Therefore, some studies consider tenderness in the absence of swollen joints as synovitis. For instance, early arthritis clinics might include patients with tenderness but no swollen joints. Others determine synovitis not by physical examination but through the use of imaging modalities such as ultrasound (US) or magnetic resonance imaging (MRI). Different ultrasound techniques are used; of these, traditional B-mode grayscale ultrasonography provides information on synovial thickness, and Doppler imaging measures the synovial blood flow. Ultrasound is easily accessible, rapid, and inexpensive. An important consideration is the reproducibility between readers. Whereas some studies found good or acceptable inter-reader agreement, other studies observed considerable variation between readers. US appears to be valuable, especially in the absence of abnormalities, and the negative predictive value of a normal US result is high. The prognostic implication of abnormal US findings is less clear. More long-term studies are needed to establish which US characteristics have a high positive predictive value for the development of persistent or erosive arthritis.

Until more data become available, the term early synovitis generally refers to synovitis that is detected by physical examination. It is important to note that early synovitis refers to a disease symptom but does not reflect any specific diagnosis. The spectrum of final diagnoses in a population of early synovitis patients is presented in Figure 72-1.

EARLY ARTHRITIS CLINICS

Early arthritis clinics are population-based inception cohorts in which patients with synovitis of early onset are included and followed prospectively. These cohorts often contain a wealth of phenotypic information, and the longitudinal design allows the study of factors related to a beneficial or severe course of the disease. Such studies might increase our understanding of the processes involved in arthritis progression. A summary of existing early arthritis cohorts is presented in Table 72-1. Inclusion criteria have changed over
Figure 72-1 Overview of diagnoses of early arthritis patients. (Presented data are based on 2284 early synovitis patients included in the Leiden Early Arthritis Clinic cohort and their diagnosis after 1 year of follow-up.) RS3PE, remitting seronegative symmetrical synovitis with pitting edema; SLE, systemic lupus erythematosus; UCTD, undifferentiated connective tissue disease.

WHAT IS UNDIFFERENTIATED ARTHRITIS?

The term undifferentiated arthritis (UA) refers to a subpopulation of early synovitis patients who do not meet criteria for other diseases, including infections, spondyloarthropathies, crystal diseases, and RA. Because it is a diagnosis of exclusion, no classification criteria for UA exist.

European early arthritis cohorts observed that 35% to 54% of the patients included do not meet criteria for other diseases and are considered as UA. In the remaining early synovitis patients, a definitive diagnosis could be established at first visits. The frequency of UA is dependent on symptom duration at the time of the first visit to a rheumatologist. The longer symptoms exist, the more likely that sufficient characteristics are evident to support a definitive diagnosis, and the lower the prevalence of UA will be. In addition, the prevalence of UA will change when classification criteria for rheumatologic diseases change. For example, when the 1987 American College of Rheumatology (ACR) criteria for RA were applied to new-presenting early synovitis patients in the context of an early arthritis clinic in Europe (e.g., Leiden, the Netherlands), about 20% of those patients directly fulfilled the 1987 ACR criteria and thus were classified as RA. At present, few studies evaluating the performance of the 2010 ACR/European League Against Rheumatism (EULAR) criteria are yet available. Initial data from the Leiden clinic suggest that this percentage is increased when the 2010 criteria for RA are applied.

Remission Rates in Undifferentiated Arthritis and Rheumatoid Arthritis

The natural disease course of UA is variable, depending on the inclusion criteria and the duration of symptoms in several inception cohorts. Spontaneous remission occurs in 40% to 55% of UA patients. In contrast, remission in RA is at most 10% to 15%. In these studies, remission was defined as the absence of swollen joints for 1 year or longer after discontinuation of eventual disease-modifying antirheumatic drug (DMARD) therapy and/or discontinuation of participation in the outpatient clinic because the patient no longer needs drug treatment.
limited number of joints. For example, the Sharp-van der Heijde method evaluates metacarpophalangeal (MCP), proximal interphalangeal (PIP), wrist, and metatarsophalangeal (MTP) joints but does not evaluate the distal interphalangeal or larger joints. Therefore, this method may be less optimal for measuring the level of joint damage in patients who present with predominant synovitis in large joints. In RA, the extent of joint destruction in small joints adequately reflects joint destruction in larger joints. Such data are missing for UA patients.

A recent study evaluated the subgroup of UA patients who developed RA later in time. Radiologic data were compared with those of patients who at first presentation of the disease directly fulfilled the 1987 ACR criteria for RA. Radiographic progression rates between these groups were not different. Health Assessment Questionnaire (HAQ) scores and Disease Activity Score (DAS) measures also were not different between these two groups. Thus, although UA patients as a group have a higher rate of spontaneous remission compared with patients with RA, the subgroup of UA patients who progress to RA have an

**Table 72-1 Overview of Early Arthritis Cohorts**

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Inclusion Criteria</th>
<th>Year of Onset</th>
<th>Includes Early Arthritis (Symptoms &lt;2 yr)</th>
<th>Includes Very Early Arthritis (Symptoms &lt;3 mo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wichita Arthritis Centre</td>
<td>Undifferentiated polyarthritis syndrome or RA</td>
<td>1973</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Norfolk Arthritis Register</td>
<td>Early inflammatory polyarthritis</td>
<td>1989</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Leiden EAC</td>
<td>Synovitis of ≥1 joint</td>
<td>1993</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Austrian Early Arthritis</td>
<td>Inflammatory arthritis</td>
<td>1995</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Registry</td>
<td>Symptom duration ≤2 wk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amsterdam</td>
<td>Synovitis of ≥2 joints</td>
<td>1995</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Leeds Early Arthritis Centre</td>
<td>Undifferentiated arthritis of the hands</td>
<td>1995</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Birmingham VEAC</td>
<td>Synovitis of at least one joint</td>
<td>2000</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>ESPOIR cohort study</td>
<td>Certain RA and UA that may develop into RA</td>
<td>2002</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Toronto Early Arthritis</td>
<td>Symptom duration between 6 wk and 12 mo</td>
<td>2003</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Cohort</td>
<td>≥2 swollen joints or 1 swollen MCP or PIP joint with one or more of the following:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>positive RF, positive ACPA, morning stiffness &gt;45 min,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>response to NSAIDs, or painful metatarsophalangeal squeeze test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Berlin EAC</td>
<td>≥2 swollen joints</td>
<td>2004</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Norwegian Very Early Arthritis (NOR-VEAC)</td>
<td>Symptom duration between 4 wk and 12 mo</td>
<td>2004</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>New Castle EAC</td>
<td>Synovitis of at least one joint or inflammatory complaints without a synovitis at</td>
<td>2006</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>physical examination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotterdam Early Arthritis</td>
<td>Synovitis ≥1 joint or ≥2 tender joints</td>
<td>2007</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>CoHort (REACH)</td>
<td>Symptom duration ≤1 yr</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The presented cohorts include not only early RA patients but also a broader range of early synovitis patients.

ACPA, anticitrullinated protein antibody; MCP, metacarpophalangeal; NSAIDs, nonsteroidal anti-inflammatory drugs; PIP, proximal interphalangeal; RA, rheumatoid arthritis; UA, undifferentiated arthritis.

**JOINT DESTRUCTION IN UNDIFFERENTIATED ARTHRITIS AND RHEUMATOID ARTHRITIS**

Validated scoring methods are available to measure joint destruction objectively. Most scoring methods assess a limited number of joints. For example, the Sharp-van der Heijde method evaluates metacarpophalangeal (MCP), proximal interphalangeal (PIP), wrist, and metatarsophalangeal (MTP) joints but does not evaluate the distal interphalangeal or larger joints. Therefore, this method may be less optimal for measuring the level of joint damage in patients who present with predominant synovitis in large joints. In RA, the extent of joint destruction in small joints adequately reflects joint destruction in larger joints. Such data are missing for UA patients.

A recent study evaluated the subgroup of UA patients who developed RA later in time. Radiologic data were compared with those of patients who at first presentation of the disease directly fulfilled the 1987 ACR criteria for RA. Radiographic progression rates between these groups were not different. Health Assessment Questionnaire (HAQ) scores and Disease Activity Score (DAS) measures also were not different between these two groups. Thus, although UA patients as a group have a higher rate of spontaneous remission compared with patients with RA, the subgroup of UA patients who progress to RA have an
equally severe course compared with patients classified as RA earlier in the disease. Studies that compare the rate of joint destruction in the total group of early UA patients versus that of patients with early RA have not been performed.

**BIOLOGIC MECHANISMS IN UA AND DETERMINANTS OF PROGRESSION TO RA**

Our understanding of the processes responsible for progression from UA to RA is far from complete. Risk factors that were identified as independent predictors for RA development may provide clues:

- **Age:** The incidence of RA is clearly age dependent with a rising incidence from 7/100,000 for the age group 18 to 34 years to 107/100,000 for the age group 75 to 84 years. It is unknown why this is so; putative mechanisms include age-related decline in cellular, humoral, and innate immunity.

- **Gender:** Women who present with UA have a twofold greater chance of developing RA compared with men presenting with UA. Sex hormones influence predisposition to autoimmune disease. In general, men are less prone than women. Androgens have anti-inflammatory effects, and estrogens have been reported to suppress arthritis as well. Both estrogen and androgen inhibit bone resorption. These results may account for increased progression in postmenopausal females.

- **Number of involved joints and C-reactive protein:** Both markers reflect the level of inflammation and are frequently reported to be associated with progression of synovitis and worse disease outcomes. Although the number of swollen joints and the level of C-reactive protein (CRP) are correlated on a group level, this is often not the case in individual patients, and both markers have their own, independent predictive value. UA patients presenting with CRP levels greater than 50 mg/L have a five times increased risk of being in an early stage of RA. Likewise, patients classified with UA who have polyarthritis have a 1.5 times higher chance of fulfilling the criteria for RA later on compared with patients with monarthritis or oligoarthritis. Moreover, the risk that an early UA patient who has more than 10 swollen joints will develop RA is three times greater than that of patients with monarthrits or oligoarthritis.

- **Autoantibodies:** ACPAs as well as rheumatoid factor (RF) can be present years before the first clinical symptom of synovitis is noted; they are a risk factor for a persistent and destructive course of synovitis. Spontaneous remission is uncommon in ACPA-positive patients. Not only the presence but also the level of ACPAs is of predictive relevance. Despite these strong associations, studies that formally prove that ACPAs themselves are causally related to disease progression are lacking.

- **Environmental factors:** Early UA patients who smoke have a higher risk of development of RA and a destructive disease course. This risk is confined to patients who carry HLA-DRB1 alleles that encode for the so-called shared epitope. Persons who smoke and also carry an HLA-DRB1 shared epitope allele are particularly prone to develop ACPAs, which subsequently are associated with disease persistency and erosiveness.

- **Genetic factors:** Apart from the HLA-DRB1 shared epitope alleles, genetic factors associated with progression from UA to RA are not clearly identified. Most genetic risk factors for RA provide risk for ACPA-positive RA when compared with healthy controls. It is not clear whether these factors are associated with progression from UA to RA as well, and if so, whether such an association is independent of the strong association between ACPAs and development of RA. Identification of new genetic factors has fueled the study of their relevance. In a population of UA patients, information on currently known genetic risk factors for RA does not improve prediction of risk for RA compared with a prediction rule based on common clinical risk factors alone.

- **Additional biomarkers:** The role of biomarkers in the diagnosis of UA is limited. Within RA, it is known that serologic levels of pro–matrix metalloproteinase (MMP)3, receptor activator of nuclear factor κB (NFκB) ligand (RANKL), and osteoprotegerin (OPG) correlate with the rate of joint destruction over time. Biomarkers that specifically reflect or predict disease progression in early UA patients thus far remain unknown.

**WINDOW OF OPPORTUNITY**

The concept of a window of opportunity suggests that there is a period early in the course of the disease when the disease process can be altered or can even be reversed with a complete return to normality. Treatment during this period might have a greater effect than treatment at a later stage in terms of halting disease progression and achieving remission. Several different aspects have been studied, including whether very early UA (<12 weeks) may be an immunopathologically distinct phase compared with later disease.

Studies that focused on synovial tissue did not reveal differences. It is difficult to come to a conclusion from negative studies because one never knows whether the relevant processes have been studied. However, studies that focused on the composition of autoantibody responses have demonstrated that profound maturation of the autoantibody response occurs early during disease.

UA patients who present to an early arthritis clinic within 12 weeks of symptom onset less often progress toward RA compared with UA patients with symptom duration greater than 12 weeks. Similarly, RA patients who at their first visit to a rheumatologist reported symptoms for less than 12 weeks had a lower rate of joint destruction over time and achieved sustained DMARD-free remission more often than RA patients who reported a period greater than 12 weeks before seeing a rheumatologist.
Support for the presence of a window of opportunity is also obtained from trials. An unblinded study of a single dose of corticosteroids in patients with mild early inflammatory arthritis (median, 20 weeks) found that the strongest predictor of disease remission at 6 months (defined as the absence of symptoms and signs in patients without anti-inflammatory treatment) was disease duration less than 12 weeks at the time of therapy. Clinical and radiologic outcomes were significantly better at 3 years in patients who started DMARD therapy within 3 months (“very early”) after disease onset compared with a median duration of 12 months at the start of treatment (“early”). Remission was achieved in 50% of the very early group compared with 15% of the early group. These data suggest that treatment in very early arthritis might have a greater effect on disease progression and thus underline the relevance of identification and treatment of arthritis in the very early disease phase.

TREATMENT OF UNDIFFERENTIATED ARTHRITIS

Although many rheumatologists have experience treating UA patients, formal data providing information on treatment of UA patients are lacking. Thus far, only a few trials have been performed in UA patients; these are summarized in Table 72-2. One trial observed an effect of steroids in early UA patients—an observation that was not reproduced in another study. Methotrexate is reportedly beneficial in UA patients and is associated with delayed progression to RA and a reduction in the rate of joint destruction. This effect appeared to be predominant among ACPA-positive UA patients but was negligible in ACPA-negative UA patients. However, when methotrexate was discontinued, the rate progression to RA was similar to the placebo arm.

Some biologics have been studied in UA patients. Although surprising lack of efficacy was observed with tumor necrosis factor (TNF) blockers, abatacept was beneficial. The heterogeneous nature of UA patients and the more favorable disease outcome and spontaneous remission rate make it difficult to observe treatment efficacy in the total UA group. Thus the placebo-treated group can have a good remission rate, and differences from the treatment group might be difficult to detect. None of the published randomized trials in UA patients addressed this problem of heterogeneity and stratified patients having a high chance of developing RA or of having a spontaneous remission.

Individualized Treatment of Undifferentiated Arthritis

UA has a variable disease course, ranging between spontaneous remission and severe destruction. Because of this, and because DMARD therapy is potentially toxic, treatment of UA patients should be personalized. The chance for individual patients to progress toward RA or to have a persistent erosive disease course can be estimated using prediction models. One of the present prediction rules (Figure 72-2A) is validated using data from early arthritis clinics from Germany, the United Kingdom, Canada, Russia, and Japan; this algorithm is currently used in daily practice in several countries. The discriminative ability of this model is high. The prediction rule consists of nine variables: age, gender, distribution of involved joints, morning stiffness, numbers of tender and swollen joints, CRP level, and presence of RF and ACPAs. It calculates the risk of developing RA for every UA patient (Figure 72-2B). Such information can facilitate the decision of whether or not to initiate DMARD therapy and might facilitate patient involvement in decision making. In general, a score of 6 or lower is related to a low chance of developing RA (91% chance of not developing RA) and may be a reason to not initiate DMARD therapy. In contrast, patients with a score of 8 or higher have an 84% chance of developing RA; this might be a reason to initiate DMARD therapy in some patients.

Can the 2010 ACR/EULAR criteria be used to identify UA patients who actually have early RA? The presence of new criteria for RA is relevant in that the absence of up-to-date classification criteria had hampered progress with regard to treatment strategies in early RA. The 1987 ACR criteria for RA were not designed to classify RA early and did not include more recent autoantibody tests or imaging modalities. Most randomized clinical trials included patients who fulfilled the 1987 criteria, hence patients with established RA. It is not clear whether the efficacy observed in these trials is the same for patients with early RA or UA.

Table 72-2  Randomized Controlled Trials in Patients with Early Undifferentiated Arthritis

<table>
<thead>
<tr>
<th>Trial</th>
<th>N</th>
<th>Treatment</th>
<th>Follow-up Duration</th>
<th>Outcome</th>
<th>Effect Compared with Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAVE trial</td>
<td>389</td>
<td>Single IM injection 120 mg methylprednisolone</td>
<td>52 wk</td>
<td>Drug-free clinical remission</td>
<td>No effect</td>
</tr>
<tr>
<td>STIVEA trial</td>
<td>265</td>
<td>3 IM injections 80 mg methylprednisolone</td>
<td>6 and 12 mo</td>
<td>Need to start DMARDs</td>
<td>At 6 mo 61% vs. 76% had started DMARDS;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>At 12 mo 10% vs. 20% had resolved disease</td>
</tr>
<tr>
<td>PROMPT trial</td>
<td>110</td>
<td>Methotrexate therapy during 12 mo</td>
<td>30 mo</td>
<td>Fulfilling the 1987 ACR criteria</td>
<td>40% vs. 53% developed RA; lower radiologic</td>
</tr>
<tr>
<td>ADJUST trial</td>
<td>56</td>
<td>Abatacept treatment during 6 mo</td>
<td>12 mo</td>
<td>Fulfilling the 1987 ACR criteria</td>
<td>progression</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46% vs. 67% developed RA; no effect on</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>radiographic progression</td>
</tr>
<tr>
<td>Saleem et al</td>
<td>17</td>
<td>Infliximab at 0, 2, 4, and 16 wk</td>
<td>26 wk</td>
<td>Clinical remission</td>
<td>No effect observed</td>
</tr>
</tbody>
</table>

ACR, American College of Rheumatology; DMARDs, disease-modifying antirheumatic drugs; IM, intramuscular; RA, rheumatoid arthritis.
Figure 72-2  Tool to predict an individual early arthritis patient’s probability of developing rheumatoid arthritis (RA). With the calculated prediction score (A), the probability can be derived (B). VAS, visual analog scale.

<table>
<thead>
<tr>
<th>1. What is the age?</th>
<th>Multiply with 0.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. What is the gender?</td>
<td>In case female: 1 point</td>
</tr>
<tr>
<td>3. How is the distribution of involved joints?</td>
<td>In case small joints hands and feet: 0.5 point</td>
</tr>
<tr>
<td></td>
<td>In case symmetric: 0.5 point</td>
</tr>
<tr>
<td></td>
<td>In case upper extremities: 1 point</td>
</tr>
<tr>
<td>Or:</td>
<td>In case upper and lower extremities: 1.5 points</td>
</tr>
<tr>
<td>4. What is the length of the VAS morning stiffness (range 0-100 mm)?</td>
<td>In case 26-90 mm: 1 point</td>
</tr>
<tr>
<td></td>
<td>In case &gt;90 mm: 2 points</td>
</tr>
<tr>
<td>5. What is the number of tender joints?</td>
<td>In case 4-10: 0.5 points</td>
</tr>
<tr>
<td></td>
<td>In case 11 or higher: 1 point</td>
</tr>
<tr>
<td>6. What is the number of swollen joints?</td>
<td>In case 4-10: 0.5 point</td>
</tr>
<tr>
<td></td>
<td>In case 11 or more: 1 point</td>
</tr>
<tr>
<td>7. What is the C-reactive protein level (mg/L)?</td>
<td>In case 5-50: 0.5 point</td>
</tr>
<tr>
<td></td>
<td>In case 51 or higher: 1.5 points</td>
</tr>
<tr>
<td>8. Is the rheumatoid factor positive?</td>
<td>If yes: 1 point</td>
</tr>
<tr>
<td>9. Are the anticitrullinated protein antibodies positive?</td>
<td>If yes: 2 points</td>
</tr>
</tbody>
</table>

Total score

SUBJECTS' PREDICTION SCORES PLOTTED VERSUS THE PREDICTED RISK TO DEVELOP RA
The new criteria have mainly been derived for classification and research purposes\(^\text{17}\) and require validation for use in clinical practice for early synovitis. The 2010 criteria should not be used as diagnostic criteria until it is demonstrated that patients who fulfill them have a high chance for a persistent or destructive disease course, making 2010 criteria positivity the basis of an argument for starting disease-modifying drugs.

### References


The references for this chapter can also be found on www.expertconsult.com.
Sjögren’s Syndrome

E. WILLIAM ST. CLAIR

KEY POINTS

Sjögren’s syndrome is divided into primary and secondary forms. The primary form occurs in approximately 0.1% to 0.6% of the general population.

The clinical hallmarks of Sjögren’s syndrome are keratoconjunctivitis sicca (dry eyes), xerostomia (dry mouth), and parotid gland swelling.

Extraglandular features of primary Sjögren’s syndrome include fatigue, Raynaud’s phenomenon, polyarthralgia/arthritis, interstitial lung disease, neuropathy, and purpura.

A chronic mononuclear cell infiltration of the lacrimal and salivary glands is the characteristic histopathologic finding.

A diagnosis of primary Sjögren’s syndrome is made by subjective and objective assessment of dry eyes and dry mouth, testing for serum antinuclear antibodies including anti-Ro/SS-A and anti-La/SS-B antibodies, and labial salivary gland biopsy.

Treatment of Sjögren’s syndrome aims to provide symptomatic improvement in the symptoms of dry eyes and dry mouth, as well as control of extraglandular manifestations of disease.

HISTORICAL PERSPECTIVE

In 1888 Johann von Mikulicz-Radecki described a case of bilateral painless swelling of the lacrimal, parotid, and submandibular glands, an entity that later bore his name. Reports soon followed showing Mikulicz disease was not a distinct pathologic entity but rather a clinical potpourri of conditions including leukemia, lymphoma, and tuberculosis. Shortly thereafter, salivary gland disease became linked with dryness of the eyes and oral mucosa when Henri Gougerot, a well-known and successful French dermatologist, wrote in 1925 about three cases of salivary gland atrophy with dryness of the eyes, mouth, and vagina. The modern concept of Sjögren’s syndrome was firmly rooted in 1933 when Henrik Sjögren, a Swedish ophthalmologist, reported a series of 19 cases of keratoconjunctivitis sicca including two cases with swelling of the major salivary glands. Over the next 2 decades, Sjögren and others published extensively about various facets of the disease that now carries his name (also called Gougerot-Sjögren disease), with most of these contributions coming from European ophthalmologists.

In 1953 Morgan and Castleman published their detailed histopathologic findings from 18 patients with enlarged lacrimal and salivary glands. This pathologic treatise consisted of cases without apparent etiology that had clinical features resembling those of Mikulicz’s disease and Sjögren’s syndrome. Notably, 15 of the 18 patients were women with lacrimal and salivary gland swelling that developed in the fifth and sixth decades of life. Morgan and Castleman found in the salivary gland tissue a consistent “lymphoid element” accompanied by a striking proliferation of myoepithelial and epithelial cells. The epithelial changes produced a characteristic narrowing or obliteration of the ductal lumen, forming cords of solid cell masses they called epimyoepithelial islands. Because of the prominence of these epithelial structures, they advanced the theory that Mikulicz disease had its origins in the ductal epithelium. Morgan and Castleman also recognized that the salivary gland pathology in their cases was similar to that of the patients described by Henrik Sjögren with keratoconjunctivitis sicca. Because Sjögren’s patients were mostly middle-aged women, Morgan and Castleman reasoned that Mikulicz disease was a subset of Sjögren’s disease but with incomplete clinical manifestations. This now classic paper had the effect of unifying Mikulicz disease and Sjögren’s syndrome into a single disease entity, a belief that dominated the field until recently.

When Joseph J. Bunim delivered the Heberden Oration at the Wellcome Foundation in London on December 2, 1960, he enlightened his audience about the latest advances in Sjögren’s syndrome. Bunim described in detail the clinical, pathologic, and laboratory findings of the 40 patients with Sjögren’s syndrome evaluated by himself, Kurt Bloch, Martin Wohl, Richard Oglesby, and Irwin Ship at the Clinical Center of the National Institutes of Health (NIH). All patients in their series had at least two of the three following features: keratoconjunctivitis sicca, xerostomia (with or without enlargement of the salivary glands), and rheumatoid arthritis. Bunim also knew from the work of others that keratoconjunctivitis sicca and xerostomia, or sicca complex, occurred in some patients with systemic lupus erythematosus, scleroderma, polymyositis, and polyarteritis nodosa. He brought home the point that the exocrinopathy extended beyond the lacrimal and salivary glands to affect the pharynx, larynx, and trachea, as well as the vagina. Many of their patients with Sjögren’s syndrome also had extraglandular manifestations such as Raynaud’s phenomenon, purpura, pulmonary infiltrates on chest radiograph, and peripheral neuropathy. Talal and Bunim took note of the diagnosis of reticulum cell sarcoma (an older term that includes non-Hodgkin’s lymphoma) in three patients and Waldenström’s macroglobulinemia in a fourth patient from the NIH cases, drawing attention for the first time to the increased risk of developing lymphoma in this disease. In this report, they speculate that a “chronic state of immunologic hyperactivity and the proliferation of immunologically competent cells producing abnormal tissue antibodies
predisposes to the relatively frequent development of malignant lymphoma.\textsuperscript{45}

The NIH group also pioneered the characterization of the serologic markers of Sjögren’s syndrome. In 1965 they reported that 12 of their 16 patients (75%) with sicca symptoms but no evidence of another connective tissue disease had serum antinuclear antibody (ANA) reactivity by indirect immunofluorescence on rat liver tissue.\textsuperscript{6} Applying the Ouchterlony plate method, they further showed that sera from 13 of 16 (81%) of these patients contained precipitating antibodies to SjD and ST, which were later to be called anti-Ro (SS-A) and anti-La (SS-B) antibodies. Moreover, Bloch and colleagues\textsuperscript{7} set the stage for our modern classification schemes by subdividing their patients at the NIH into primary and secondary Sjögren’s syndrome. Secondary Sjögren’s syndrome was the category reserved for patients with sicca symptoms in the setting of another connective tissue disease such as rheumatoid arthritis, systemic lupus erythematosus, scleroderma, or dermatomyositis, whereas primary Sjögren’s syndrome was the designation when the sicca complex occurred in the absence of another connective tissue disease.

In the late 1960s and early 1970s, the investigation of labial salivary gland biopsy for the diagnosis of Sjögren’s syndrome led to the development of grading systems for quantifying the intensity of tissue inflammation. In 1968 Chisholm and Mason proposed a grading system for labial salivary gland biopsies that used a simple grading scale (0 to 4) in which grades 0, 1, and 2 represented absent, slight, and moderate mononuclear cell infiltrates, respectively.\textsuperscript{8} The higher grades of 3 and 4 corresponded to focal inflammatory cell scores of 1 and greater than 1 per 4 mm\textsuperscript{2} of tissue, where a focus equated to an aggregate of 50 or more mononuclear cells. In this initial study, 9 of the 10 patients with Sjögren’s syndrome were classified as grade 3 (n = 3) or grade 4 (n = 6). By comparison, biopsies had lower grades when they were taken from patients with other chronic inflammatory conditions, as well as controls. For example, among 10 patients with rheumatoid arthritis, the highest biopsy grade was only 3. Twenty other subjects with a variety of other rheumatic diseases (e.g., osteoarthritis, reactive arthritis, psoriatic arthritis, scleroderma) had biopsies with grades of 0 or 1. Although it was a small study by today’s standards, this work inspired others to validate and refine these findings. Later, postmortem biopsies from 116 controls had other shortcomings limiting their clinical use.

With a growing interest in Sjögren’s syndrome, investigators began to develop classification criteria for comparison of results across studies. In the 1980s, many groups proposed classification criteria for Sjögren’s syndrome, aiming to identify patients with a sicca complex caused by a chronic autoimmune disorder. These proposed criteria incorporated a combination of the following items: symptoms of the ocular and oral components; objective measures of lacrimal and salivary gland involvement; a focus score greater than or equal to 1 from a labial salivary gland biopsy; and the presence of serum autoantibodies. Many investigators from Europe, the United States, and Japan contributed to this effort, leading to the general acceptance in 2002 of the revised version of the European criteria proposed by the American-European Consensus group.\textsuperscript{12} The International Sjögren’s Syndrome Registry, funded in 2003, is collecting data that will inform the natural history of Sjögren’s syndrome. It will also support validation of new criteria sets that may be developed from the baseline data of the more than 1200 patients across the world with suspected primary or secondary Sjögren’s syndrome.

**DEFINITIONS AND CLASSIFICATION CRITERIA**

The clinical hallmarks of Sjögren’s syndrome are keratoconjunctivitis sicca and xerostomia, or the sicca complex. The term “keratoconjunctivitis sicca” is derived from Latin, and its translation is “dryness of the cornea and conjunctiva.” Xerostomia refers to the subjective symptoms of dry mouth. Sjögren’s syndrome is subdivided into primary and secondary Sjögren’s syndrome. The category of secondary Sjögren’s syndrome is reserved for patients with keratoconjunctivitis sicca or xerostomia, or both, in the setting of another connective tissue disease or chronic inflammatory process such
as rheumatoid arthritis, systemic lupus erythematous, polymyositis, systemic sclerosis (scleroderma), or granulomatosis with polyangiitis (formerly Wegener's granulomatosis). Patients are diagnosed with primary Sjögren's syndrome if they manifest signs and symptoms of keratoconjunctivitis sicca and xerostomia in the absence of another connective tissue disease provided they meet serologic or histopathologic criteria. Because primary Sjögren's syndrome is associated with extraglandular features that overlap with those of other connective tissue diseases, it may be difficult in some cases to distinguish from patients with secondary Sjögren's syndrome. The challenge in distinguishing between primary and secondary forms of the disease arises most often in patients with primary Sjögren's syndrome and overlapping features of lupus such as rash, arthritis, and leukopenia. This distinction is often a matter of semantics depending on the preference for “lumping” as opposed to “splitting.”

Other terms used to describe Sjögren's syndrome are autoimmune exocrinopathy and autoimmune epithelitis. In Sjögren's syndrome, the term exocrinopathy receives emphasis because of the generalized glandular involvement causing dysfunction of the lacrimal and salivary glands, as well as the apocrine sweat glands of the skin and the submucosal glands of the nose, pharynx, larynx, large airways, and vagina. The term epithelitis gains footing from the uniformly activated epithelial cells omnipresent in the lacrimal and salivary glands and other sites of glandular involvement.

International agreement has been reached on the classification of Sjögren's syndrome. By the early 1980s, several criteria sets had been proposed for the classification of Sjögren's syndrome including the Copenhagen criteria, the Japanese criteria, the Greek criteria, and the California criteria. Common to each of these criteria sets was the requirement for objective evidence of keratoconjunctivitis sicca and salivary gland involvement. They differed, however, in their item content and weighting, as well as the methods for assessing salivary gland involvement. For example, some of the criteria relied on whole salivary flow (Copenhagen), whereas other criteria relied on parotid flow rate (Greek, California). Only two of the criteria sets discriminated between primary and secondary Sjögren's syndrome (Copenhagen and Greek).

Movement toward consensus began when the Epidemiology Committee of the European Community conducted a multicenter study involving 26 centers from 12 countries aimed at developing criteria for the classification of Sjögren's syndrome. This group first agreed on the items that should be included in the classification criteria for Sjögren's syndrome and then tested their operational characteristics in a large sample of patients with a clinical diagnosis of primary Sjögren's syndrome (n = 246), secondary Sjögren's syndrome (n = 201), other connective tissue diseases without Sjögren's syndrome (n = 113), and healthy controls (n = 133). Two sets of three questions were selected from a larger body of questions that best correlated with the presence of keratoconjunctivitis sicca and xerostomia, respectively, as judged by clinical experts. The results from objective tests were analyzed by univariate analyses, yielding a tentative list of items based on their sensitivity and specificity for correct classification of the diagnosis. The following six items were chosen for the classification criteria: (I) ocular symptoms; (II) oral symptoms; (III) ocular signs (Schirmer-I-test ≤5 mm/5 min or Rose Bengal score ≥4 by the van Bijsterveld scoring system); (IV) histopathologic features (focus score ≥1 on labial salivary gland biopsy); (V) objective evidence of salivary gland involvement by at least one abnormal test (salivary scintigraphy, parotid sialography, or unstimulated salivary flow rate ≤1.5 mL/15 min); and (VI) at least one of the following serum autoantibodies: anti-Ro/SS-A or anti-La/SS-B antibodies, ANAs, or rheumatoid factor. Exclusion criteria were pre-existing lymphoma, acquired immunodeficiency syndrome, sarcoidosis, and grafted-versus-host disease. The presence of four of six criteria (accepting only a positive test for anti-Ro/SS-A or anti-La/SS-B antibodies for item VI) had good sensitivity (93.5%) and specificity (94%) for correctly classifying a patient with primary Sjögren's syndrome. It was suggested to be the optimal combination for efficient classification of this condition. For classification of secondary Sjögren's syndrome, the best combination was found to be a positive response for items I or II plus a positive result for any two of items II, IV, and V, resulting in a sensitivity of 85.1% and specificity of 93.9% for this diagnosis. With a positive response to items I or II and only one positive response to items III to V, the sensitivity increased to 93.6% but the specificity dropped to 71.6%.

These preliminary European classification criteria for Sjögren's syndrome were validated in a second study of similar design. Cases of primary Sjögren's syndrome (n = 81), secondary Sjögren's syndrome (n = 76), other connective tissue diseases without Sjögren's syndrome (n = 54), and controls (n = 67) were assembled by expert clinicians at 16 centers from 10 countries to validate the initial results. In this study, it was decided not to use items III (a) (Schirmer-I-test) and V (c) (unstimulated whole salivary flow) to classify patients older than age 60 because the tear and salivary flow by these measures were significantly reduced in the elderly controls from the original study. The results proved to be similar to those of the earlier study, showing a sensitivity of 97.5% and specificity of 94.2% for correctly classifying primary Sjögren's syndrome and a sensitivity of 97.3% and specificity of 91.8% for correctly classifying secondary Sjögren’s syndrome. A limitation of this type of study design is the possibility that the sensitivity and specificity of these criteria may be artificially inflated by a “circular bias” deriving from the results of the diagnostic testing on the original selection of patients. These preliminary European classification criteria have also been criticized because a patient may be classified with primary Sjögren's syndrome despite a negative biopsy and the absence of serum anti-Ro/SS-A or anti-La/SS-B antibodies. Thus patients may be classified as having primary Sjögren's syndrome without evidence of an immune basis for their condition.

This conceptual impasse was overcome by the European Study Group on Classification Criteria for Sjögren's Syndrome with a group of American experts. From their previous study, they selected a cohort of 76 patients with primary Sjögren's syndrome, 41 patients with another connective tissue disease but without Sjögren's syndrome, and 63 controls and tested three different combinations of items from the original classification criteria using a receiving operator curve (ROC) analysis. To determine the best possible prediction model, the optimal classifiers were compared with ROC analysis for the three different combinations: C point...
(positivity for any four of the six items); C* point (positivity of any four of the six items, excluding cases negative for both items IV and VI); and D point (positive for any of the four objective criteria). The C point and the C* point showed the same accuracy (92.7%); however, the C* point had a lower specificity than the C point (89.5% vs. 97.4%), but a higher specificity (95.2% vs. 89.4%), and was preferred on the basis of the goal of selecting criteria with a high probability of excluding patients without the disease. The sensitivity and specificity of the D point were 84.2% and 95.2%, respectively, and judged to be comparable with the C* point and acceptable for classification purposes. Therefore patients may now be classified as having primary Sjögren's syndrome if they fulfill four of the six items including either a positive labial salivary gland biopsy or a positive test for serum anti-Ro(SS-A) and/or anti-La(SS-B) antibodies. This combination avoids the problem of classifying a patient with primary Sjögren's syndrome in the absence of a positive labial salivary gland biopsy or a positive test for anti-Ro(SS-A) and/or anti-La(SS-B) antibodies. Alternatively, they may meet the criteria by satisfying any four of the objective items, which was a highly unusual scenario in their data set because nearly all patients with primary Sjögren's syndrome manifested symptoms of dry eyes and/or dry mouth.

The revised classification criteria proposed by the American-European Consensus Group are shown in Table 73-1. It was further specified that the Schirmer-I test be performed with anesthesia and that other ocular dyes such as lissamine green (for conjunctival staining) and fluorescein (for corneal staining) be allowed as replacements for Rose Bengal dye, which was not available in many countries. A positive labial salivary gland biopsy was further defined according to the rules set forth by Daniels and Whitcher. In this case, a positive biopsy must show evidence of focal lymphocytic sialadenitis (FLS), which they defined as dense aggregates of 50 or more lymphocytes in perivascular or periductal locations. These aggregates, or

### Table 73-1 Revised International Classification Criteria for Sjögren's Syndrome (SS)

<table>
<thead>
<tr>
<th>I. Ocular symptoms: a positive response to at least one of the following questions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Have you had daily, persistent, troublesome dry eyes for more than 3 months?</td>
</tr>
<tr>
<td>2. Do you have a recurrent sensation of sand or gravel in the eyes?</td>
</tr>
<tr>
<td>3. Do you use tear substitutes more than 3 times a day?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Oral symptoms: a positive response to at least one of the following questions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Have you had a daily feeling of dry mouth for more than 3 months?</td>
</tr>
<tr>
<td>2. Have you had recurrently or persistently swollen salivary glands as an adult?</td>
</tr>
<tr>
<td>3. Do you frequently drink liquids to aid in swallowing dry food?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Ocular signs—that is, objective evidence of ocular involvement defined as a positive result for at least one of the following two tests:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Schirmer’s I test, performed without anaesthesia (≤5 mm in 5 minutes)</td>
</tr>
<tr>
<td>2. Rose bengal score or other ocular dye score (≥4 according to van Bijsterveld’s scoring system)</td>
</tr>
</tbody>
</table>

| IV. Histopathology: in minor salivary glands (obtained through normal-appearing mucosa), focal lymphocytic sialadenitis, evaluated by an expert histopathologist, with a focus score greater than 1, defined as a number of lymphocytic foci (which are adjacent to normal-appearing mucous acini and contain more than 50 lymphocytes) per 4 mm² of glandular tissue |

<table>
<thead>
<tr>
<th>V. Salivary gland involvement: objective evidence of salivary gland involvement defined by a positive result for at least one of the following diagnostic tests:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Unstimulated whole salivary flow (≤1.5 mL in 15 minutes)</td>
</tr>
<tr>
<td>2. Parotid sialography showing the presence of diffuse sialectasias (punctate, cavitary or destructive pattern), without evidence of obstruction in the major ducts</td>
</tr>
<tr>
<td>3. Salivary scintigraphy showing delayed uptake, reduced concentration, and/or delayed excretion of tracer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VI. Autoantibodies: presence in the serum of the following autoantibodies:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Antibodies to Ro(SSA) or La(SSB) antigens, or both</td>
</tr>
</tbody>
</table>

### Revised Rules for Classification

**For Primary SS**

In patients without any potentially associated disease, primary SS may be defined as follows:

a. The presence of any 4 of the 6 items is indicative of primary SS, as long as either item IV (Histopathology) or VI (Serology) is positive.

b. Any 3 of the 4 objective criteria items (i.e., items III, IV, V, and VI) are present.

c. The classification tree procedure represents a valid alternative method for classification, although it should be more properly used in clinical-epidemiologic surveys.

**For Secondary SS**

In patients with a potentially associated disease (for instance, another well-defined connective tissue disease), the presence of item I or item II plus any 2 from among items III, IV, and V may be considered as indicative of secondary SS.

### Exclusion Criteria

- Past head and neck radiation treatment
- Hepatitis C infection
- Acquired immunodeficiency syndrome
- Pre-existing lymphoma
- Sarcoidosis
- Graft-versus-host disease
- Use of anticholinergic drugs (since a time shorter than fourfold the half-life of the drug)

**EPIDEMIOLOGY**

Primary Sjögren’s syndrome ranks among the most common of the autoimmune diseases, with a prevalence rate ranging from 0.1% to 4.6%. However, the epidemiologic data are confounded by variations in the ages of the study populations and differences in the classification criteria used for case identification. For example, the application of the Copenhagen criteria yields a prevalence rate for Sjögren’s syndrome 2.7-fold greater than the California criteria, whereas the preliminary European classification criteria captures 2- to 3-fold more patients than the Copenhagen criteria. Bowman and colleagues estimated the prevalence of primary Sjögren’s syndrome at 0.1% to 0.6% in a community from the United Kingdom using the revised criteria proposed by the American-European Consensus Group. In a study from Greece employing these same criteria, the age-adjusted mean annual incidence and prevalence rates were 5.3 (confidence interval [CI], 4.5 to 6.1) per 105 persons (0.5 for men and 10.1 for women) and 92.8 per 100,000 persons (8.4 for men and 177.4 for women), respectively. The incidence and prevalence rates of primary Sjögren’s syndrome are significantly higher in women than men (e.g., approximately 20:1), with a peak incidence in the fifth and sixth decades of life. Primary Sjögren’s syndrome occurs infrequently in children with onset as early as 5 years of age. The revised criteria of the American-European Consensus Group may be less sensitive for classifying primary Sjögren’s syndrome in children than adults owing to differences between these two age groups in the clinical presentation of this disease.

The prevalence of secondary Sjögren’s syndrome associated with rheumatoid arthritis, systemic lupus erythematosus, and systemic sclerosis has been estimated to be 17.1%, 8% to 20%, and 14% respectively. To be classified as secondary Sjögren’s syndrome, patients in these studies had at minimum symptoms of keratoconjunctivitis sicca or xerostomia and objective evidence of lacrimal or salivary gland involvement. None of these studies used the preliminary European classification criteria for secondary Sjögren’s syndrome.

**GENETICS AND PATHOGENESIS**

Central to the pathogenesis of primary Sjögren’s syndrome is the dysregulation of T cells and B cells, with key contributions from innate pathways of inflammation. Many of these same concepts may apply to the pathogenesis of secondary Sjögren’s syndrome; however, most investigation into disease mechanisms has focused on patients with the primary form of the disease, and many of the immunologic features of secondary Sjögren’s syndrome are driven by the pathogenic mechanisms underlying the associated disorders.

In primary Sjögren’s syndrome, a common theme among models of disease pathogenesis is loss of immunologic tolerance to self-antigens. This failure to recognize self manifests in primary Sjögren’s syndrome as the production of serum autoantibodies indicative of the loss of B cell tolerance. Because the appearance of serum autoantibodies may precede the onset of clinical disease, the loss of immune tolerance appears to be permissive but not sufficient to induce clinical disease. T cells react with several
candidate self-antigens have also been found infiltrating the labial salivary glands of patients with primary Sjögren’s syndrome. However, the timing of the evolution of these aberrant T cell responses is unclear with respect to disease onset. Environmental or stochastic factors or both are likely at play in triggering the chronic inflammatory response in the context of a genetically predisposed innate and adaptive immune system.

Because a disproportionately high number of the cases of primary Sjögren’s syndrome occur in women, the search for environmental candidates has centered on abnormal regulation of estrogens and androgens. However, no major differences have been found between patients with primary Sjögren’s syndrome and healthy controls in serum levels of sex steroid hormones. In addition, treatment with the hormone dehydroepiandrosterone, which acts on the androgen receptor, has shown no clinical efficacy in women with primary Sjögren’s syndrome. Among possible viral triggers, Epstein-Barr virus (EBV) and cytomegalovirus (CMV) have received attention in primary Sjögren’s syndrome due to their suppressive effects on T cell immunity and their ability to establish persistent infection. In one study, EBV deoxyribonucleic acid (DNA) was found by immunochemical staining in salivary gland biopsies to localize in acinar and ductal epithelial cells; however, the available evidence does not otherwise support a direct role for EBV infection in the pathogenesis of this disease. A 94-bp fragment of coxsackievirus ribonucleic acid (RNA) was also shown to be differentially expressed in salivary gland biopsies from patients with primary Sjögren’s syndrome compared with controls, but these results were not replicated in a later study. As yet, the identity of a possible viral trigger remains elusive, if indeed one exists.

It has been difficult to estimate the genetic risk for developing primary Sjögren’s syndrome owing to the absence of large twin studies. However, the existence of many large families with two or more members with primary Sjögren’s syndrome argues strongly for a genetic component to disease pathogenesis. Primary Sjögren’s syndrome is considered to be a complex genetic disorder, similar to the genetic susceptibility of systemic lupus erythematosus and rheumatoid arthritis. It is now clear from genetic studies of human autoimmune diseases that multiple genes contribute to disease risk and that individually each gene confers only modest effects on disease susceptibility. The exception to this rule is the relatively strong signal associated with the human leukocyte antigen (HLA) locus on human chromosome 6p21.3. In populations of European descent, confirmed HLA associations with primary Sjögren’s syndrome include DRB1*0301 (DR3), DRB1*1501 (DR2), DQA1*0103, DQA1*0501, DQB1*0201, and DQB1*0601. The disease-associated polymorphisms located in the DRB1*0301 and DRB1*1501 loci account for 90% of the HLA genetic contribution. The HLA locus appears to play a major role in the pathogenesis of autoantibody responses associated with primary Sjögren’s syndrome. In patients with primary Sjögren’s syndrome, higher titers of anti-Ro/SS-A and anti-La/SS-B antibodies have been linked to heterozygosity for the DQA1 and DQB1 alleles.

The genetic susceptibility to primary Sjögren’s syndrome will likely include inheritance of genes identified as risk factors in other autoimmune diseases. Familial clustering of cases of primary Sjögren’s syndrome with systemic lupus erythematosus, rheumatoid arthritis, and systemic sclerosis, as well as other autoimmune diseases, are consistent with this premise. There have been no large-scale genome-wide association studies in primary Sjögren’s syndrome. Most efforts to identify disease susceptibility genes in primary Sjögren’s syndrome have taken the candidate gene approach. Genes in the type I interferon (IFN) pathway have been the focus of several candidate gene studies because they are highly expressed in the peripheral blood and salivary glands of patients with primary Sjögren’s syndrome compared with controls. In a small cohort study, an increased risk of primary Sjögren’s syndrome was associated with a single nucleotide polymorphism in the splicing sequence of exon B of the gene encoding IFN regulatory factor 5 (IRF5). IRF5 is among the nine IRFs that signal through Toll-like receptors (TLRs) and is essential for inducing responses through TLR4, 7, and 9. Given the possible inciting roles of viral RNA and DNA, the endosomal TLR7 and TLR9 pathways may be important in primary Sjögren’s syndrome for the activation of the type I IFN pathway. A 5-bp insertion/deletion polymorphism (CGGCG insertion/deletion) in the promoter region of the IRF5 transcript has also been linked to primary Sjögren’s syndrome. This polymorphism appears to have functional significance because reovirus (double-stranded RNA virus)—infected salivary gland epithelial cells in culture bearing this polymorphism produce higher levels of the IRF5 transcript.

Other susceptibility genes have been identified in primary Sjögren’s syndrome, but the findings have not yet been widely replicated. In a case-control study, an increased risk for primary Sjögren's syndrome was associated with a variant haplotype of the transcription factor Signal Transducers and Activator of Transcription 4 (STAT4). This STAT4 polymorphism has also been shown to increase the risk for systemic lupus erythematosus and rheumatoid arthritis. STAT4 is a key intracellular signaling molecule involved in IL-12 and IL-23 signaling and is known to promote the development of T helper I (Th1) and T helper 17 (Th17) responses. Polymorphisms in the STAT4 and IRF5 genes appear to be additive in the risk for developing primary Sjögren’s syndrome. In other studies, Nordmark and colleagues have identified three gene loci associated with an increased risk of primary Sjögren’s syndrome—the early B cell factor (EBF1) gene, the interval encompassing family with sequence similarity 167 member A and B lymphoid tyrosine kinase (FAM167A-BLK), and the tumor necrosis family member 4 (TNFSF4, or the OX40L gene). All three genes are involved in B cell development, which is of particular interest due to the hyperactivated state of B cells in this disease.

The human genome contains conserved noncoding elements with functional sequences including regulatory motifs in promoters and untranslated regions of genes. Some of these conserved noncoding elements encode microRNAs, a novel means for regulation of gene expression. MicroRNAs influence both innate and adaptive immunity and have been shown to play a role in late B cell differentiation and development, as well as the establishment of B cell tolerance. Early studies have shown that microRNA expression patterns in salivary gland tissue can...
Table 73-2  Mouse Models of Sjögren’s Syndrome*

<table>
<thead>
<tr>
<th>Mouse Model</th>
<th>Phenotype</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spontaneous Disease Models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(NZB)/NZW F1 mice</td>
<td>Progressive focal sialadenitis</td>
<td>Glandular involvement F &gt; M</td>
</tr>
<tr>
<td>MRL/lpr</td>
<td>Lymphocytic infiltration of lacrimal and salivary glands; anti-Ro/SS-A and MR3 antibodies; oligoclonal expansion of T cells and IgA and IgM production in the salivary glands</td>
<td>Normal secretory function; mRNAs for IL-1 and TNF expressed in salivary glands before onset of sialadenitis</td>
</tr>
<tr>
<td>NOD and its derivatives</td>
<td>Lymphocytic infiltration of lacrimal and salivary glands; NOD.H2h4 strain (but not parental NOD strain) secretes high levels of anti-Ro/SS-A and anti-La/SS-B antibodies; reduced glandular function</td>
<td>Mice also develop diabetes; exchange of H2 haplotype from H2g7 to H2q (NOD.Q) or H2p (NOD.P) does not affect the frequency of sialadenitis; disruption of ICA69 locus prevents lacrimal gland inflammation and reduces salivary gland inflammation; NOD IFN-γ−/− mice do not develop glandular disease; blockade of LT-αR signaling pathway reduces salivary gland infiltrates and improves salivary gland function</td>
</tr>
<tr>
<td>NFS/sld</td>
<td>Lymphocytic infiltration of lacrimal and salivary glands; anti-α-fodrin antibodies</td>
<td>Aberrant immune responses to α-fodrin; mice develop autoimmune lesions in other organ systems</td>
</tr>
<tr>
<td><strong>Experimentally Induced Models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carbonic anhydrase (PL/J mice)</td>
<td>Lymphocytic infiltration of salivary glands; antibodies to carbonic anhydrase</td>
<td>Disease induction requires multiple injections of peptide emulsified in Freund’s adjuvant</td>
</tr>
<tr>
<td>Ro peptides (Balb/c mice)</td>
<td>Lymphocytic infiltration of salivary glands; glandular hypofunction, anti-Ro/SS-A and anti-La/SS-B antibodies</td>
<td></td>
</tr>
<tr>
<td><strong>Transgenic or Knockout Models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Id3−/−</td>
<td>Lymphocytic infiltration of lacrimal and salivary glands; adoptive transfer experiments; anti-Ro/SS-A and anti-La/SS-B antibodies; salivary hyposecretion</td>
<td>Id3 gene involved in TCR-mediated T cell indicate a role for T cells in the development of disease; treatment with anti-CD20 antibodies ameliorates disease</td>
</tr>
<tr>
<td>PI3K−/−</td>
<td>Lymphocytic infiltration of the lacrimal glands; anti-Ro/SS-A and anti-La/SS-B antibodies</td>
<td>No anti-Ro/SS-A or anti-SS-B antibodies; also develop lupus manifestations and anti-DNA antibodies and RF</td>
</tr>
<tr>
<td>BAFF transgenic</td>
<td>Lymphocytic infiltration of lacrimal and salivary glands; unique population of marginal zone B cells in salivary glands</td>
<td>IL-14 is a growth factor for B cells; mice develop large B cell lymphomas later in life; role for LT-α in the salivary gland inflammation; infiltrate primarily B cells with relatively few CD4+ and CD8+ T cells</td>
</tr>
<tr>
<td>IL-14α transgenic</td>
<td>Lymphocytic infiltration of lacrimal and salivary glands; hypergammaglobulinemia; &lt;25% of animals develop anti-Ro/SS-A and anti-La/SS-B antibodies; glandular hyposecretion; mild immune complex-mediated renal disease and lymphocytic interstitial pneumonitis</td>
<td></td>
</tr>
<tr>
<td>IL-12 transgenic</td>
<td>Lymphocytic infiltration of lacrimal and salivary glands; anti-La/SS-B antibodies; glandular hyposecretion; increased acinar cell volume</td>
<td>Mice also develop thyroiditis and lung pathology</td>
</tr>
</tbody>
</table>

*See references 58 and 59 for further details.

BAFF, B cell activating factor; F, female; ICA69, islet cell antigen 69; Id3, protein inhibitor of DNA binding 3; IFN-γ, interferon-γ; IL-1, interleukin-1; LT-α, lymphotoxin-α; LT-βR, lymphotoxin β receptor; M, male; MR3, muscarinic receptor subtype 3; NOD, nonobese diabetic; NZB/NZW F1, New Zealand Black x New Zealand White F1 (mouse hybrid); PI3K, phosphoinositide 3-kinase; TNF, tumor necrosis factor.

distinguish patients with Sjögren’s syndrome from controls, suggesting a pathologic role for dysregulated microRNA expression in the regulation of the chronic inflammatory response.77

Epigenetics refers to the study of inherited changes in gene expression caused by mechanisms other than a change in DNA sequence of that gene. Its principal mechanisms include remodeling of chromatic structure by DNA methylation and histone modification. Thus far, virtually nothing is known about the epigenetics of primary Sjögren’s syndrome.

Many insights have been gained about disease mechanisms in primary Sjögren’s syndrome from studies of humans with this autoimmune disease, as well as animal models. Experiments in model systems enable hypotheses to be examined in ways not possible in humans and have contributed to understanding the immunoregulatory disturbances underlying the clinical expression of disease (Table 73-2).38,58

Whereas these animal models have shed light on disease mechanisms, it has been possible to perform highly significant research in humans with primary Sjögren’s syndrome because of the accessibility of labial salivary gland tissue for immunohistopathologic analysis. Studies of labial salivary gland biopsies from patients with this disease have shown approximately 90% of the infiltrating cells are composed of CD4+ T lymphocytes and B lymphocytes, with the remainder an admixture of plasma cells, CD8+ T cells, FoxP3+ regulatory cells, CD56+ natural killer (NK) cells, and macrophages, as well as myeloid and plasmacytoid dendritic cells (DCs)60 (Figure 73-1). Most of the infiltrating T cells bear the memory phenotype (CD45RO) and display a restricted T cell receptor (TCR) repertoire representing...
several different clonotypes across multiple Vβ families. The proportion of B cells in the infiltrate increases with the severity of the inflammatory lesion.

The infiltrating mononuclear cells tend to coalesce around ducts and blood vessels, and in more severe inflammatory lesions they may form aggregates organized into germinal center (GC)-like structures. The GC-like structures display well-circumscribed mononuclear cell infiltrates with B and T cell components, Ki-67+ proliferating cells, CD21/CD35+ follicular DC networks, and CD31+ high endothelial venules (HEVs). CXCL13 expression by epithelial cells, HEVs, and within germinal center-like structures together with the expression of CXCL12 and CCL21 provides a salivary gland microenvironment capable of attracting and retaining B cells. The myeloid DCs and macrophages, the classic antigen-presenting cells, are mostly found in proximity to the ductal epithelium, where they have been shown to secrete tumor necrosis factor (TNF), interleukin-6 (IL-6), IL-10, IL-12, and IL-18. Minor salivary glands also contain a small number of plasmacytoid DCs, the main producers of the type I IFNs. A robust type I IFN signature has been detected both in the salivary glands and peripheral blood of patients with primary Sjögren’s syndrome. The generation of type I IFNs by plasmacytoid DCs, among the first lines of defense against viral infection, is partially dependent on signals through TLR7 and TLR9.

The analysis of T cell cytokines in labial salivary gland biopsies from patients with primary Sjögren’s syndrome suggests a predominantly Th1- and Th17-driven response. Studies of the expression of cytokine messenger RNAs (mRNAs) in salivary gland tissue show mainly upregulation of IL-2 and IFN-γ, the Th1-specific cytokines, and lesser quantities of IL-4, IL-5, and IL-13, the Th2-specific cytokines. Th17 cells, so named because they produce IL-17, are also found in the minor salivary glands; the salivary gland microenvironment is also rich in transforming growth factor (TGF)-β, IL-6, and IL-23, which are cytokines known to promote the development of this subset. In addition, the salivary gland infiltrates contain a generous infiltrate of FoxP3+ regulatory T cells. These cells typically show suppressive behavior, but as yet they have an uncertain role in regulating the chronic inflammatory lesion.

Several features of primary Sjögren’s syndrome implicate B cells in disease mechanisms. B cells are the source of autoantibodies, but they also activate T cells through their ability to present antigen, are stimulated to secrete proinflammatory and anti-inflammatory cytokines, and facilitate the organization of secondary and tertiary lymphoid tissue.
The frequent presence of hypergammaglobulinemia, circulating immune complexes, mixed monoclonal immunoglobulin M (IgM) cryoglobulinemia, and serum autoantibodies in patients with primary Sjögren’s syndrome implies B cells are in a dysregulated state. Most B cells in the salivary gland tissue exhibit a memory phenotype, express somatically hypermutated immunoglobulin V genes, and show preferential changes in V \_i gene usage and the length of CDR3 from V \_i gene rearrangements, which are findings characteristic of an antigen-driven response. Salivary gland B cells appear to be responding specifically to antigen, although the site of antigen encounter and activation (salivary gland vs. secondary lymphoid tissue) remains uncertain. Germinial center–like structures found in some of the salivary gland biopsies provide an opportune environment for the selection and differentiation of antigen-driven B cell responses. The distribution of peripheral blood B cell subsets in primary Sjögren’s syndrome differs from healthy controls, showing increased proportions of CD27⁻ (naive) B cells and decreased proportions of CD27⁺ (memory) B cells. Among several possibilities, these changes in the peripheral blood may result from abnormal trafficking of CD27⁺ memory B cells to the target tissues, aberrations in B cell development, or both.

Several intrinsic B cell defects have also been found in patients with primary Sjögren’s syndrome including abnormal retention of preswitch Ig transcripts in circulating post-switch CD27⁺ memory B cells and enhanced B cell signaling due to altered kinetics of B cell receptor translocation to lipid rafts. Moreover, studies in patients with primary Sjögren’s syndrome have shown increased serum and salivary gland tissue levels of B cell activating factor (BAFF), a B cell prosurvival factor, as well as upregulated salivary gland expression of lymphotixin (LT)-β mRNA. LT-β is required for the formation of lymph nodes and germinal centers, while the heterodimer LT-α/LT-β can induce the development of ectopic germinal center–like structures. LT-α in soluble form induces the secretion of IFNs and chemokines, which are elevated in the salivary gland tissue from patients with primary Sjögren’s syndrome. Levels of IL-14, another B cell growth factor, also appear to be increased in the serum and saliva of patients with primary Sjögren’s syndrome. Blocking LT-β receptor signaling in NOD mice ablates the lymphoid organization in the salivary glands and improves their function, further evidence that LTs may play a role in the pathogenesis of Sjögren’s syndrome. Finally, abnormal B cell behavior is implied in primary Sjögren’s syndrome by the predisposition of this disease toward the development of non-Hodgkin’s B cell lymphoma. Interestingly, IL-14 transgenic mice develop large B cell lymphomas later in life.

Anti-Ro(SS-A) and anti-La(SS-B) antibodies do not appear to have a pathogenic role in disease mechanisms despite their diagnostic and prognostic significance. The stimulating antigen(s) is unknown. An aberrant response to self may be provoked by altered expression of autoantigens. Several models have been proposed to explain the altered expression of autoantigens including differential expression of protein isofoms, post-translational modification, and abnormal autoantigen presentation via apoptotic blebs, exosomes, or heat shock protein–mediated cross-priming. Ro(SS-A) and La(SS-B) proteins do in fact show upregulated expression in the vicinity of the immunopathologic lesion in primary Sjögren’s syndrome, where they may induce local immune responses. Saliva from patients with this disease has been shown to contain anti-Ro(SS-A) and anti-La(SS-B) antibodies. Although this finding may be interpreted to reflect local production of autoantibodies, it may also reflect extravasation of proteins from blood vessels into the inflamed tissues.

Among the other autoantibody specificities, those directed against the muscarinic receptor (MR) have attracted the most interest because of their possible role in causing glandular hypofunction. The MR family of receptors is bound by acetylcholine, which mediates the effects of the preganglionic and postganglionic parasympathetic nerve fibers that primarily regulate salivary flow. The muscarinic type 3 receptor (M3R), the target of autoantibodies in primary Sjögren’s syndrome, is the subtype that predominantly controls salivary flow. Two lines of evidence support the hypothesis that anti-M3R antibodies reduce exocrine gland function. Firstly, serum immunoglobulins from patients with primary Sjögren’s syndrome have been shown to bind MR3 receptors on acinar cell membranes; they also have been shown to inhibit acetylcholine-evoked Ca²⁺ responses in a salivary gland cell line. This Ca²⁺-sensitive response is tightly regulated through intracellular signaling pathways that open the Cl⁻ channels on the apical membrane, resulting in an osmotic gradient and movement of water into the duct lumen. Secondly, passive transfer of the antibody has been shown to cause glandular hypofunction in NOD mice, an animal model of Sjögren’s syndrome. However, the assays for detecting serum anti-M3R antibodies have produced variable results depending on the methodology. For this reason, it has been difficult to determine the diagnostic sensitivity and specificity of serum anti-M3R antibodies in primary Sjögren’s syndrome and the possible relationship of these autoantibodies to glandular hypofunction. Other autoantigens have been identified in patients with primary Sjögren’s syndrome including α-fodrin, poly(ADP)ribose polymerase, carbonic anhydrase, and ICA69 protein. Despite initial enthusiasm for the possible etiologic role of α-fodrin in primary Sjögren’s syndrome, recent studies have shown that serum anti-α-fodrin antibodies are not highly specific for this diagnosis. Autoantibodies to carbonic anhydrase, poly(ADP)ribose polymerase, and ICA69 protein occur in only a minority of patients with primary Sjögren’s syndrome and therefore do not appear to be attractive candidates for the initiating autoantigens.

In primary Sjögren’s syndrome, the glandular epithelial cell appears to be an active participant in the abnormal immune response. In the salivary gland, the mononuclear cells preferentially congregate around ductal epithelium. The epithelial cells show upregulated expression of HLA class I and II molecules; adhesion molecules such as intercellular adhesion molecule-1 (ICAM-1; CD54), vascular cell adhesion molecule-1 (VCAM-1; CD106) and E-selectin; and co-stimulatory molecules such as CD80 and CD86; they also produce high levels of cytokine mRNAs for IL-1, IL-6, IL-12, IL-18, and TNF, as well as BAFF. Ductal epithelial cells have also been shown to express...
Figure 73-2  Two models for the pathogenesis of salivary gland inflammation in Sjögren’s syndrome. A, Glandular hypofunction is explained by tissue loss secondary to immune attack, resulting in cytotoxic cell death and apoptosis. Epithelial cells likely play a central role in this process by several mechanisms: antigen-presentation and T cell activation; production of proinflammatory cytokines such as interleukin-1 (IL-1), IL-6, tumor necrosis factor (TNF), IL-12, and IL-23; and secretion of proteases. Epithelial cells also upregulate expression of Fas (FAS) and Fas ligand (FASL), cell surface molecules involved in the activation of apoptotic pathways. Other immune cells such as T cells, B cells, macrophages, and dendritic cells serve to amplify the chronic inflammatory response. In model B, glandular hypofunction results from downregulation of receptor-mediated secretion of salivary fluid into the ductal lumen. Acetylcholine binds to muscarinic receptor type 3 (MR3) on the surface of acinar cells, stimulating production of the second messenger inositol 1,4,5-triphosphate (IP$_3$). IP$_3$ in turn diffuses through the cytoplasm until it binds its receptor, IP$_3$R, on the endoplasmic reticulum. This interaction causes calcium to be released into the cytoplasm, which opens Ca$^{2+}$-sensitive chloride channels on the apical membrane of the cell. Electrochemical neutrality is preserved when Na$^+$ follows chloride across the membrane, while the osmotic gradient propels the water into the ductal lumen. Mechanisms of immune-mediated glandular hypofunction might include inhibition of acetylcholine release by cytokines, increased breakdown of acetylcholine in the epithelial space by upregulated production of cholinesterases, blockade of M3R by autoantibodies, inhibition of intracellular signaling pathways involved in the fluid secretory process, or altered expression of aquaporin 5 (AQP5), which appears to be primarily responsible for water movement through the apical cell membrane. See reference 96 for more details. BAFF, B cell activating factor; ICAM-1, intercellular adhesion molecule-1; MHC, major histocompatibility complex; MMP-9, matrix metalloproteinase-9; PLC, phospholipase C.
mRNAs for several different proinflammatory chemokines including CXCL13, CCL7, CCL21, and CCL22. Cultured salivary gland epithelial cells also express functional TLR3 and TLR7 molecules, endowing a capability to sense pathogens and endogenous molecules produced by injured tissues. Activated epithelial cells are likely driving the aberrant innate and adaptive immune responses not only in the lacrimal and salivary glands but also in the lungs and kidneys, as well as other extraglandular sites.

At least two models exist to explain the glandular hypofunction in Sjögren’s syndrome. In one model, it may be hypothesized that the glandular tissue is destroyed by an immune onslaught perpetuated by persistent exposure to self-antigens or other environmental stimulants (e.g., viral infection), leading to apoptosis of acinar epithelial cells and the irreversible loss of salivary gland function (Figure 73-2A). A weakness of this model is that studies suggest epithelial cells rarely undergo apoptosis in the salivary gland tissue despite the fact they upregulate expression of mediators of cell death such as Fas, Fas ligand, and Bax (B cell lymphoma 2–associated X protein). Ligand binding to CD40 expressed on epithelial cells in the salivary gland tissue also leads to Fas-mediated cell death by down-regulating c-FLIP (cellular FLICE-like inhibitor protein), an inhibitor of Fas-mediated cell death. Because no studies have yet shown that an imbalance truly exists between the rates of glandular proliferation and damage, the relevance of this model to glandular hypofunction is uncertain. It also does not explain why many patients with markedly diminished salivary flow retain substantial amounts of normal-appearing acinar tissue in their salivary glands.

A second model assumes apoptosis is not a significant factor in the loss of acinar epithelium and instead posits that glandular function is inhibited (and not destroyed) by immune-mediated mechanisms (Figure 73-2B). This model implies a reversibility component to the loss of salivary flow and a pathologic process interrupting M3R activation. Possible mechanisms for such inhibitory effects include a reduction in acetylcholine release, increased breakdown of acetylcholine in the epithelial space (e.g., acetylcholine must diffuse 100 nm from its release point at the nerve ending to the receptor on the cell), and blockade of M3R by antibodies. Relatively little direct evidence supports any of these mechanisms except for the experiments showing that anti-M3R antibodies may block receptor function. Nevertheless, it is possible that the cytokine milieu in the inflamed glandular tissue interferes with the mechanisms controlling acinar gland function. In the NOD mouse, for example, impairment of salivary gland secretion does not appear to be readily explained by the destruction of acinar glands. Regardless, the “immune destructive” and “immune inhibited” models are not mutually exclusive and may contribute to glandular hypofunction at various stages of the disease.

**CLINICAL FEATURES**

**Keratoconjunctivitis Sicca**

Chronic inflammation of the lacrimal glands diminishes secretion of aqueous tears, which if severe, may destroy the conjunctival and bulbar epithelium. Aqueous tear deficiency produces a dry eye, causing symptoms of grittiness or foreign body sensation, burning, photophobia, and eye fatigue. Routine eye examination usually reveals a reduction in the tear flow, as measured by the Schirmer-I test. Additional findings may include the absence of tears in the conjunctival sac and dilated bulbar conjunctival vessels. Thick mucus may also be seen in the inner canthus of the eye. Slit-lamp examination allows for more detailed visualization of the corneal and conjunctival surface. Following instillation of lissamine green dye or fluorescein onto the ocular surface, slit-lamp examination may expose devitalized cells or epithelial defects, respectively, signs of corneal and conjunctival damage. Severe dryness may result in corneal abrasion or ulceration.

The anatomy of the tear film informs the differential diagnosis of a dry eye (Table 73-3). The tear film consists...
<table>
<thead>
<tr>
<th>Condition</th>
<th>Associated Clinical/Radiologic/Pathophysiologic Features</th>
<th>Immunologic/Histopathologic Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evaporative (Excessive Water Loss from the Ocular Surface)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meibomian gland deficiency</td>
<td>Ocular redness and burning; increased tear break-up time and tear hyperosmolarity; secondary to posterior blepharitis, acne rosacea, seborrheic dermatitis, and drugs (e.g., retinoids)</td>
<td></td>
</tr>
<tr>
<td>Cicatricial pemphigoid‡</td>
<td>Scarring of the meibomian orifices, loss of goblet cells</td>
<td></td>
</tr>
<tr>
<td>Poor lid congruity or low blink rate</td>
<td>Proptosis (poor lid congruity), Parkinson’s disease (low blink rate)</td>
<td></td>
</tr>
<tr>
<td>Vitamin A deficiency</td>
<td>Impaired goblet cell development</td>
<td></td>
</tr>
<tr>
<td><strong>Dry Mouth (Hyposalivation)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amyloidosis</td>
<td>Nephrotic syndrome</td>
<td>ANA negative; monoclonal gammopathy; biopsy often shows amyloid deposits</td>
</tr>
<tr>
<td>Hemachromatosis</td>
<td>Cirrhosis</td>
<td>Labial salivary gland biopsy with heavy hemosiderin deposits in the ductal epithelium and loss of acinar glands</td>
</tr>
<tr>
<td>Chronic GVHD</td>
<td>Salivary gland dysfunction less common than lacrimal gland dysfunction; may be associated with oral mucosal involvement of chronic GVHD</td>
<td>Periductal inflammation, fibrosis, and glandular atrophy</td>
</tr>
<tr>
<td>Chronic hepatitis C infection</td>
<td>See above</td>
<td>Labial salivary gland biopsy shows chronic inflammation similar to that of Sjögren’s syndrome</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>Type I or type II diabetes</td>
<td>Lack of functional acinar cells; damage to ducts, blood vessels, and nerves</td>
</tr>
<tr>
<td>Radiation therapy</td>
<td>Side effects after radiotherapy for head and neck cancer</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systemic drugs†</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parotid Gland Enlargement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarcoïdosis*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Igg4-related diseases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multicentric Castleman’s disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetic mellitus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcoholism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV lymphoepithelial lesion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benign tumors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malignant carcinoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary B cell lymphoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary T cell lymphoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculous duct obstruction</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Table 73-3 Differential Diagnosis of Primary Sjögren’s Syndrome—cont’d</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*May coexist with primary Sjögren’s syndrome.
†Common offending drugs include: anticholinergics (e.g., antihistamines, tricyclic antidepressants, antispasmodics), clonidine, diuretics, isotretinoin, estrogen replacement therapy, amiodarone.
‡May also destroy the lacrimal gland and cause lacrimal duct obstruction late in the clinical course.
ANA, antinuclear antibodies; GVHD, graft-versus-host disease; HIV, human immunodeficiency virus; HTLV-1, human T lymphotropic virus type 1; IL-6, interleukin-6; ILD, interstitial lung disease; MALT, mucosa-associated lymphoid tissue.
of three major layers: the outer lipid layer, the middle aqueous layer, and the inner mucin layer. In addition to Sjögren’s syndrome, conditions associated with infiltration of the lacrimal glands (e.g., sarcoidosis) and diminished tear flow (e.g., medications, aging, and estrogen deficiency) may decrease aqueous tear flow. The lipid layer derives from the meibomian glands and traps the aqueous tear film on the eyeball and protects it from rapid evaporation. Meibomian gland dysfunction, or posterior blepharitis, produces dry eyes from rapid evaporation of tears; it may be present with aqueous tear deficiency and be an aggravating factor in patients with keratoconjunctivitis sicca. Meibomian gland dysfunction is often associated with ocular rosacea and seborrheic dermatitis, two conditions encountered often in clinical practice that lead to symptoms of dry eyes. Lipid degradation resulting from meibomian gland inflammation may produce free fatty acids, which are irritating to the ocular surface and may cause punctate keratopathy. The mucin layer originates from the goblet cells of the conjunctiva and, if deficient, leads to an uneven distribution of the tear film over the surface of the eye. Vitamin A deficiency and Stevens-Johnson syndrome are examples of conditions associated with an abnormal mucin layer.

**Xerostomia**

The changes in the quality and quantity of the saliva are responsible for the signs and symptoms of xerostomia. Although symptoms of dry mouth are relatively common in the general population, they are usually more severe in Sjögren’s syndrome and cause incessant difficulties with chewing and swallowing dry food, altered taste (metallic, salty, or bitter), and prolonged speaking. Patients with xerostomia may have problems wearing dentures. Despite complaints of a dry mouth, many patients will appear to have a normal oral examination owing to residual salivary flow. Others with more severe hypofunction will manifest a dry, sticky, or erythematous oral mucosa.

Two complications of xerostomia are important to the care of patients with xerostomia, especially those with severe deficits in salivary flow. Xerostomia often results in rampant dental caries, cracked teeth, and loose fillings. Oral candidiasis, the other frequent complication, typically manifests as the atrophic variant, which is characterized by erythema and atrophy of the oral mucosa and filiform papillae on the dorsum of the tongue, with angular cheilitis. Sometimes, a thin, white exudate may appear on the surface of the tongue. The “thrushlike” variant of oral candidiasis is seen much less frequently in patients with xerostomia except in the face of recent antibiotic therapy.

About one-quarter of patients with primary Sjögren’s syndrome suffer enlargement of the parotid or submandibular glands during the course of their disease. Chronic swelling is usually painless and may be unilateral or bilateral; it is often diffuse and firm by palpation.Transiently painful episodes of acute swelling and tenderness may also punctuate the clinical course. Acute swelling of the major salivary glands probably results most often from dried mucus obstructing the lumens of the major ducts; it usually subsides within a few days with conservative therapy. Rarely, bacterial infections may cause acute salivary gland swelling and should be considered as a possible etiology if the patient has a fever or other constitutional complaints. Asymmetric gland enlargement with palpable hard nodules that are increasing in size may indicate a neoplasm such as a lymphoma.

**Involvement of Other Exocrine Glands**

Glandular hypofunction may also affect the nasal passages (meatal obstruction from dried mucus), larynx (hoarseness), trachea (cough), vagina (dyspareunia), and skin (pruritus), producing symptoms of dryness. Virtually any exocrine gland may be involved in this disease.

**Extraglandular Manifestations**

Nearly three-quarters of patients with primary Sjögren’s syndrome manifest signs or symptoms of extraglandular disease. Extraglandular involvement is more likely to occur in patients with serum anti-Ro/SS-A and anti-La/SS-B antibodies, as well as hypergammaglobulinemia, cryoglobulinemia, and hypocomplementemia. However, only about 25% of patients with primary Sjögren’s syndrome develop moderate or severe extraglandular disease.

Fatigue, a complex and multifaceted phenomenon, occurs in approximately 70% of patients with primary Sjögren’s syndrome. It is also a symptom of depression, chronic anxiety, fibromyalgia, and sleep deficit, as well as a side effect of certain medications. The United Kingdom Sjögren’s Interest Group has developed an instrument called the Profile of Fatigue and Discomfort–Sicca Symptoms Inventory (PROFAD-SSI) to specifically measure both somatic (needing rest, poor starting, low stamina, weak muscles) and mental fatigue (poor concentration, poor memory) in this disease. In 547 patients with a confirmed diagnosis of primary Sjögren’s syndrome, somatic fatigue was the dominant predictor of physical function and general health. The relative contribution of behavioral and cognitive variables to fatigue was studied in 94 patients with primary Sjögren’s syndrome. Although depression was associated with higher levels of fatigue in this study, it was not present in most patients who reported fatigue. The link between fatigue and “biologic disease activity” is not fully understood in primary Sjögren’s syndrome. Increased serum cytokines such as IL-6 and type I IFNs, as well as neuroendocrine and autonomic dysfunction are postulated to be contributing factors to the physical and mental aspects of fatigue in this setting.

Raynaud’s syndrome has been reported in 13% to 33% of patients with primary Sjögren’s syndrome, often preceding the onset of sicca symptoms by several years. Digital ulcers occur only rarely.

Among the dermatologic manifestations, the most common are xerosis, or dry skin; eyelid dermatitis; and angular cheilitis. In addition, many patients develop a variety of other cutaneous manifestations including annular erythema, purpura, and urticarial vasculitis. Annular erythema has been described in several forms: a donut-ring-like erythema with an elevated border (type I); a subacute cutaneous lupus erythematosus (SCLE)-like lesion with marginally scaled polycyclic erythema (type II); and a papular insect bite–like erythema (type III). Histopathologically, these lesions are characterized by a deep perivascular lymphocytic infiltrate without the epidermal changes associated with lupus.
in some cases, immunoglobulin and complement deposition is observed along the basement membrane with liquefaction degeneration in the basal layer of the involved skin. The type I lesion appears to be specific for primary Sjögren’s syndrome, occurring predominantly in Asian but not Western populations.

Cutaneous vasculitis may be expressed in several different forms including palpable purpura, erythematous papules or macules, and ulcers, with lesions predominantly located in the lower limbs. Such lesions are frequently accompanied by other extraglandular manifestations of disease such as arthritis, peripheral neuropathy, Raynaud’s syndrome, anemia, elevated erythrocyte sedimentation rate, hypergammaglobulinemia, serum rheumatoid factor, and serum anti-Ro/SS-A and anti-La/SS-B antibodies. In one study, 27% and 50% of the cases of small vessel, cutaneous vasculitis were present with and without cryoglobulinemia, respectively; 21% of cases were classified as urticarial vasculitis. Livedo reticularis has also been described in this setting.

Polyarthralgia occurs frequently in patients with primary Sjögren’s syndrome. In a retrospective study, articular symptoms in primary Sjögren’s syndrome had a prevalence rate of 45%. Although most patients complain only of polyarthralgia, a subset may develop objective signs of synovitis, which is nonerosive, symmetric, and polyarticular of waxing and waning intensity. Joint symptoms may precede the diagnosis of primary Sjögren’s syndrome in up to one-third of cases. In two separate studies, anticyclic citrullinated peptide (anti-CCP) antibodies were detected in serum from 7.5% and 9.9% of patients with primary Sjögren’s syndrome. However, in only one of these studies was their presence closely associated with synovitis, and in neither study did it appear that serum anti-CCP positivity was associated with radiographic erosions or progression to rheumatoid arthritis.

Involvement of the airways and lung parenchyma in primary Sjögren’s syndrome may take several forms including xerotrachea and xerobronchitis; nonspecific interstitial pneumonitis (NSIP); lymphocytic interstitial pneumonitis (LIP), now considered to be a subset of NSIP; usual interstitial pneumonitis (UIP); bronchiolitis; and lymphoma (reviewed in reference 110). The estimated prevalence of pulmonary involvement in primary Sjögren’s syndrome varies depending on the thoroughness of the evaluation. In one study involving 123 patients with primary Sjögren’s syndrome, 11.4% showed pulmonary signs or symptoms and/or impaired pulmonary function with abnormal chest computed tomography (CT) findings at the time of evaluation.

Evidence of small airways dysfunction is found often in asymptomatic patients with normal radiologic studies. In symptomatic patients, NSIP appears to be the predominant type of lung involvement. The diagnosis can often be made on the basis of clinical presentation, pulmonary function tests (PFTs), and abnormal chest CT findings. PFTs in patients with NSIP show a restrictive pattern with reduced diffusion capacity of lung carbon monoxide (DLCO). Chest CT scans in this condition reveal ground-glass opacities and a reticular nodular pattern. Bronchoalveolar lavage (BAL), which is not usually required for diagnosis, shows evidence of alveolar inflammation, with elevated neutrophil or lymphocyte counts, or both. LIP is also associated with a restrictive pattern on PFTs. Chest CT findings are ground-glass opacities and thin-walled cysts, with centrilobular nodules, interlobular septal thickening, and bronchovascular bundle thickening (Figure 73-3A). Microscopically, the lung biopsy from patients with LIP shows a diffuse interstitial infiltrate composed of lymphocytes, plasma cells, and histiocytes, which expand the interlobular and alveolar spaces (Figure 73-3B). The PFTs in patients with UIP also show a restrictive pattern like NSIP and LIP, but the chest CT findings of lower lobe fibrosis, honeycombing, and traction bronchiectasis distinguish it from these other two types. Histopathologically, follicular bronchiolitis is
A case has also been described of a patient with primary Sjögren’s syndrome who developed esophageal achalasia in the setting of a sensory ataxic neuropathy, which damaged the myenteric plexus and led to the esophageal abnormality. In addition, chronic atrophic gastritis, which may cause dyspeptic symptoms, has been reported in a small number of cases of primary Sjögren’s syndrome.

Patients with primary Sjögren’s syndrome may develop liver enzyme abnormalities for a variety of reasons, most often from an associated disorder such as hepatitis C virus infection, autoimmune hepatitis, primary biliary cirrhosis, or a nonspecific hepatitis. Bowel symptoms such as abdominal pain and constipation occur more commonly in patients with primary Sjögren’s syndrome than healthy controls, but their etiology is often obscure.

Neurologic abnormalities are protean in primary Sjögren’s syndrome, with varied patterns of peripheral and central nervous system involvement. The prevalence of central nervous system involvement that can be directly attributed to primary Sjögren’s syndrome is likely to be in the range of 1% to 2%, but much higher rates have been reported with more liberal case definitions. For example, higher prevalence rates are noted when mood disturbances and minor cognitive and affective disturbances are included in the definition of central nervous system (CNS) involvement. Neuropsychiatric symptoms such as depression and minor cognitive disturbances occur in approximately one-third of patients with primary Sjögren’s syndrome. However, they are nonspecific clinical disorders that occur frequently in the general population and in patients with other chronic diseases. Rarely, cases have been described in which patients with primary Sjögren’s syndrome develop severe cognitive dysfunction. Brain magnetic resonance imaging (MRI) scans in these instances typically show nonspecific T2-weighted, high-intensity signals in the white matter that have an uncertain relationship to the clinical CNS findings. Focal CNS deficits have also been rarely described in patients with primary Sjögren’s syndrome and include optic neuropathy, hemiparesis, movement disorders, cerebellar syndromes, recurrent transient ischemic attacks, and motor neuron syndrome. Spinal cord syndromes resembling multiple sclerosis such as transverse myelitis and progressive myelopathy have also been reported in patients with primary Sjögren’s syndrome. Patients with primary Sjögren’s have been described, too, with neuromyelitis optica in association with serum anti-aquaporin-4 antibodies.

Peripheral nervous system involvement is among the most common of the extraglandular features of primary Sjögren’s syndrome. In a cross-sectional study, peripheral neuropathy was diagnosed in 17 (27%) of 62 patients with primary Sjögren’s syndrome on the basis of a conventional neurologic examination. However, only 34 (55%) in this group had abnormal nerve conduction velocity studies including 19 (31%) with a motor neuropathy, 8 (13%) with a sensory neuropathy, and 7 (11%) with a sensorimotor neuropathy. Some of the others with normal nerve conduction velocity studies may have had a small fiber neuropathy. Other peripheral neuropathies that have been reported in patients with primary Sjögren’s syndrome are cranial neuropathies, autonomic neuropathies, and multiple mononeuropathies. Most cases of peripheral neuropathy are dominated by sensory symptoms and typically do not progress to cause motor weakness.
Beyond lower extremity purpura, systemic vasculitis appears to be a rare manifestation of extraglandular disease. Patients can develop a small vessel vasculitis and cryoglobulinemia in the absence of hepatitis C virus infection and a medium-sized vessel vasculitis with features ranging from multiple mononeuropathies to ischemic bowel.

**Lymphoma**

Non-Hodgkin’s lymphoma (NHL) is a complication of primary Sjögren’s syndrome with important prognostic significance. In a recent study, the prevalence of NHL in this disease was 4.3%, with a median time of approximately 7.5 years from the diagnosis of primary Sjögren’s syndrome to the development of NHL. Several histopathologic types of NHL have been described in association with primary Sjögren’s syndrome including marginal zone B cell lymphoma, follicular cell lymphoma, diffuse large B cell lymphoma, and lymphoplasmacytoid lymphoma. Marginal zone B cell lymphoma, which is a family of low-grade B cell lymphomas, is by far the predominant type of NHL associated with this chronic autoimmune disease.

The mucosa-associated lymphoid tissue (MALT) lymphoma, a type of marginal zone B cell lymphoma, occurs mostly in chronic autoimmune diseases such as primary Sjögren’s syndrome. It develops at extranodal sites in relation to mucosae or glandular epithelium such as the lacrimal and salivary glands, lung, gastrointestinal tract, and skin. The earliest histopathologic feature of MALT lymphoma is the finding of monocytoid B cells surrounding the epithelium. Immunohistochemical staining of these lesions shows their clonality in revealing monotypic infiltrates of either Igκ or Igλ-light chains. With progression, these lesions show increasing proliferation of neoplastic cells, replacement of reactive follicles, and ductal dilatation. In primary Sjögren’s syndrome, MALT lymphomas most often develop in the salivary glands, but they may also develop at other extranodal sites, especially the lung and gastrointestinal tract. In primary Sjögren’s syndrome, a fivefold increased risk for lymphoma is conferred by the presence of parotid gland enlargement, splenomegaly, lymphadenopathy, neutropenia, cryoglobulinemia, or low C4.

**Associated Diseases**

Primary Sjögren’s syndrome has been associated with a higher risk for the development of other autoimmune diseases including thyroid disease, autoimmune hepatitis, primary biliary cirrhosis, and celiac disease. Although initial studies found a higher frequency of thyroid disease in patients with primary Sjögren’s syndrome than controls, a recent, larger study failed to show a statistically significant difference in the rate of thyroid disease between the two groups. Autoimmune hepatitis and primary biliary cirrhosis occur in less than 5% of patients with primary Sjögren’s syndrome, but more precise estimates of their prevalence are not available owing to the absence of well-controlled studies addressing this question. One exception is a Hungarian study showing a 10-fold higher rate of celiac disease in patients with primary Sjögren’s syndrome compared with healthy controls.

**DIAGNOSIS AND DIAGNOSTIC TESTS**

**Keratoconjunctivitis Sicca and Xerostomia**

A diagnostic suspicion of Sjögren’s syndrome is initially raised by complaints of dry eyes and dry mouth. However, many patients do not openly volunteer this information because they do not feel it is important enough to bring to the attention of the physician. Therefore it is essential to inquire about these symptoms if they do not come up during initial history taking. In patients with symptoms of dry eyes and dry mouth, the next step is to confirm the diagnosis by objective testing. An easy method for evaluating aqueous tear flow is the Schirmer’s-I test, which is performed by inserting a sterile piece of filter paper in the middle to lateral third of the lower eyelid and measuring the distance tears elute over 5 minutes. A distance of 5 mm or less is the usual cutoff for abnormally low tear production. Although a Schirmer-I test is subject to an approximately 20% rate of false negativity, normal tear flow in the face of ocular foreign body sensations suggests an alternative diagnosis such as blepharitis (see later). Patients with moderate-to-severe symptoms of dry eyes are usually referred to an ophthalmologist for slit-lamp examination. With a slit-lamp, the ophthalmologist can carefully examine the ocular surface, evaluating for any signs of damage. Instillation of lissamine green or fluorescein dye onto the surface of the eye displays the integrity of the conjunctival and corneal surface. Lissamine green dye stains epithelial surfaces lacking mucin (Figure 73-4), whereas the fluorescein dye targets areas of cellular disruption on the ocular surface. Rose Bengal dye, which stains dead or degenerated cells, is no longer preferred for evaluating the ocular surface due to its toxic effects on the cornea.

Several methods may be used for objective assessment of dry mouth, or xerostomia. Only a minority of patients with xerostomia show obvious signs of a dry mucosa on oral examination (e.g., absent sublingual salivary pool, thick and sticky saliva). Sialometry is used to test for functional impairment in salivary flow and refers to the measurement of salivary flow from individual glands (parotid, submandibular, or sublingual) or the mouth as a whole. An unstimulated whole salivary flow rate less than or equal to 1.5 mL/15 min meets the criterion for xerostomia according to the classification criteria developed by the American-European Consensus Group (see Table 73-1). For the collection procedure, the patients should keep their heads tilted forward and swallow once to clear the mouth of excess saliva. At this point, the 15-minute collection period is initiated and then subjects expectorate, as needed, accumulating saliva into a preweighed 50-cm³ cryovial. Samples are weighed on an analytic balance to determine the volume (1 g = 1 mL) of saliva.

Sialography, another technique for evaluating xerostomia, visualizes radiographic patterns of duct obstruction in the major salivary glands. It is most useful for differentiating an inflammatory process from a neoplasm. This technique calls for injection of radiographic contrast dye into the salivary duct followed by serial radiographs to image the pattern of dye flow. A water-based contrast is strongly preferred to an oil-based contrast because the latter may damage the adjacent salivary gland tissue. Sialography is not routinely...
employed in clinical practice because it is an invasive procedure with complications including duct rupture, pain, and infection. Scintigraphy is a radionuclide technique for measuring salivary gland function. After sodium radiolabeled pertechnetate technetium is injected into the blood, it is absorbed into the salivary gland and secreted into the mouth, allowing for determination of the salivary flow rate. The diagnostic sensitivity and specificity of salivary gland scintigraphy for the diagnosis of primary Sjögren’s syndrome has been estimated to be 75% and 78%, respectively.

Salivary scintigraphy is not widely available for routine testing.

Ultrasonography and MRI have been examined for their ability to detect anatomic abnormalities in the salivary glands of patients with primary Sjögren's syndrome. In comparison with normal individuals, the detection of parenchymal inhomogeneity by ultrasonography in two or more major salivary gland had a sensitivity of 63% and a specificity of 99% for identifying patients with primary Sjögren's syndrome. Although these results are promising, this technique requires further validation and investigation using appropriate disease controls. MRI with sialography may be more sensitive than ultrasonography for detecting glandular structural changes, but further studies are necessary to validate this technique as well.

**Labial Salivary Gland Biopsy**

Labial salivary gland biopsy has long been considered the gold standard for diagnosing primary Sjögren's syndrome. However, in clinical practice, its use is often reserved for patients in whom the diagnosis remains unclear after thorough clinical and laboratory evaluation. The biopsy is usually performed by an oral surgeon or otolaryngologist, or other appropriately trained individual. This minor procedure calls for removal of four or more salivary gland lobules through a small incision in the inner lip. A biopsy is considered to be positive if histopathologic analysis shows a focus score of greater than or equal to 1 per 4 mm² of tissue, where a focus is defined as a cluster of 50 or more lymphocytes. Using a focus score of greater than or equal to 1 as the cutoff, the sensitivity and specificity of labial salivary gland biopsy for the diagnosis of primary Sjögren’s syndrome was reported to be 83.5% and 81.8%, respectively. The interpretation of labial salivary gland biopsies is subject to considerable inter-reader variability depending on the experience of the reader. Therefore it is recommended that the biopsy slides be read by an experienced pathologist or other specialist with an appreciation for the nuances of interpretation.

**Laboratory Evaluation**

Most patients with primary Sjögren’s syndrome test positive for serum ANAs. In a large study from Spain, ANAs were detected in sera from 85% of patients with primary Sjögren's syndrome. Approximately one-half and one-third of patients with primary Sjögren’s syndrome in this study had anti-Ro(SS-A) and anti-La(SS-B) antibodies, while about 50% of patients tested positive for rheumatoid factor. There also appears to be a small subset of patients (<5%) with primary Sjögren’s syndrome that test negative for serum anti-Ro(SS-A) and/or anti-La(SS-B) antibodies, but positive for serum anti-centromere antibodies. In one study, such patients with serum anticentromere antibodies had a lower prevalence of dry eyes, hypergammaglobulinemia, and anti-Ro(SS-A) and anti-La(SS-B) antibodies than other patients with primary Sjögren's syndrome, as well as a higher prevalence of Raynaud’s syndrome.

About 5% to 10% of patients with primary Sjögren's syndrome have low blood levels of C3 and C4. For example, 9% of the Spanish cohort referred to earlier had low C3 or C4 levels (13%). About the same proportion of patients with primary Sjögren’s syndrome has either type II or III cryoglobulinemia or a monoclonal gammopathy. In type II cryoglobulinemia, the monoclonal gammopathy is usually an IgMκ that has rheumatoid factor activity when the patient resides in a Western or European country; patients of Japanese decent express a higher prevalence of IgA and IgG monoclonal gammopathies. Hematologic abnormalities are observed in about 5% to 15% of patients with primary Sjögren’s syndrome and include leukopenia and thrombocytopenia.

**Approach to Diagnosis**

The clinical diagnosis of primary Sjögren’s syndrome is highly likely if a patient with symptoms of dry eyes and dry
mouth has objective evidence of keratoconjunctivitis sicca and/or xerostomia in conjunction with a positive test for serum anti-Ro/SS-A and/or anti-La/SS-B antibodies. It may not be necessary to obtain objective evidence of salivary gland involvement if the other clinical and laboratory features are consistent with this diagnosis. In some cases, a clinical diagnosis of primary Sjögren’s syndrome may be made in the absence of serum anti-Ro/SS-A and/or anti-La/SS-B antibodies but a moderately or strongly positive test for ANAs. Although accepting a positive ANA alone as the serologic criterion lowers diagnostic specificity, the other clinical and laboratory features can be taken into consideration when making a diagnosis of primary Sjögren’s syndromes as opposed to another connective tissue disease or related condition. Because approximately 25% to 50% of patients with primary Sjögren’s syndrome test negative for anti-Ro/SS-A and anti-La/SS-B antibodies, a positive test for ANAs may often be the only evidence of an immune-mediated process. Serum ANAs of low titer have relatively little diagnostic value in this setting.

Biopsy of the labial salivary glands may be warranted in cases where signs and symptoms of keratoconjunctivitis sicca and xerostomia are not accompanied by convincing serologic evidence of autoimmunity. The labial salivary gland biopsy is used in this scenario to confirm that glandular hypofunction is associated with a chronic inflammatory process and to explore alternative diagnoses such as sarcoidosis or amyloidosis. In most of these situations, relatively few labial salivary gland biopsies will prove to be positive (focus score ≥1), although a negative result probably reduces the likelihood of primary Sjögren’s syndrome to less than 5%.

The diagnosis of primary Sjögren’s syndrome in children and adolescents may require special consideration. Compared with adults, children and adolescents present more often with recurrent parotitis, which is not included as a criterion in the recent classification schemes. It has been shown that the inclusion of recurrent parotitis in the classification criteria validated by the American-European Consensus Group increases the sensitivity of diagnosing primary Sjögren’s syndrome in children and adolescents from 39% to 76%.^10^

**Differential Diagnosis**

The differential diagnosis of primary Sjögren’s syndrome is broad owing to the variety of conditions that also cause dry eyes and mouth, salivary gland enlargement, and a similar profile of extraglandular manifestations (see Table 73-3). It is important to realize that most patients diagnosed with keratoconjunctivitis sicca by an ophthalmologist do not have primary Sjögren’s syndrome, despite a deficiency in aqueous tear flow. The etiology of keratoconjunctivitis in such cases is often obscure and presumably related to aging or hormonal changes or other degenerative processes. Evaporative tear loss secondary to meibomian gland dysfunction is another common cause of dry eyes. Many anticholinergic medications are contributing factors to the sicca complaints. Because approximately 25% of the general population has symptoms of dry mouth, it is not surprising that dry eyes and dry mouth often coexist in the same patient in the absence of a systemic disease such as Sjögren’s syndrome.

Chronic parotitis in association with the sicca complex strongly suggests the diagnosis of Sjögren’s syndrome or possibly another systemic disease. Sarcoidosis often comes up in the differential diagnosis of primary Sjögren’s syndrome and may even coexist with it. Bilateral diffuse enlargement of the parotid glands from sialadenosis, a noninflammatory process, also may occur in patients with diabetes mellitus who suffer from dry eyes and mouth. In addition, severe hypertriglyceridemia, chronic liver disease, and alcoholism have been associated with sialadenosis and enlargement of the major salivary glands.

A relatively new clinical entity, termed IgG4-related syndrome or IgG4 positive multiorgan lymphoproliferative syndrome, is emerging as an important consideration in the differential diagnosis of primary Sjögren’s syndrome. IgG4-related syndromes are uniquely characterized by high serum levels of IgG4 and tissue biopsies showing a marked infiltration of IgG4+ plasma cells coupled with fibrosis or sclerosis. Compared with primary Sjögren’s syndrome, IgG4+ related syndromes do not show the same female predominance and are associated with a lower frequency of dry eyes and mouth, arthralgia, and serum ANA positivity, as well as a higher rate of autoimmune pancreatitis.\(^{118}\)

**TREATMENT**

The treatment of primary Sjögren’s syndrome aims at reducing the signs and symptoms of dry eyes and dry mouth and ameliorating systemic manifestations (Figure 73-5). Sicca complaints may be lessened in general by withdrawal of medications with drying effects. For the treatment of dry eyes, the patient is advised to avoid central heating and air conditioning, windy environments, and medications that reduce tear and saliva production. Tear supplements are available over the counter in various formulations of different viscosities. Low-viscosity tears are administered more often than high-viscosity tear formulations, which are most useful at bedtime for prolonged lubrication during the sleeping hours. Artificial tears are also available with and without preservatives. Preservative-based tears may worsen dry eyes if they are instilled more than four times per day owing to their toxic effects on the ocular surface. More frequent instillation of tears warrants use of formulations without preservatives.\(^{139}\)

An ophthalmic preparation of cyclosporine 0.05% (Restasis) has been approved by the U.S. Food and Drug Administration for the treatment of keratoconjunctivitis sicca. In clinical trials, ophthalmic cyclosporine 0.05% has been shown to reduce the signs and symptoms of dry eyes including a statistically significant improvement in tear flow as measured by the Schirmer-I test. Many patients have trouble tolerating ophthalmic cyclosporine because of its burning effects.

Tears may be conserved by blocking the nasolacrimal drainage channels with temporary or permanent occlusion of the punctae, a relatively simple procedure performed by an ophthalmologist. This procedure may be performed in patients with persistent symptoms of dry eyes despite use of artificial tears. Plugs are initially placed in the two inferior punctae because 90% of the tears drain through these channels. Plugs may be later employed to block the two superior...
oral secretagogues are available by prescription that stimulate saliva and tear flow. Pilocarpine (Salogen), which acts by stimulating the M3R, is given at a dose of 5 mg three to four times daily and has been shown in controlled clinical trials to improve subjective and objective measures of both dry eyes and dry mouth.\textsuperscript{140,141} Cevimeline (Evoxac) also stimulates the M3R and has been similarly shown in trials to improve subjective and objective measures of dry eyes and mouth.\textsuperscript{142} Common side effects of these two secretagogues are sweating and flushing, which derive from their mechanism of action. Other possible side effects include visual disturbances, increased urination, nausea, abdominal pain, and diarrhea. Due to their pharmacologic properties, pilocarpine and cevimeline are contraindicated in patients with iritis, narrow angle glaucoma, and moderate-to-severe asthma.

Because xerostomia predisposes to dental carries and broken teeth, daily flossing, frequent dental visits, and regular fluoride applications are recommended for prevention purposes. Oral candidiasis, another complication of xerostomia, may be treated for 10 to 14 days with a nystatin elixir, clotrimazole troches, or fluconazole. Clotrim-

---

**Figure 73-5** Treatment algorithm for Sjögren’s syndrome. The treatment of Sjögren’s syndrome usually requires a multidisciplinary approach involving rheumatologists, ophthalmologists, dentists/oral surgeons, otolaryngologists, and other subspecialists depending on the extent of extraglandular disease. In all cases, it is prudent to minimize the use of medications that can exacerbate the symptoms of dryness such as antihistamines, antidepressants, muscle relaxers, and other drugs with anticholinergic properties. The treatment of extraglandular disease is individualized according to the site and severity of organ system involvement. The approaches indicated in the algorithm for the treatment of extraglandular disease are not supported by evidence from randomized, controlled trials, but rather from expert opinion based on retrospective case series and clinical experience. AZA, azathioprine; CNS, central nervous system; CYC, cyclophosphamide; KCS, keratoconjunctivitis sicca; LIP, lymphocytic interstitial pneumonitis; MMF, mycophenolate mofetil; MTX, methotrexate; NSAIDs, nonsteroidal anti-inflammatory drugs; NSIP, nonspecific interstitial pneumonitis; PNS, peripheral nervous system.

---

1. Somatic fatigue, arthralgia/arthritis, myalgia, palpable purpura without skin ulceration
2. NSIP or LIP, interstitial nephritis, PNS involvement with motor weakness, systemic necrotizing vasculitis, CNS involvement with focal deficits or severe cognitive dysfunction

---

punctae. Excessive tearing is the most frequent complication of punctal occlusion, which is the rationale for a trial of temporary blockade with dissolvable collagen plugs. Long-term blockage may be achieved using silicone plugs or surgery by cauterization or a laser.

Blepharitis, which commonly occurs in patients with dry eyes, often complicates a dry eye state. Patients with blepharitis may benefit from warm compresses and lid scrubs using an eyelid detergent such as Johnson’s Baby Shampoo mixed 1:1 with water. More refractory cases of blepharitis may be managed with chronic doxycycline therapy. Occasionally, the tear film may contain excessive mucous debris that can be dispersed by instillation of a 10% acetylcysteine ophthalmic solution, a mucoytic agent that can be compounded by a pharmacy.

A dry mouth may be managed by replacing existing saliva or stimulating residual salivary flow. Various over-the-counter preparations of artificial saliva are available containing hypermellose or methylcellulose. However, they provide limited relief in most cases because of their short duration of action. Patients may stimulate salivary flow by sucking on sugarless candy or chewing sugarless gum. Two oral secretagogues are available by prescription that stimulate saliva and tear flow. Pilocarpine (Salogen), which acts by stimulating the M3R, is given at a dose of 5 mg three to four times daily and has been shown in controlled clinical trials to improve subjective and objective measures of both dry eyes and dry mouth. Cevimeline (Evoxac) also stimulates the M3R and has been similarly shown in trials to improve subjective and objective measures of dry eyes and mouth.

Common side effects of these two secretagogues are sweating and flushing, which derive from their mechanism of action. Other possible side effects include visual disturbances, increased urination, nausea, abdominal pain, and diarrhea. Due to their pharmacologic properties, pilocarpine and cevimeline are contraindicated in patients with iritis, narrow angle glaucoma, and moderate-to-severe asthma.

Because xerostomia predisposes to dental carries and broken teeth, daily flossing, frequent dental visits, and regular fluoride applications are recommended for prevention purposes. Oral candidiasis, another complication of xerostomia, may be treated for 10 to 14 days with a nystatin elixir, clotrimazole troches, or fluconazole. Clotrim-
azole cream is beneficial for the treatment of angular cheilitis, which typically accompanies oral candidiasis.

Patients with primary Sjögren’s syndrome also suffer from dryness of the lips, skin, and nasal passages. This problem may be addressed by the frequent application of moisturizers, lip balm, and nasal saline spray. The oily content of vitamin E softgel capsules may be used topically on the skin or the lips for its moisturizing effects. Vaginal dryness with dyspareunia is a frequent complaint of women with primary Sjögren’s syndrome. This condition may respond to moisturizers and treatment with a topical estrogen cream. Vaginal lubricants may be necessary for treatment of dyspareunia.

Currently, the therapeutic armamentarium of primary Sjögren’s syndrome lacks a proven disease-modifying drug. Clinical experience suggests that symptoms of fatigue, myalgia, and arthralgia/arthritis may respond favorably to treatment with hydroxychloroquine; however, the efficacy of this approach has not been proven in a randomized, controlled clinical trial. Care is taken to avoid doses of hydroxychloroquine exceeding 6.5 mg/kg/day to minimize the risk of retinal toxicity. When this drug is used long term, annual ophthalmologic examinations are recommended to monitor for possible retinal toxicity.

Corticosteroids and other immunosuppressive drugs are often employed for the treatment of organ-threatening extra-glandular disease. NSIP and LIP are usually treated with high doses of corticosteroids and other immunosuppressive agents such as azathioprine, mycophenolate mofetil, or cyclophosphamide. Because virtually no controlled data are available supporting the use of these agents for NSIP or LIP, this approach is necessarily empiric and demands judicious monitoring and close follow-up. Bronchiolitis typically improves with high doses of corticosteroids alone, but close follow-up is mandatory for treatment of relapses. Patients who are asymptomatic and have had PFTs showing an isolated mild reduction in DLCO or evidence of small airway disease may be carefully followed without treatment.

Conservative measures may be the only therapy for a sensory peripheral neuropathy where medications such as gabapentin and analgesics may help to control the aggravating and painful neuropathic symptoms. High doses of corticosteroids may produce transient improvement in the symptoms of a peripheral neuropathy, but it is unclear if they provide any long-term benefits and if the benefits outweigh the risks of this treatment. Clinically demonstrable motor loss calls for a more aggressive treatment approach using corticosteroids and other immunosuppressive agents. Some patients with severe motor involvement may benefit from treatment with intravenous immunoglobulin.

A minority of patients with primary Sjögren’s syndrome develop recurrent lower extremity purpura. These lesions usually cause burning and stinging and may be associated with lower extremity edema, but they rarely ulcerate. Although moderate to high doses of corticosteroids afford symptomatic relief in most cases, subsequent tapering to low doses and withdrawal usually leads to recurrence of the purpura. Corticosteroid therapy may be avoided in the absence of severe and uncontrollable symptoms or progression to skin ulceration. Support stockings provide symptomatic relief in many cases and can be used in combination with analgesics, antihistamines, or nonsteroidal anti-inflammatory drugs to manage the painful symptoms. The patients are usually left with postinflammatory changes in the skin, which may be an acceptable outcome given the long-term side effects of corticosteroids.

Biologic therapies have been investigated recently for their clinical efficacy and safety in primary Sjögren’s syndrome. Infliximab and etanercept, which are TNF inhibitors, have failed in controlled clinical trials to demonstrate therapeutic benefit. Preliminary studies of rituximab and epratuzumab have produced mixed results so far, and larger studies will be necessary to provide more convincing evidence of clinical efficacy and safety for this indication. Other biologics are on the horizon and will likely be investigated for the treatment of primary Sjögren’s syndrome in the near future.

The interest in finding a disease-modifying therapy for primary Sjögren’s syndrome has led to the initial development of disease activity and damage measures, which will require subsequent validation in clinical trials.

OUTCOME

Overall mortality is not increased in patients with primary Sjögren’s syndrome compared with the general population, although the subgroup of patients with extra-glandular disease has an increased risk of morbidity and death. Compared with the general population, Ioannidis and colleagues showed that mortality was not significantly higher in a cohort of 723 patients with primary Sjögren’s syndrome. In this study the patients were subdivided into types I and II according to their risk for complications and death. Approximately 20% of the patients fell into the high-risk, type I group, whereas the remainder in the type II group had no increase in the risk for complications or death. In the type I group, patients presenting with palpable purpura and low C4 levels had a higher risk for long-term complications and death. Similar results were found in another study showing no increase in all-cause mortality among individuals with primary Sjögren’s syndrome. Hypocomplementemia has been confirmed as a risk factor for adverse outcomes in another cohort of 336 patients with primary Sjögren’s syndrome from Spain.

The disease mechanisms of primary Sjögren’s syndrome are being probed with increased interest. The labial salivary gland tissue is readily accessible for biopsy, and the new technologies are dramatically accelerating understanding of the complex pathways intertwining the aberrant innate and adaptive immune responses in this disease. Genetics and genomics are only beginning to permeate the scientific investigation of primary Sjögren’s syndrome, and predictably this gap will be filled in the near future. An unanswered question of major import is the cause of the glandular hypofunction. In the earlier stages of disease, it appears that glandular damage does not adequately explain the impaired salivary flow, suggesting that autoantibodies, the cytokine milieu, or both, or other mediators are contributing to these functional abnormalities. Further studies are necessary to address this question owing to the potential reversibility of glandular dysfunction and the nature and timing of therapeutic interventions to be tested in the future.
Although primary Sjögren’s syndrome cannot yet claim a disease-modifying therapy, much can be done for patients with this disease by effectively managing the sicca component, ameliorating fatigue and joint pain with hydroxychloroquine therapy, and monitoring for complications. Although biologics such as rituximab may yet prove to be useful for treating Sjögren’s syndrome, new therapeutic targets illuminated by translational research will shape the discovery of disease-modifying therapies down the road. Proof of therapeutic efficacy will require instruments that quantify disease activity and damage that are only now under development. Finding that first drug with the properties of disease modification will go a long way toward moving the field forward and expanding the therapeutic options for patients with primary Sjögren’s syndrome in the decades to come.

CONCLUSIONS

Primary Sjögren’s syndrome occurs nearly as often as rheumatoid arthritis and greater than 10 times more frequently than systemic lupus erythematosus. Therefore it ranks second in prevalence only to rheumatoid arthritis among the autoimmune disorders in the domain of rheumatology. Many patients with mild forms of primary Sjögren’s syndrome probably never come to the attention of the rheumatologist, but rather are managed exclusively by primary care physicians, ophthalmologists, otolaryngologists, and dentists. Others may never be referred for a diagnosis because of the frequent misperception that little can be done for patients with Sjögren’s syndrome. However, primary Sjögren’s syndrome is a systemic autoimmune disease associated with lung, neurologic, and renal involvement; vasculitis; and an increased risk for a low-grade B cell lymphoma, and the rheumatologist is aptly trained to provide an accurate diagnosis and deal effectively with the multisystem complications of this disease.

Selected References


References


Ankylosing spondylitis and reactive belong to the family of spondyloarthritis. Reactive arthritis (ReA), is clearly distinct in that it is induced by an episode of acute infection. Accordingly, the pathogenesis of these two diseases will be described in separate sections. For ankylosing spondylitis, it is commonly recognized that two major but “uncoupled” processes coexist inflammation and abnormal bone formation. The process of bone formation in this disease has become a major research frontier.

**PATHOGENESIS OF ANKYLOSING SPONDYLITIS**

The causes of ankylosing spondylitis are multifactorial and involves a number of interlinking pathways. A considerable number of critical factors have been identified. They can be classified into two groups. One group consists of mediators that are midstream in the pathways leading to inflammation. These mediators are targets of current therapies. The second group of critical factors contributing to ankylosing spondylitis is genetics. How should the degree of contribution of each factor be estimated? For therapeutics, the usefulness of each therapy will be ranked by a statistical value known as *effect size*. In pharmacology, effect size reflects the size of the difference in outcome between groups receiving a drug compared with a group receiving placebo, taking into account the degree of variability within groups. The higher the effect size, the more effective is the therapy. For ranking the genes, we will use the statistical value known as the *population attributable risk*. For each gene, this value represents the incidence of disease in a population that would be eliminated if the gene is absent. The higher the population attributable risk, the greater the contribution to the disease.

**Causes of Ankylosing Spondylitis That Are Targets of Therapies**

It has been known for more than three decades that most patients with ankylosing spondylitis demonstrate a highly satisfactory response to treatment with nonsteroidal anti-inflammatory drugs (NSAIDs), including the cyclooxygenase-2 (COX-2) inhibitors (coxibs). The effect size for spinal pain is 1.11, which means that there is a 77% probability that a randomly selected NSAID-treated patient will show a better response than a randomly selected placebo-treated patient. This effect size is much greater than in patients with nonspecific chronic low back pain. Because the molecular targets of most NSAIDs are COX-1 and COX-2, and the target of the coxibs is restricted to COX-2, it can be concluded that one of the pathways causing spinal pain in ankylosing spondylitis is the COX-2 pathway. Indeed, the COX-2 pathway generates prostaglandins and thromboxanes, which are strongly proinflammatory. The response of spinal pain in ankylosing spondylitis to NSAIDs is indirect evidence that inflammation is responsible for the pain. In fact, the spinal pain of ankylosing spondylitis is designated as being “inflammatory” rather than “mechanical.” The presence of spinal inflammation is validated by magnetic resonance imaging (MRI) observations. Although we know that the COX-2 pathway is responsible in part for the symptoms of ankylosing spondylitis, no information is available as to what triggers...
this pathway or which downstream mediators are most important.

When responses to NSAIDs are not satisfactory, some patients will be given a trial of a disease-modifying antirheumatic drug (DMARD). Most of the conventional DMARDs useful for rheumatoid arthritis such as methotrexate are much less effective in ankylosing spondylitis. Similar to rheumatoid arthritis, for patients who are resistant to treatment with NSAIDs or conventional DMARDs, there is a high rate of clinical response to biologically generated agents that target tumor necrosis factor (TNF). Currently, several of these biologics have been approved for the treatment of ankylosing spondylitis in the United States. All are very effective in controlling symptoms. The effect size for etanercept, for example, is 2.25, which means that more than 90% of randomly selected patients will respond better than another randomly selected patient who is treated with placebo. Hence there is little doubt that TNF is a major player in causing the symptoms of ankylosing spondylitis. TNF is a cytokine generated via innate and adaptive immunities by several types of cells, including macrophages, T lymphocytes, and mast cells. TNF is not a factor with a single pathway inducing a single event in a single type of cell. Rather, it is pleiotropic in affecting many cell types, and it induces a network of cytokines and chemokines and other mediators of inflammation. Although it is certain that TNF is a major early upstream mediator in ankylosing spondylitis, no agreement has been reached as to which cell types are specifically responsible for generating TNF, or which processes are disease-specific targets of TNF.

Information derived from the most effective modalities of treatment currently available for ankylosing spondylitis does not provide any concrete clues regarding fundamental causes. As with most chronic diseases, the fundamental causes must be a combination of environmental and genetic factors. What we first need to estimate in ankylosing spondylitis is the degree of contribution of environmental versus genetic factors.

### ASSESSMENT OF DEGREE OF CONTRIBUTION BY ENVIRONMENTAL VERSUS GENETIC FACTORS AND IDENTIFICATION OF GENETIC FACTORS

Degrees of contribution of environmental versus genetic factors have been estimated from the degree of concordance among twins. According to those analyses, genetics contributes to more than 90% of the total cause of ankylosing spondylitis. In addition, if there are environmental causes, they are probably ubiquitous, such as enteric bacteria. Indeed, rats carrying the human ankylosing spondylitis-causing transgene HLA-B27 do not develop arthritis when bred in a germ-free environment. However, arthritis will develop once they are transferred to a regular environment. It is thought that the development of arthritis in these rats is related to commensal bacteria, especially in the gastrointestinal tract. In human ankylosing spondylitis, the only enteric bacterium that has been incriminated is Klebsiella pneumoniae. As a group, patients with ankylosing spondylitis have higher mean antibody titers to Klebsiella compared with control subjects. However, evidence supporting Klebsiella as a cause of ankylosing spondylitis is far less strong than that which supports, for example, that streptococcal infections are a cause of rheumatic fever.

Over the past decade, most of the researchers studying the pathogenesis of ankylosing spondylitis have focused on identification of arthritis-causing genes. According to statistical modeling based on family studies, the disease is caused by up to nine genes with multiplicative interactions among loci. The most important gene, HLA-B27, was discovered in 1973. It contributes about 30% of the heritability of the disease. About three decades later, with the development of the technique of genome-wide association study of nonsynonymous single-nucleotide polymorphisms (SNPs), at least four other loci encoding known structural gene sequences have now been reported. In the Wellcome Trust Case Control Consortium and the Australo-Anglo-American Spondylitis Consortium studies, when these genes are ranked according to their degree of contribution, they are listed as follows: HLA-B27, ERAP1 (endoplasmic reticulum aminopeptidase 1, previously known as aminopeptidase-regulating tumor necrosis factor receptor shedding 1, abbreviated as ARTS-1), IL-23R (interleukin 23 receptor), IL-1R2 (interleukin 1 receptor, type II), and ANTXR2 (anthrax toxin receptor, also known as capillary morphogenesis protein 2, or CMG2). The population attributable risks for the first three are 90%, 26%, and 9%, respectively. Those for IL-1R2 and ANTXR2 are probably less than 5%. Hence, HLA-B27 is the largest contributor.

### How HLA-B27 Induces Ankylosing Spondylitis

The most important gene in ankylosing spondylitis is HLA-B27. In most populations, it is present in more than 90% of patients and in less than 10% of the general population. Rats carrying the HLA-B27 transgene develop arthritis and colitis. Using gene terminology from microbiology, it is described as being “essential” for ankylosing spondylitis. Because the association between HLA-B27 and ankylosing spondylitis has been known for more than three decades, the search for how HLA-B27 causes ankylosing spondylitis has been the Holy Grail of many scientists, and an enormous number of publications have been generated. HLA-B27 is an allele of the HLA-B locus of the human leukocyte antigen (HLA) class I antigens. Multiple subtypes of HLA-B27 have been defined. Most experiments attempting to identify the mechanisms of how HLA-B27 mediates arthritis have been carried out with the more common B27*05 and B27*04 subtypes. Multiple hypotheses have been reported regarding how HLA-B27 molecules mediate arthritis. At least three hypotheses are currently under active investigation.

### Arthritogenic Peptide Hypothesis

The arthritogenic peptide hypothesis is based on the classic structure and the canonical function of HLA-B alleles. As in all HLA class I molecules, a classic HLA-B27 molecule is a trimeric complex of a polymorphic HLA class I heavy chain together with a monomorphic light chain (β2-microglobulin [β2m]) and a single highly variable peptide (Figure 74-1). Most of the peptides are nanomers that
conform to a strict motif, usually with arginine as the second amino acid. The motif of peptides that bind to HLA-B27 is different from that of peptides that bind to other HLA-B alleles or to HLA-A or -C molecules. These polymorphisms probably are driven by evolution to survey against ever changing pathogens such as the influenza viruses. The canonical function of HLA-A and -B molecules is to present peptides derived from such intracellular pathogens to CD8+ T lymphocytes to generate a protective adaptive immune response. The reason why this is vulnerable to autoimmunity is that most of the peptides complexed with HLA-A and -B molecules are derived from self-proteins. In health, the hosts are tolerant to all these self-peptides.

The arthritogenic peptide hypothesis postulates that in the case of ankylosing spondylitis, there is a breakdown of tolerance to certain self-peptides, and this breakdown is a consequence of mimicry between the self-peptides and certain pathogen-derived and arthritis-causing peptides. Identification of these arthritogenic peptides demands a great deal of fundamental information concerning the quaternary structures of the HLA-peptide complexes, as well as the dynamics of these structures, the peptide repertoires of candidate pathogens, the repertoires of self-peptides, and last, the repertoires of T cell receptors. At this point, enormous detail has been generated concerning the structures of HLA-peptide complexes. Reactivity against several self-peptides has also been reported. Many yet untested candidate peptides have recently been identified and will require testing in the future, perhaps with newly developed immunoproteomic techniques.

**Free Heavy Chain Hypothesis**

The free heavy chain hypothesis is based on the observation that HLA-B27 molecules can exist on the cell surface as free heavy chains free of the more usual association with β2m or peptides with the classic HLA-B27 motif (see Figure 74-1). These free heavy chains exist as stable dimers and are capable of engaging allele-specific receptors on natural killer (NK) cells and T lymphocytes. Receptors for HLA-B27 free heavy chain dimers are the KIR3DL1*001 allele of KIRs (a family of immunoglobulin-like receptors on NK cells and certain subsets of T cells) and the LILRA1 and LILRB2 alleles of LILRs (a family of leukocyte immunoglobulin-like receptors on NK cells, T and B cells, and cells of the myeloid lineage). The free heavy chain hypothesis postulates that engagement of these receptors will generate arthritis-causing events. This hypothesis is supported by actual observations of ankylosing spondylitis patients. Such HLA-B27 heavy chain dimers are found on the cell surfaces of mononuclear cells, along with expansion of KIR3DL2-expressing NK and CD4+ T cells responding to such dimers.

**Unfolded Protein Hypothesis**

The unfolded protein hypothesis is different from the previous two in that it is concerned not with activities of HLA-B27 on the cell surface, but with activities inside the cells (see Figure 74-1). Like most surface proteins, the HLA-B27 heavy chains are synthesized linearly into a processing organelle inside the cell by the endoplasmic reticulum (ER). At first, these newly synthesized proteins do not have any conformation and are described as being "unfolded." Then, they are stepwise driven into a series of conformations through sequential complexing with a corresponding series of ER chaperones. These ER chaperones generate conformations in their target proteins by binding to their inappropriately exposed hydrophobic domains, as well as to underglycosylated residues. Through a series of cycles of binding and release with a series of ER chaperones, formation of disulfide bonds, and pairing with the β2m and a peptide, an HLA-B27 molecule will mature into a quaternary and stable form, which then will be transported to the cell surface.

This sequence of events takes place over a period of time. Compared with a few other common HLA alleles, HLA-B27 heavy chains have more prolonged retention times inside the ER. In addition, with HLA-B27, more partially misfolded and unfolded forms are found inside the ER compared with other HLA alleles. Reasons include aberrant disulfide bond formation and multimer formation among the heavy chains, among others. These partially folded or misfolded HLA-B27 proteins remain for a time sequestered inside the ER, being complexed with chaperones such as the "immunoglobulin heavy chain-binding protein."
How Non–Major Histocompatibility Complex Genes Modify HLA-B27 Physiology

Each of the three hypotheses of how HLA-B27 causes ankylosing spondylitis is based to a large extent on in vitro observations, to a lesser extent on observations in animal models, and even less on patient-derived samples. They are working hypotheses in the sense that they are not yet capable of being translated to the bedside for diagnosis or for development of therapeutics. However, strong support has appeared recently from studies of the biology of ankylosing spondylitis–associated genes identified by the genome-wide association study.

Probably the most important non–major histocompatibility complex (MHC) gene discovered so far is ERAP1. The association of ERAP1 with ankylosing spondylitis has been replicated in multiple populations, including white and nonwhite ethnicities. Equally significant, in one familial study, disease association has been extended to a haplotype of ERAP1 and ERAP2, indicating that the disease association is with a biology common to ERAP1 and ERAP2. ERAP1 and ERAP2 are metalloproteinases sharing considerable structural identity with one another. As described earlier, for HLA-B27 to mature into a trimolecular complex inside the endoplasmic reticulum, each heavy chain has to become associated with the monomorphic β2m and a polymorphic peptide. These peptides are derived from proteins that are degraded in the proteasomes outside the endoplasmic reticulum and are transported into the endoplasmic reticulum, often with lengths exceeding the HLA class I peptide motif. ERAP1 and ERAP2 are aminopeptidases, which degrade the peptides to suitable lengths to be accommodated into the HLA class I heavy chain–β2m complex. It is conceivable that alteration of the expression or the structures of these aminopeptidases can change the peptide repertoire of HLA-B27 or can lead to events postulated by the free heavy chain or the unfolded protein response hypotheses.

How Non–Major Histocompatibility Complex Genes Modify the Cytokine Network

As was described in a previous section, a major mediator of ankylosing spondylitis is TNF. TNF is situated relatively upstream mechanistically in a vast network of arthritis-causing cytokines such as IL-1, -6, and -17. This network is a self-organizing complex kept in balance with self-regulating inhibitors. Three of the non-MHC ankylosing spondylitis genes can potentially perturb such balance in the cytokine network. The activities of two of the ankylosing spondylitis genes—IL-1R and IL-23R—are self-evident. IL-23R is especially appealing to the pathogenesis of ankylosing spondylitis because it leads to the generation of IL-17, a cytokine that probably holds as pivotal a position as TNF. The other non-MHC ankylosing spondylitis gene, ERAP1, apparently has dual biologic activity. Its activities inside the ER have already been described earlier. When present on the cell surface, ERAP1 aids in release of the receptors for TNF, IL-1, and IL-6. It is possible that polymorphisms of ERAP1 can modify the balance of the cytokine network through this receptor-releasing activity, which favors the development of ankylosing spondylitis.
The link between inflammation and ankylosis, however, remains poorly understood.

Anatomic and Molecular Bases of Ankylosis

Histomorphologic studies of the spine are relatively rare owing to difficulty in accessing the tissues. Based on studies performed decades ago, several different mechanisms of new bone formation have been proposed. Both endochondral and direct bone (also known as membranous) formation appear to play a role. These two different types of bone formation have been best studied during development. The process of endochondral bone formation is critical for the development of the long bones and starts with proliferation and condensation of precursor cells. These cells subsequently differentiate into chondrocytes, which produce an extracellular matrix rich in type II collagen. Articular chondrocytes appear not to further differentiate and have a stable phenotype. Developmental and growth plate chondrocytes, on the other hand, progress into hypertrophic chondrocytes. These cells produce a matrix rich in type X collagen. This matrix gets calcified, is invaded by newly forming vessels, becomes broken down, and is gradually replaced by bone with active osteoblasts as the cellular component. Membranous or direct bone formation is important for the development of the skull and for cortical bone growth.

Unlike endochondral bone formation, in membranous bone formation, progenitor cells directly differentiate into osteoblasts. In ankylosing spondylitis, the shape of the syndesmophytes and their volume suggest that endochondral bone formation is more important than membranous bone formation. However, growth factors that steer endochondral and direct membranous bone formations belong to the same families, and the differentiation process of the progenitor cells is likely determined by the specific localization of the cells and local balances between growth factors. A third process, chondroid metaplasia, has been proposed as an additional mechanism for ankylosing spondylitis. In this process, the chondrocyte matrix is calcified and thus becomes a bone-like tissue. In summary, the temporospatial relationship between inflammation and new bone formation in AS is not clear from these studies. Recent work shows that both classical innate and adaptive immune systems might also participate.

From the molecular perspective, most data concerning ankylosing spondylitis have been derived from mouse models. Based on the view that new bone formation will recapitulate to some extent bone development and growth, a critical role for bone morphogenetic protein (BMP) and Wnt signaling has been proposed (Figure 74-2). In a mouse model of enthesitis, Noggin, a BMP antagonist, inhibits enthesal systems, proinflammatory cytokines inhibit chondrogenesis, and osteogenesis. In other joint diseases such as rheumatoid arthritis, TNF likely contributes to inhibition of repair by upregulating antagonists of the Wnt signaling pathway such as DKK1. Different and contrasting but not mutually exclusive hypotheses have been put forward to explain the apparent paradox of an inflammatory disease associated with new tissue formation. Two of these support the concept that inflammation in AS can subside spontaneously, thereby leaving windows of opportunity for tissue repair ultimately leading to ankylosis. The third hypothesis proposes that chronic inflammation and progressive ankylosis are linked, but that they are in molecular terms largely independent processes. In this hypothesis, cell stress and damage at the enthesis are proposed as disease-initiating events.

It remains unclear whether very early treatment of patients with anti-TNF drugs will have an impact on structural progression. It is possible that future therapeutics based on modulation of BMP and Wnt signaling could have an additional effect on ankylosis, but given the critical role of these pathways in the homeostasis of different organs, simple antagonist or agonist approaches may have toxicity issues. The single observation that celecoxib has an effect on structural disease progression indicates that modulation of bone formation is possible. In this case, it is hypothesized that this anti-inflammatory drug has a direct effect on bone formation, and that this particular effect is not linked to symptom control. A major problem in interpreting the results of therapeutic studies so far is that it is not clear which particular patients are prone to new bone formation. Research conducted to identify this specific at-risk patient group will be critical. Moreover, these efforts may lead to better insights into initiating pathophysiologic mechanisms of the disease.

### Relationship between Inflammation and New Tissue Formation

The specific phenotype of ankylosing spondylitis with progressive spinal ankylosis remains enigmatic. Other diseases such as fibrodysplasia ossificans progressiva, diffuse idiopathic skeletal hyperostosis, and sometimes osteoarthritis, are also characterized by excessive bone formation. These facts and interindividual variability among AS patients suggest that genetic factors may play a role in AS.

Specific localization of new bone formation at enthesal sites suggests that biomechanical factors may play an important role. Bony spurs are found at enthesal sites in degenerative lesions and often are associated with strain. Whether there is a direct relationship between inflammation and new bone formation remains controversial. In most experimental systems, proinflammatory cytokines inhibit chondrogenesis and osteogenesis. In other joint diseases such as rheumatoid arthritis, TNF likely contributes to inhibition of repair by upregulating antagonists of the Wnt signaling pathway such as DKK1. Different and contrasting but not mutually exclusive hypotheses have been put forward to explain the apparent paradox of an inflammatory disease associated with new tissue formation. Two of these support the concept that inflammation in AS can subside spontaneously, thereby leaving windows of opportunity for tissue repair ultimately leading to ankylosis. The third hypothesis proposes that chronic inflammation and progressive ankylosis are linked, but that they are in molecular terms largely independent processes. In this hypothesis, cell stress and damage at the enthesis are proposed as disease-initiating events.

It remains unclear whether very early treatment of patients with anti-TNF drugs will have an impact on structural progression. It is possible that future therapeutics based on modulation of BMP and Wnt signaling could have an additional effect on ankylosis, but given the critical role of these pathways in the homeostasis of different organs, simple antagonist or agonist approaches may have toxicity issues. The single observation that celecoxib has an effect on structural disease progression indicates that modulation of bone formation is possible. In this case, it is hypothesized that this anti-inflammatory drug has a direct effect on bone formation, and that this particular effect is not linked to symptom control. A major problem in interpreting the results of therapeutic studies so far is that it is not clear which particular patients are prone to new bone formation. Research conducted to identify this specific at-risk patient group will be critical. Moreover, these efforts may lead to better insights into initiating pathophysiologic mechanisms of the disease.

### PATHOGENESIS OF REACTIVE ARTHRITIS

Reactive arthritis (ReA)—a sterile synovitis precipitated by an extra-articular infection—has been a challenge for resolving pathogenesis definitively. ReA is the quintessential paradigm of host susceptibility interacting with
an environmental trigger; as such, it sets a conceptual framework that underlies a broad range of rheumatic diseases. Indeed, ReA occupies an interesting middle ground between autoimmune and autoinflammatory inflammatory diseases. Studies in ReA and undifferentiated oligoarthritis indicate that about 50% of such cases can be attributed to a specific pathogen by a combination of culture and serology, the predominant organisms being Salmonella, Yersinia, and Chlamydia. Species-specific analysis of serologic responses to pathogens may further enhance this detection rate. A recent population-based study sought to define the epidemiology and clinical spectrum of ReA following culture-confirmed infection with bacterial enteric pathogens in the United States. Telephone interviews were conducted with persons who had culture-confirmed Campylobacter, Escherichia coli O157, Salmonella, Shigella, or Yersinia infection in Minnesota and Oregon between 2002 and 2004. The estimated incidence of ReA following culture-confirmed Campylobacter, E. coli O157, Salmonella, Shigella, or Yersinia infection in Oregon was 0.6 to 3.1 cases per 100,000.

**Cytokines in Reactive Arthritis**

Analysis of T cell subsets with respect to cytokine profiles is a further method for studying the link between infection and reactive arthritis. Exposure to different pathogens can stimulate at least two patterns of cytokine production by CD4+ T cells (T helper [Th]1 and Th2). Th1 cells mediate a protective role against intracellular pathogens. Thus, it would seem appropriate that these cells would be central in clinical complications of such infections. In a study of 11 patients with ReA, it was observed that stimulation of synovial fluid mononuclear cells resulted in secretion of low amounts of interferon (IFN)-γ and TNF but high amounts of IL-10. IL-10 was responsible for suppression of IFN-γ and TNF as judged by the effect of adding IL-10 or anti–IL-10 to cells. Enhanced production of TNF and IFN-γ is observed in chronic ReA when compared with acute ReA or rheumatoid arthritis. TNF dependence of chronic inflammation in the spine draws indirect support from the dramatic changes that follow the institution of anti-TNF therapy in these patients. In a study of 11 patients with ReA, it was observed that stimulation of synovial fluid mononuclear cells resulted in secretion of low amounts of interferon (IFN)-γ and TNF but high amounts of IL-10. IL-10 was responsible for suppression of IFN-γ and TNF as judged by the effect of adding IL-10 or anti–IL-10 to cells. Enhanced production of TNF and IFN-γ is observed in chronic ReA when compared with acute ReA or rheumatoid arthritis. TNF dependence of chronic inflammation in the spine draws indirect support from the dramatic changes that follow the institution of anti-TNF therapy in these patients. A study investigated TNF production in healthy subjects with previous Yersinia-triggered ReA. The study comprised HLA-B27+ subjects with ReA (B27+ ReA+), in contrast to B27+ReA− and B27 ReA− subjects. It was observed that B27+ReA+ supernatants had higher TNF levels than B27+ReA− supernatants after stimulation with the phorbol ester, phorbol myristate acetate (PMA), and the calcium ionophore, A23187. Patients who have recovered from Yersinia ReA show enhanced TNF production, which may be regulated at the level of...

---

**Figure 74-2** Roles of bone morphogenetic proteins (BMPs) and WNTs in endochondral bone formation. A, Physiologic endochondral bone formation is stimulated by BMPs. WNTs play a supportive role for BMPs. However, some WNTs have a negative effect on early chondrocyte differentiation. B, In the presence of inflammation, tumor necrosis factor (TNF) may stimulate BMP signaling as well as expression of DKK1, which acts as a WNT antagonist. The balance between TNF, BMP, and WNT signaling may determine the onset and progression of ankylosis.
monocyte adhesion. A recent study measured levels of several proinflammatory and immunoregulatory cytokines in synovial fluid and sera from patients with ReA and undifferentiated spondyloarthritis, in comparison with rheumatoid arthritis and osteoarthritis.61 Synovial fluid concentrations of IL-17, IL-6, transforming growth factor (TGF)-β, and IFN-γ were significantly higher in patients with ReA as compared with RA patients. Synovial fluid levels of IL-10 were comparable, but the ratio of IFN-γ/IL-10 was significantly higher in ReA patients than in rheumatoid arthritis patients. IL-17, IL-6, IL-10, and IFN-γ synovial fluid levels were significantly higher than paired serum levels in ReA patients. These findings suggest that Th17 cells, as well as Th1 cells, could be playing a major role in the chronic inflammation seen in ReA.

Innate Immunity and Reactive Arthritis

Recognition that critical host determinants are operational as the first line of defense against pathogens, and that these determinants precede adaptive immunity, represents a major advance in our understanding of host-microbial interactions. ReA represents the paradigm of a dynamic interaction between genetically defined host susceptibility and an environmental trigger. Thus, it is particularly relevant to examine innate immunity of the host because these events, which set the stage for the subsequent development of adaptive immunity, occur early in the course of a challenge by an arthritogenic pathogen. A recent outbreak of salmonellosis was accompanied by ReA and afforded investigators the opportunity to study genetic susceptibility to ReA, which traditionally has been difficult in examining sporadic cases of ReA.57 Using this cohort and exposed non-ReA individuals as controls, a recent study examined the role of innate immunity in ReA.58 Genotyping was performed using two TLR2 (rs5743708 and rs5743704) and two TLR4 (rs4986790 and rs4986791) SNPs. No TLR4 exon variants were associated with any clinical events accompanying the Salmonella infection. In contrast, one of the rare TLR2 SNPs (rs5743708; R753Q) was associated with the development of ReA. The TLR2 variant 753Q was not detected in any of the infected individuals with an uncomplicated course. Another TLR2 variant, 631H, was associated with articular symptoms in infected men. This is the first demonstration that genetic variants of TLR2, but not TLR4, are associated with clinical ReA.

The contribution of TLR interactions in Chlamydia has also been studied recently. Chlamydia trachomatis—inactive elementary bodies (EBs) and the following antigens were tested for their ability to induce the production of proinflammatory cytokines by human monocytes/macrophages and THP-1 cells: purified lipopolysaccharide, recombinant heat shock protein (rHsp)70, rHsp60, rHsp10, recombinant polypeptide encoded by open reading frame 3 of the plasmid (rOmp3), recombinant macrophage infectivity potentiator (rMip), and recombinant outer membrane protein 2 (rOmp2). Aside from EB, rMip displayed the greatest ability to induce release of IL-1β, TNF, IL-6, and IL-8. Stimulating pathways appeared to involve TLR2/TLR1/TLR6 with the help of CD14 but not TLR4. These data support a role for Mip lipoprotein in the pathogenesis of C. trachomatis—induced inflammatory responses and heighten attention on innate immunity during Chlamydia infection.

Response of Reactive Arthritis to Antibiotic Treatment

Recognition that C. trachomatis and C. pneumoniae species may exist in a persistent metabolically active infective state in the synovium has suggested that persistent chlamydiae may be susceptible to antimicrobial agents. A recent study undertook a 9-month, double-blind prospective trial to assess a 6-month course of combination antibiotics as treatment for Chlamydia-induced ReA.66 Treatment arms consisted of doxycycline and rifampin, azithromycin and rifampin, and placebo. After 6 months of treatment, 17 of 27 subjects (63%) on antibiotics were responders compared with 3 of 15 (20%) on placebo. Antibiotic treatment was associated with significant improvement in the modified swollen joint count and tender joint count. Significantly more subjects became polymerase chain reaction negative at month 6 in the active therapy group than in the placebo group. These data suggest that a 6-month course of combination antibiotics might serve as effective therapy for chronic Chlamydia-induced ReA. Further studies will be needed to determine whether this approach is effective in ReA triggered by other pathogens.

References


The references for this chapter can also be found on www.expertconsult.com.
Ankylosing spondylitis (AS) belongs to the group of diseases known as the spondyloarthropathies or, better, spondyloarthritides. This group of disorders constitutes a family of related but heterogeneous conditions rather than a single disease with different clinical manifestations (Tables 75-1 and 75-2).

Radiographic sacroiliitis is considered a hallmark in AS and is found in more than 90% of patients. The inflammation of the sacroiliac (SI) joints and the spine eventually may lead to bony ankylosis. The term ankylosing spondylitis is derived from the Greek roots ankylos, or “bent” (although it now usually implies fusion or adhesions), and spondylos, or “vertebral disk.” Ankylosis of the spine tends to appear in late stages of the disease and does not occur in many patients with mild disease.

Many patients with AS have onset of back pain during their third decade of life. It takes, on average, 6 to 8 years between the onset of back pain and establishing a definite diagnosis of AS. This diagnostic delay in the majority of patients results mainly from the relatively late appearance of definite radiographic sacroiliitis on conventional plain radiographs,23 which is a requirement for diagnosis according to the modified New York criteria. Active sacroiliitis on magnetic resonance imaging (MRI) has been shown to predict the later appearance of sacroiliitis on radiographs.45 Thus many patients at an early stage of AS typically present with characteristic clinical symptoms of AS but may not show definite sacroiliitis on radiographs. Therefore they may not be classified as AS cases. In such patients with suspected early AS, sacroiliitis may best be detected on MRI.1 It has been proposed that a substantial proportion of such patients will develop radiographic sacroiliitis (i.e., structural damage of SI joints) with time and will progress to definite AS, but this hypothesis requires further evaluation in prospective studies. On the other hand, this also implies that a proportion of patients will remain at this (nonradiographic) stage of disease, with inflammation on MRI at some point, but without radiographically detectable damage over the subsequent years.5 Therefore in order to describe these patients correctly, the term early AS has been dropped, with terms such as preradiographic axial spondyloarthritis1 or, more recently, nonradiographic axial spondyloarthritis4 seeming to be more appropriate (Figure 75-1). The term axial spondyloarthritis comprises both nonradiographic AS and classic AS (according to the modified New York criteria) and is now considered by ASAS (Assessment of SpondyloArthritis international Society) as the preferred terminology for patients with predominantly spinal disease.

The total prevalence of axial spondyloarthritis has been estimated at two to three times that of AS according to the modified New York criteria.1 In the concept of axial spondyloarthritis, the disease AS represents the tip of the iceberg. Patients with axial spondyloarthritis who have normal SI joints on conventional radiographs often show inflammatory changes on appropriate MRI and may or may not progress to sacroiliitis according to the New York criteria. Such patients may have active disease that needs appropriate treatment. They respond to therapy with biologics. The course and natural history of axial spondyloarthritis without radiographic sacroiliitis is not yet well known. However, although this chapter focuses on radiographic AS, many aspects also apply to patients with axial spondyloarthritis. The interval between the first complaints of the disease and the time of a definite diagnosis may be as long
as 4 to 9 years. This affects how disease duration is defined. Important components of the definition of disease duration are provided in Table 75-3.

CLASSIFICATION

Criteria for Ankylosing Spondylitis and Axial Spondyloarthritis

The diagnosis of AS is based on clinical features. The disease is “primary” or “idiopathic” if no associated disorder is present; it is “secondary” if the disease is associated with psoriasis or chronic inflammatory bowel disease. In daily practice, a presumptive clinical diagnosis of AS is usually supported by evidence of sacroilitis on conventional pelvic radiographs; indeed, many think of AS as a symptomatic sacroilitis. The presence of sacroilitis does not necessarily indicate the presence of AS, however. Moreover, although radiographic sacroilitis is frequent in AS, it is by no means an early or obligate manifestation of the disease. Contrary to the New York criteria, radiographic evidence of sacroilitis is not obligatory to fulfill the Rome criteria (Table 75-4). Both sets were primarily intended for use in epidemiologic studies. Lack of either sensitivity or specificity led to a modification of the New York criteria for AS (see Table 75-4). Two criteria—limitation of lumbar spine motion and limitation of chest expansion—appear to reflect disease duration; they are usually not present in early disease. It should be stressed that classification criteria are usually not useful for early diagnosis owing to a lack of sensitivity. In particular, in the early phase of AS, conventional SI radiographs may be normal. MRI is useful to diagnose the disease with predominantly axial manifestations before the presence of radiographic sacroilitis. To encompass both pre-radiographic AS and classic AS (according to the modified New York criteria), the Assessment of SpondyloArthritis international Society (ASAS) has proposed classification criteria for axial spondyloarthritis for patients with back pain of at least 3 months’ duration and age at onset of complaints before 45 years (Table 75-5). The term axial spondyloarthritis has been introduced as an umbrella for the entire spectrum of spondyloarthritis patients with predominant axial involvement, irrespective of the presence of structural damage on radiographs. Thus axial spondyloarthritis includes nonradiographic axial spondyloarthritis and classical AS (fulfilling the modified New York criteria). According to the ASAS classification criteria for axial spondyloarthritis, a patient with chronic back pain and age at onset before age 45 can be classified as having axial spondyloarthritis if sacroilitis on imaging (radiographs or MRI) is present plus at least one further spondyloarthritic feature, or, in the absence of sacroilitis on imaging, if HLA-B27 plus at least two further spondyloarthritic features are present (see Table 75-5). The sensitivity of the entire set of ASAS criteria for axial spondyloarthritis was 83%, and the specificity was 84%. In the ASAS study on these new criteria, 30% of all patients diagnosed as having axial spondyloarthritis had definite radiographic sacroilitis and fulfilled the modified New York criteria for AS. Therefore two-thirds were classified as nonradiographic axial spondyloarthritis.

EPIDEmiOLOGY

Prevalence

The prevalence of AS closely parallels the frequency of HLA-B27. This holds true for those B27 subtypes that are associated with the disease, but it is not true for populations in which the HLA-B27*06 subtype that lacks a strong association with AS occurs frequently such as the Indonesian population.

Table 75-2 Clinical Characteristics of Spondyloarthritis

| Typical pattern of peripheral arthritis | Predominantly of lower limb, asymmetric |
| Absence of rheumatoid factor | Absence of subcutaneous nodules and other extra-articular features of rheumatoid arthritis |
| Overlapping extra-articular features characteristic of the group (e.g., anterior uveitis) | Significant familial aggregation |
| Association with HLA-B27 |

Table 75-3 Components of Disease Duration

| Onset of axial AS manifestations (inflammatory back pain) |
| Onset of extra-axial AS manifestations (peripheral arthritis, enthesitis) |
| Onset of associated spondyloarthropathic diseases (acute anterior uveitis, inflammatory bowel disease, psoriasis) |
| Time since diagnosis of AS by health care provider |

Table 75-4  Criteria for Ankylosing Spondylitis

Rome, 1961

Clinical Criteria
1. Low back pain and stiffness for more than 3 mo, not relieved by rest
2. Pain and stiffness in thoracic region
3. Limited motion in lumbar spine
4. Limited chest expansion
5. History or evidence of iritis or its sequelae

Radiographic Criterion
6. Radiograph showing bilateral sacroiliac changes characteristic of ankylosing spondylitis (this excludes bilateral osteoarthritis of sacroiliac joints)

Definite Ankylosing Spondylitis
Grade 2 or more bilateral sacroiliitis with at least one clinical criterion
Or
At least four clinical criteria

New York, 1966

Diagnostic Criteria
1. Limitation of lumbar spine motion in all three planes: anterior flexion, lateral flexion, extension
2. Pain at dorsolumbar junction or in lumbar spine
3. Limitation of chest expansion to 2.5 cm or less measured at level of fourth intercostal space

Grading of Radiographs
Normal, 0; suspicious, 1; minimal sacroiliitis, 2; moderate sacroiliitis, 3; ankylosis, 4

Definite Ankylosing Spondylitis
Grade 3 or 4 bilateral sacroiliitis with at least one diagnostic criterion
Or
Grade 3 or 4 unilateral or grade 2 bilateral sacroiliitis with diagnostic criterion 1 or with criteria 2 and 3

Probable Ankylosing Spondylitis
Grade 3 or 4 bilateral sacroiliitis with no diagnostic criteria

Modified New York, 1984

Criteria
1. Low back pain of at least 3 months’ duration improved by exercise and not relieved by rest
2. Limitation of lumbar spine in sagittal and frontal planes
3. Chest expansion decreased relative to normal values for age and sex
4. Bilateral sacroiliitis grade 2 to 4
5. Unilateral sacroiliitis grade 3 or 4

Definite Ankylosing Spondylitis
Unilateral grade 3 or 4, or bilateral grade 2 to 4 sacroiliitis and any clinical criterion


Among whites, the estimated prevalence rate of AS as defined by the modified New York criteria ranges from 68 per 100,000 population older than 20 years in the Netherlands to 197 per 100,000 in the United States.3,12 The prevalence of clinical AS in France is 150 per 100,000 adults, whereas in Norway it is 210 per 100,000 adults.18,19 The prevalence of the disease in Finland is similar, with a figure of 150 per 100,000 people.20

Higher prevalence rates have been reported in central Europe. An epidemiologic study from Berlin reported a prevalence figure of 0.86%.21 In the general population, AS is likely to develop in about 1% to 2% of HLA-B27+ adults who have a disease-associated B27 subtype, although there may be regional or geographic differences. For example, in northern Norway, AS may develop in 6.7% of HLA-B27+ people.22,23

The disease is much more common among HLA-B27+ first-degree relatives of HLA-B27+ AS patients; roughly 10% to 30% of them have signs or symptoms of AS.9 In fact, a positive family history of AS is a strong risk factor for the disease.

Incidence

There is no adequate evidence that the incidence of AS has changed in the past few decades. Clinical features, age of onset, and survival time have remained stable.23 One study revealed an overall age and gender-adjusted incidence of 7.3 per 100,000 person-years. This U.S. figure compares quite well with the Finnish study, which revealed a stable incidence of 8.7 (95% confidence interval [CI], 6.4 to 11.0) per 100,000 people aged 16 or older.16

Racial Distribution

AS occurs in all parts of the world, but there are race-related differences in prevalence. This might reflect differences in the distribution of HLA-B27 among races. Approximately 90% of white patients with AS possess HLA-B27, whereas AS and HLA-B27 are nearly absent (prevalence of B27 < 1%) in African blacks and Japanese. In African-Americans, owing to racial admixture with whites, 2% possess B27, but only about 50% of black patients with AS possess B27. Correspondingly, African-Americans are affected far less frequently than American whites.

Table 75-5  ASAS Classification Criteria for Axial Spondyloarthritis (SpA) (in Patients with Back Pain ≥ 3 Months and Age at Onset < 45 Years)

<table>
<thead>
<tr>
<th>SpA Features</th>
<th>Sacroiliitis on Imaging plus ≥ 1 SpA feature OR HLA-B27 plus ≥ 2 other SpA features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflammatory back pain</td>
<td>Active (acute) inflammation on MRI highly suggestive of sacroiliitis associated with SpA</td>
</tr>
<tr>
<td>Arthritis</td>
<td>Sacroiliitis according to modified New York criteria</td>
</tr>
<tr>
<td>Enthesitis (heel)</td>
<td>Dactylitis</td>
</tr>
<tr>
<td>Uveitis</td>
<td>Crohn’s disease/ulcerative colitis</td>
</tr>
<tr>
<td>Family history for SpA</td>
<td>Family history for SpA</td>
</tr>
<tr>
<td>HLA-B27</td>
<td>Elevated CRP*</td>
</tr>
</tbody>
</table>

*Elevated CRP is considered a SpA feature in the context of chronic back pain.

ASAS, Assessment of SpondyloArthritis international Society; CRP, C-reactive protein; HLA-B27, human leukocyte antigen-B27; IBP, inflammatory back pain; MRI, magnetic resonance imaging; NSAIDs, nonsteroidal anti-inflammatory drugs.

Burden of Disease

AS is associated with a considerable burden to the patient and society. Apart from the axial and articular manifestations, extra-articular manifestations such as uveitis and acute anterior uveitis and comorbidities such as inflammatory bowel disease and psoriasis contribute to the burden of disease. In addition, a large proportion of patients has spinal osteoporosis, leading to vertebral fractures and thoracic kyphosis. All these features result in a decreased quality of life. Disease status scores for physical functioning and disease activity correlate clearly with psychologic scores for anxiety and depression. The impact of AS can also be seen in various aspects of employment, ranging from requiring assistance at work to increased sick leave and withdrawal from the workforce. Apart from the impact on labor force participation, AS patients have an important impact on health care and non-health care resource utilization, resulting in mean total costs (direct and productivity) of about $6700 to $9500 per year per patient when applying the human capital approach to calculate productivity costs.

The burden of illness increases with duration of disease. Because the burden due to AS reduces quality of life, and because all types of costs associated with AS result from loss of function and disease activity, early diagnosis and treatment are necessary to prevent or reduce functional decline and improve patient outcome.

PATHOGENESIS

The precise cause of AS is still unclear, but several novel insights are currently emerging. First, several new hypotheses have been proposed to explain the major genetic contribution of HLA-B27. Second, other genetic risk factors have been identified by genome-wide association studies. Third, there is increasing evidence that AS may be driven by abnormal innate immune reactions rather than by autoantigen-specific T and/or B cell reactivity. And finally, the molecular mechanisms of new bone formation and their relationship with inflammation have started to emerge.

HLA-B27

The dominant role of genetic factors is highlighted by data demonstrating disease concordance in 75% of monozygotic twins compared with 13% of nonidentical twins, familial aggregation, and population data demonstrating associations with HLA-B27. About 90% of white AS patients are HLA-B27+. HLA-B27 is the first described and major genetic risk factor for AS but, despite the near-pervasive nature of this association, it has been estimated that B27 contributes only 16% of the total genetic risk. In humans, there appears to be a hierarchy of association of the more than 45 as yet known subtypes of HLA-B27 with AS, ranging from strong association with HLA-B*2705 to weak association with HLA-B*2709. The direct pathogenic role of HLA-B27 is evidenced by the spontaneous spondyloarthritis-like disease in rats overexpressing HLA-B27.

The main function of human leukocyte antigen (HLA) class I molecules such as HLA-B27 is to present peptides to CD8+ T cells and, accordingly, the arthritogenic peptide hypothesis proposes that CD8+ T cells are activated by bacterial antigens, for example in the gut, and after recirculation are reactivated in the joint by cartilage or other autoantigens. A number of studies have provided some partial evidence for this concept. This hypothesis has been thoroughly questioned by the fact that HLA-B27 transgenic rats develop severe spondyloarthritis-like disease even in the absence of CD8+ T cells. Two alternative hypotheses have been proposed more recently, suggesting that HLA-B27 has other specific features independent of antigen presentation that may directly contribute to disease. One theory indicates that HLA-B27 has a special tendency to form heavy chain homodimers that, when present on the cell surface, can trigger direct activation of natural killer (NK) cells through recognition via killer immunoglobulin receptor (KIR)-like receptors.

Alternatively, HLA-B27 has a special propensity to misfold in the endoplasmic reticulum and thereby to promote an unfolded protein response, which in turn modulates the functional behavior and cytokine production of myeloid cells. The latter two hypotheses imply a role for altered innate immune responses rather than autoantigen-driven acquired immunity in the pathogenesis of AS.

Population and family studies have shown that HLA-B60 increases susceptibility to AS. This applies to both HLA-B27+ and HLA-B27- persons.

Non-Human Leukocyte Antigen Genes

Two observations clearly suggest a role for non-HLA genetic risk factors. First, there is an increased risk for disease in B27+ first-degree relatives of AS probands (10% to 20%) compared with B27+ individuals in the general population (2% to 5%). Second, disease concordance is 75% in identical twins versus 27% in HLA-B27 concordant dizygotic twins. Genome-wide association studies have now allowed the identification of several of these factors. There is a robust association with single-nucleotide polymorphisms (SNPs) in endoplasmic reticulum aminopeptidase-1 (ERAP-1), an enzyme involved in the trimming of peptides for loading in MHC molecules. There is now emerging evidence for a gene-gene interaction between ERAP-1 and HLA-B27, but it remains unknown how the polymorphisms may affect antigen presentation, HLA-B27 homodimer formation, and misfolding. A second strong association is found with SNPs in IL-23R, a genetic feature shared with Crohn’s disease and psoriasis. Together with the suggested but not yet confirmed association with STAT3 and JAK2 polymorphisms, this suggests the potential involvement of an interleukin (IL)-17 response in AS. Definite association was also found with gene deserts on chromosome 2p15 and 21q22, which warrants further investigation of noncoding ribonucleic acid (RNA) and epigenetic effects in the pathophysiology of this disease. Besides these definite associations, genome-wide association studies have suggested potential associations with tumor necrosis factor (TNF) receptor 1 (TNFR1), the signaling molecule TNFR1-associated death domain protein (TRADD), the TNF superfamily cytokine TNFSF15, IL-1A and the IL-1 receptor 2 (IL-1R2), IL-12B, the vascular morphogenesis protein gene anthrax toxin receptor 2 (ANTXR2), and the innate
immune receptor caspase recruitment domain family, member 9 (CARD9).56

These associations point toward a potential pathogenic role for altered innate immune responses, TNFR1 signaling, and IL-1.

The importance of TNF to the pathogenesis of AS is highlighted in the phenotype of a transgenic mouse model that overexpresses TNF. These animals develop sacroiliitis characterized by the formation of osteoclasts and granulation tissue.

**Autoimmunity versus Autoinflammation**

Both the alternative roles for HLA-B27 besides antigen presentation and the genetic risk factors suggest that AS may not be primarily driven by a canonical autoantigen-specific T and/or B cell reactivity. Indeed, AS does not share genetic risk factors such as PTPN22 or cytotoxic T lymphocyte–associated antigen-4 (CTLA-4) with other autoimmune diseases and lacks disease-specific autoantibodies. On the basis of the predilection of the disease for sites rich in cartilage, it has been proposed that cartilage antigens may be the primary target of an autoimmune response in AS. This hypothesis is supported by several animal models of peripheral and axial arthritis on the basis of the induction of autoimmunity to antigens present in cartilage and fibrous tissue such as aggrecan and versican.51,52 T cell responses toward the G1 domain of aggrecan have also been observed in human AS, but similar responses were seen in other inflammatory joint disorders, implying a nonspecific response to joint damage.53 Finally, biologic therapies targeting B (such as anti-CD20) or T (such as CTLA4-Ig) cell pathways have limited to no efficacy in AS in proof-of-concept trials.54,55 Taken together, little strong evidence supports a primary T cell– or B cell–driven autoimmune process in AS.

An alternative hypothesis proposes that AS would be driven by abnormal reactivity of innate immune cells such as macrophages, neutrophils, and mast cells. This hypothesis is supported by several lines of evidence. First, genetic associations with CARD9 and the IL-1 pathway strongly suggest involvement of the inflammasome. Second, histologic studies demonstrated increased infiltration with innate immune cells, but not T, B, or dendritic cells in the peripheral joint and gut of patients with AS.56-58 Third, human AS and its experimental models have a predilection for sites exposed to either microbial or mechanical stress. The latter aspect has led to the concept that AS and other forms of spondyloarthritides are primarily characterized by inflammation of the so-called synovio-entheseal complex.59 In this perspective, it is interesting to note that the fibrocartilage protein versican, which could be released during mechanical stress and induces spondylitis in BALB/c mice, is also a ligand for the innate immune receptor TLR2. Formal proof that innate immune alterations may trigger AS remains, however, awaited.

Independently of the exact origin of the inflammation, it is clear that several proinflammatory cytokines are pivotal in the downstream effector mechanisms. TNF plays a crucial role in the disease process as evidenced by the successful introduction of TNF blockade for this disease. Accordingly, transgenic overexpression of TNF in mice leads not only to severe, destructive polyarthritis but also to sacroiliitis.60 However, it remains incompletely understood which cells are the main producers of TNF in AS, in which form (soluble vs. transmembrane) and through which receptor (TNFR1 or TNFR2) TNF exerts its pathogenic effects, and which are the main target cells. In line with the genetic data, experimental data suggest a role for TNFR1 signaling in stromal cells.61 Besides TNF, the genetic data and emerging data form functional studies on the unfolded protein response implicate IL-23 and thus possibly IL-17 as important cytokines in the pathogenesis of AS.62

Clinical trials targeting the IL-23/IL-17 pathway in AS are currently being performed.

**Structural Remodeling and Ankylosis**

A crucial and largely unexplained aspect of the pathogenesis of AS is the occurrence of structural remodeling and new bone formation, ultimately leading to ankylosis. Human and experimental studies have revealed three important concepts. First, the remodeling phenotype of AS cannot be explained by the absence of joint destruction because erosive disease is evidenced by histology and radiology and because the cellular and molecular machinery for cartilage and bone destruction is present and operative at the sites of disease.63,64

Second, structural remodeling and osteoproliferation are dependent on endochondral bone formation. This process is governed by several molecular pathways including bone morphogenetic protein (BMP) and Wnt signaling.65,66

There is emerging evidence that several inhibitors of the Wnt pathway are dysfunctional in human spondyloarthritides and that this correlates with new bone formation.67 Third, osteoproliferation is not critically dependent on inflammation because, despite clinical efficacy on signs and symptoms, TNF blockade fails to halt radiologic progression in AS and both processes can be uncoupled in experimental models.58,69

On the other hand, vertebral sites showing signs of inflammation at baseline have a higher risk for subsequent osteoproliferation on follow-up.70 The exact relationship between inflammation and osteoproliferation needs to be further clarified because it may have important consequences for treatment.

**PATHOLOGY**

Characteristic pathologic features of AS include inflammation in axial joints, large peripheral joints, and entheses associated with inflammation in subchondral bone marrow. Reparation is also characteristic in terms of the development of chondroid metaplasia, followed by calcification of cartilage and formation of bone, particularly in the axial joints. Fat metaplasia in the axial skeleton at sites of prior inflammation is frequently observed on MRI.

**Axial Skeleton**

Detailed histopathologic studies of axial involvement in AS are limited owing to the inaccessibility of biopsy material of both SI joints and the spine. A controlled study of SI biopsies from AS patients at various stages of disease and
controls showed cellular infiltration with lymphocytes, macrophages, and plasma cells in the synovium and subchondral marrow as the earliest features of disease. Later features include the development of pannus extending from both synovium and subchondral bone marrow, with erosion of articular cartilage and its replacement by granulation tissue. Osteoclast formation and erosion of subchondral bone account for the typical widening of the joint spondyloarthritic seen on plain radiography. Enthesitis is also evident in later stages of disease at the insertion of the posterior capsule. Reparative changes include cartilage metaplasia at sites of active inflammation, followed by its calcification and then replacement by endochondral bone, leading to obliteration of the joint space by ankylosis. Pararticular changes include bone sclerosis and fat replacement of bone marrow. Regarding the spine, pathologic data are limited except for a number of recent studies of apophyseal joints. Immunohistologic analysis shows subchondral lymphocyte infiltrates with CD4+ and CD8+ T cells, together with hypervascularization and foci of CD68+ osteoclastic cells.

Peripheral Skeleton

Regarding peripheral joint manifestations of AS, histopathologic studies have assessed surgical samples of affected hip joint, synovial biopsies of knee and ankle joint, and enthesis. Involvement of the hips is characterized by subchondral granulation tissue and osteoclast formation in the femoral heads and acetabulum that is associated with degradation of overlying articular cartilage. Synovial studies have revealed that the type of inflammation is strongly different from that observed in RA, with increased vascularity, increased infiltration with innate immune cells, and absence of specific features of T and B cell autoimmunity.

Interestingly, these studies revealed a similar pattern of inflammation in AS in comparison with other subtypes of spondyloarthritis. Enthesial involvement most frequently occurs at sites rich in fibrocartilage such as the Achilles tendon. Inflammation and chronic cellular infiltration of soft tissues are relatively sparse but may be extensive within the adjacent subchondral bone, particularly in B27+ individuals. A comparative study of the subchondral marrow from knee and hip joint entheses showed that AS patients clearly differ from those with RA and osteoarthritis with respect to the frequency of marrow inflammation, infiltration with CD8+ T cells, and presence of hyperosteoclastic erosive lesions.

<table>
<thead>
<tr>
<th>Table 75-6</th>
<th>Aspects of Inflammatory Back Pain in Ankylosing Spondylitis and Axial Spondyloarthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Onset of complaints before age 45</strong></td>
<td><strong>Duration of symptoms more than 3 mo (chronic pain)</strong></td>
</tr>
<tr>
<td><strong>Located at the lower back</strong></td>
<td><strong>Alternating buttock pain</strong></td>
</tr>
<tr>
<td><strong>Awaking due to back pain during the second half of the night</strong></td>
<td><strong>Morning stiffness for at least 30 min</strong></td>
</tr>
<tr>
<td><strong>Insidious onset of complaints</strong></td>
<td><strong>Improvement with exercises</strong></td>
</tr>
<tr>
<td><strong>No improvement of back pain with rest</strong></td>
<td><strong>Improvement with use of nonsteroidal agents</strong></td>
</tr>
</tbody>
</table>

The pain is initially felt primarily deep in the gluteal region, is dull in character, is difficult to localize, and is insidious in onset. The pain can be severe at this early phase of the disease; it localizes in the SI joints but is occasionally referred toward the iliac crest or greater trochanteric region or down the dorsal thigh. Radiation of buttck pain may suggest root compression of the sciatic nerve. The buttck pain typically alternates from side to side. Coughing, sneezing, or other maneuvers that cause a sudden twist of the back may accentuate pain. Although the pain is often unilateral or intermittent at first, within a few months it usually becomes persistent and bilateral and the lower lumbar area becomes stiff and painful. The pain is associated with a feeling of low back stiffness that is worse in the morning and may awaken the patient from sleep, particularly during the second half of the night. Many patients do not differentiate between low back pain and stiffness. The morning stiffness may last up to 3 hours. Both the stiffness and the pain tend to be eased by a hot shower, an exercise program, or physical activity; they do not improve with rest. Fatigue as a result of chronic back pain and stiffness may be an important problem and can be accentuated by sleep disturbances due to these symptoms.

Chest Pain

With subsequent involvement of the thoracic spine (including costovertebral and costotransverse joints) and the occurrence of enthesitis at the costosternal and manubriostral joints, patients may experience chest pain accentuated by coughing or sneezing, which is sometimes characterized as “pleuritic.” The chest pain is often associated with tenderness over the sternocostal or costosternal junctions. Mild to moderate reduction of chest expansion is often detectable in an early stage of AS. Chest pain occurs relatively often in HLA-B27+ relatives, even in the absence of radiographic evidence of sacroilitis.

<table>
<thead>
<tr>
<th>Table 75-7</th>
<th>Diagnostic Features of Ankylosing Spondylitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic inflammatory spinal pain</td>
<td>Chest pain</td>
</tr>
<tr>
<td>Alternate buttock pain</td>
<td>Acute anterior uveitis</td>
</tr>
<tr>
<td>Synovitis (predominantly of lower limbs, asymmetric)</td>
<td>Enthesitis (heel, planter)</td>
</tr>
<tr>
<td>Radiographic sacroilitis</td>
<td>Positive family history of ankylosing spondylitis</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>Chronic inflammatory bowel disease</td>
</tr>
</tbody>
</table>

CLINICAL MANIFESTATIONS

Skeletal Manifestations

Low Back Pain and Stiffness

Back pain is an extremely common symptom, occurring in up to 80% of the general population. Therefore it is important to note that back pain in AS and axial spondyloarthritits has special features that differentiate it from mechanical back pain. In clinical practice inflammatory back pain is often not well recognized. Back pain is the most prevailing diagnostic feature of AS (Table 75-7).


**Tenderness**

Extra-articular tenderness at certain loci is a prominent complaint in some patients. These lesions are due to enthesitis. Common tender sites are the costosternal junctions, spinous processes, iliac crests, greater trochanters, ischial tuberosities, tibial tubercles, and heels (Achilles tendinitis or plantar fasciitis). Radiographically, bone spurs may develop at these sites.

**Joints**

The girdle or “root” joints (hips and shoulders) are the most frequently involved extra-axial joints in AS, and pain in these areas is the presenting symptom in up to 15% of patients. Shoulder involvement, but especially hip involvement, may cause considerable physical disability. Coexisting disease in the lumbar spine often contributes significantly to disability of the lower extremities. Hips and shoulders are involved at some stage of disease in up to 35% of patients. Hip disease is more common in Algeria, India, and Mexico. It is relatively more common as a presenting manifestation if the disease starts in childhood (juvenile AS). In boys 8 to 10 years of age, hip disease as a manifestation of juvenile AS is the most frequent type of chronic arthritis. These children with hip disease are mostly HLA-B27+, and they are serologically negative for antinuclear antibodies.

The knee joint may also be affected in AS, often as an intermittent effusion. The temporomandibular joint is involved in about 10% of patients.

**Extraskeletal Manifestations**

Constitutional symptoms such as fatigue, weight loss, and low-grade fever occur frequently. Other extraskeletal manifestations are more localized.

**Eye Disease**

Acute anterior uveitis or iridocyclitis is the most common extra-articular manifestation of AS, occurring in 25% to 30% of patients at some time during the course of the disease. There is no clear relationship between activity of the articular disease and this extra-articular manifestation. The onset of eye inflammation is usually acute and typically unilateral, but the attacks may alternate. The eye is red and painful, with visual impairment. Photophobia and increased lacrimation may be present. If the eye remains untreated or if treatment is delayed, posterior synchiae and glaucoma may develop. Most attacks subside in 4 to 8 weeks without sequelae if early treatment is provided. Acute anterior uveitis is more common in B27+ than B27- patients with AS. Relatives who have acute anterior uveitis seem to be at higher risk for AS. The calculated incidence of acute anterior uveitis in a Swiss family study was 89 attacks per 1000 patient-years for AS patients, but only 8 per 1000 person-years among healthy B27+ relatives.

**Cardiovascular Disease**

Cardiac involvement may be clinically silent or may cause considerable problems. Manifestations of cardiac involvement include ascending aortitis, aortic valve incompetence, conduction abnormalities, cardiomegaly, and pericarditis. In rare situations, aortitis may precede other features of AS. Aortic incompetence was noted in 3.5% of patients who had the disease for 15 years and in 10% after 30 years. Inflammation and dilation of the aorta are the main causes of aortic valve incompetence. Cardiac conduction disturbances are seen with increasing frequency with the passage of time, occurring in 2.7% of those with disease of 15 years’ duration and in 8.5% after 30 years. Both aortic incompetence and cardiac conduction defects occur twice as often in patients with peripheral joint involvement. In AS the prevalence of myocardial infarction is increased (4.4% in AS patients compared with 1.2% in the general population in a Dutch study).

**Neurologic Involvement**

Neurologic complications of AS can be caused by fracture, instability, compression, or inflammation. Traffic accidents or minor trauma can cause spinal fractures. The C5-C6 or C6-C7 level is the most commonly involved site.

As in RA, atlantoaxial joint subluxation, atlanto-occipital subluxation, and upward subluxation of the axis may occur in AS as a consequence of instability resulting from the inflammatory process. Spontaneous anterior atlantoaxial subluxation is a well-recognized complication in about 2% of patients and manifests with or without signs of spinal cord compression. It is observed more commonly in patients with spondylitis and peripheral arthritis than in those with exclusively axial involvement.

Causes of neurologic complications due to compression include ossification of the posterior longitudinal ligament (which may lead to compressive myelopathy), destructive intervertebral disk lesions, and spinal stenosis.

The cauda equina syndrome is a rare but serious complication of long-standing AS. The syndrome affects lumbosacral nerve roots. This gives rise to pain and sensory loss, but...
frequently there are also urinary and bowel symptoms. Gradual onset of urinary and fecal incontinence, impotence, saddle anesthesia, and occasionally loss of ankle jerks occurs. Motor symptoms, if present, are usually mild. CT and MRI allow the accurate noninvasive diagnosis of this complication of AS.\(^9\) No compressive lesions exist. Arachnoiditis and arachnoid adhesions may be important in the pathogenesis.

**Renal Involvement**

IgA nephropathy has been reported in many patients with AS. These patients often have an elevated immunoglobulin (IgA) level (93%) and renal impairment (27%) at presentation.\(^9\) Microscopic hematuria and proteinuria may occur in up to 35% of patients. The significance of these findings in terms of subsequent deterioration of renal function is unclear.\(^9\) Amyloidosis (secondary type) is a rare complication. Amyloid deposits detected through abdominal subcutaneous fat aspiration are not invariably associated with a poor renal prognosis.\(^9\)

**Osteoporosis**

Osteopenia is seen in the early stages of AS.\(^9\) In patients with this disease, osteoporotic deformities of the thoracic spine contribute significantly to abnormal posture, particularly fixed hyperkyphosis.\(^9\) Radiographic damage to the cervical and lumbar spine, thoracic wedging, and disease activity are determinants of hyperkyphosis in AS.\(^9\) An increased occiput-to-wall distance is associated with vertebral fractures. The prevalence of symptomatic osteoporotic spinal fractures is increased in AS.\(^10\) Neurologic complications occur rather frequently, even after minor trauma.\(^10\) Proper assessment of bone density in the spine is difficult in the presence of syndesmophytes because they may give rise to falsely high values. The true fracture risk and complication rate in early and late disease and the relation to disease activity are not yet known. Currently, it is unclear whether any specific antiosteoporotic therapy to prevent spinal fractures is effective.

**PHYSICAL FINDINGS**

**Spinal Mobility**

To arrive at an early diagnosis, the physician must perform a thorough physical examination. On examination of the spine, there may be some limitation of motion of the lumbar spine as elicited by forward flexion, hyperextension, or lateral flexion. Early loss of the normal lumbar lordosis is often the first sign and is easily assessed on inspection. The Schober test (or its modifications) is useful to detect any limitation of forward flexion of the lumbar spine, although it is typically normal in early disease. As the patient stands erect, one mark is placed with a pen on the skin overlying the fifth lumbar spinous process (usually at the level of the posterosuperior iliac spine or the “dimple of Venus,” and another mark is placed 10 cm above in the midline. The patient is then asked to bend forward maximally without bending the knees. In healthy people, the distance between the two marks on the skin should increase as the skin stretches. If the distance between both marks does not reach 15 cm, this indicates reduced lumbar spine mobility. Lateral flexion may also be diminished, and spinal rotation may cause pain.

**Chest Expansion**

Mild to moderate reduction of chest expansion is often detectable in early stages of AS. Normal values are age and sex dependent, and there is considerable overlap between normal values and those obtained from AS patients. Reduction below 5 cm in young persons with an insidious onset of chronic, inflammatory low back pain strongly suggests AS. Chest expansion should be measured on maximal inspiration after forced maximal expiration at the level of the fourth intercostal spondyloarthritic in males and just below the breasts at the xiphisternal level in females.

**Enthesitis**

Examination of the ischial tuberosities, greater trochanters, spinous processes, costochondral and manubriosternal junctions, supraspinatus insertion, and iliac crests can determine the presence of enthesitis. Heel pain, especially when getting out of bed, is a characteristic manifestation of Achilles and plantar fasciitis enthesitis.

**Sacroiliitis**

Direct pressure over the SI joints may elicit pain, as may special testing maneuvers, although the latter lack specificity and sensitivity. These signs may also be negative in early disease or may become negative in late stages as inflammation is replaced by fibrosis or bony ankylosis.

**Posture**

Over the course of the disease, the patient may lose normal posture. Involvement of the cervical spine is manifested by pain and limitation of neck movement. A forward slope of the neck can be detected by having the patient stand against a wall and try to position his or her occiput against it. After many years of progression in patients with severe disease, the entire spine may become increasingly stiff, with loss of normal posture from gradual loss of lumbar lordosis and the development of thoracic kyphosis.\(^9\)\(^\text{99}\) The abdomen becomes protuberant; breathing is primarily by diaphragmatic action. These typical deformities usually evolve after disease duration of 10 years or more.

**LABORATORY TESTS**

Generally, routine blood tests are not helpful. A normal erythrocyte sedimentation rate (ESR) or normal C-reactive protein (CRP) level does not exclude active disease. An elevated ESR or CRP is reported in up to 75% of patients, but it may not correlate with clinical disease activity.\(^10\) In an unselected patient population, an elevated ESR and CRP was present in 45% and 38%, respectively, of patients with spinal disease only, compared with 62% and 61%, respectively, of patients with peripheral arthritis with or...
without inflammatory bowel disease. Neither ESR nor CRP is superior in assessing disease activity. A mild normochromic anemia may be present in 15% of patients. Elevation of serum alkaline phosphatase (derived primarily from bone) is seen in some patients but is unrelated to disease activity or duration. Some elevation of serum IgA is frequent in AS. Its level correlates with acute-phase reactants. Active disease is associated with decreased lipid levels, particularly high-density lipoprotein cholesterol, resulting in a more atherogenic lipid profile.

**IMAGING STUDIES**

**Conventional Radiography**

The typical radiographic changes of AS are seen primarily in the axial skeleton, especially in the SI, discovertebral, apophyseal, costovertebral, and costotransverse joints. They evolve over many years, with the earliest, most consistent, and most characteristic findings seen in the SI joints. However, otherwise typical AS has been described in the absence of radiographic evidence of sacroiliitis. The radiographic findings of sacroiliitis are usually symmetric and consist of blurring of the subchondral bone plate, followed by erosions and sclerosis of the adjacent bone. The changes in the synovial portion of the joint (i.e., the lower two-thirds of the joint) result from inflammatory synovitis and osteitis of the adjacent subchondral bone. The cartilage covering the iliac side of the joint is much thinner than that covering the sacral side. Therefore the erosions and subchondral sclerosis are typically seen first and tend to be more prominent on the iliac side.

In the upper one-third of the SI joint, where strong intraarticular ligaments hold the bones together, the inflammatory process may lead to similar radiographic abnormalities. Progression of the subchondral bone erosions can lead to pseudowidening of the SI joint space. Over time, gradual fibrosis, calcification, interosseous bridging, and ossification occur. Erosions become less obvious, but the subchondral sclerosis persists, becoming the most prominent radiographic feature.

Ultimately, usually after several years, there may be complete bony ankylosis of the SI joints, with resolution of bony sclerosis. It is practical to grade radiographic sacroiliitis according to the New York criteria (Table 75-8).

Table 75-8  Grading of Sacroiliitis: New York Criteria

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>normal</td>
</tr>
<tr>
<td>1</td>
<td>suspicious</td>
</tr>
<tr>
<td>2</td>
<td>minimal sacroiliitis</td>
</tr>
<tr>
<td>3</td>
<td>moderate sacroiliitis</td>
</tr>
<tr>
<td>4</td>
<td>ankylosis</td>
</tr>
</tbody>
</table>

Bony erosions and osteitis (“whiskering”) at sites of osseous attachment of tendons and ligaments are frequently seen, particularly at the calcaneus, ischial tuberosities, iliac crest, femoral trochanters, supraspinatus insertion, and spinous processes of the vertebrae. In the early stages of the evolution of syndesmophytes, there is inflammation of the superficial layers of the annulus fibrosus, with subsequent reactive sclerosis and erosions of the adjacent corners of the vertebral bodies. This combination of destructive osteitis and repair leads to “squaring” of the vertebral bodies. This squaring is associated with gradual ossification of the annulus fibrosus and eventual “bridging” between vertebrae by syndesmophytes. There are often concomitant inflammatory changes, ankylosis in the apophyseal joints, and ossification of the adjacent ligaments. In a number of patients, this may ultimately result in a virtually complete fusion of the vertebral column (“bamboo spine”).

Hip involvement may lead to symmetric, concentric joint spondylarthritic narrowing, irregularity of the subchondral bone with subchondral sclerosis, osteophyte formation at the outer margin of the articular surface, and, ultimately, bony ankylosis of these joints.

Several validated scoring methods are available to quantify structural damage in AS: the Bath AS radiology index (BASRI), the Stoke AS spondylitis score (SASSS), and the modified SASSS. The BASRI includes scores for the cervical and lumbar spine, as well as the SI joints. A similar score for the hips is also available. The SASSS evaluates the lumbar spine only; the modified SASSS assesses the anterior, cervical, and lumbar spine. These scoring methods are most suited for use in clinical trials and observational studies.

**Computed Tomography and Magnetic Resonance Imaging**

The conventional plain pelvic radiograph is still the initial tool for the evaluation of SI joints in patients with inflammatory low back pain. This technique, however, lacks sensitivity in early sacroiliitis because it only detects structural abnormalities that are the consequence of inflammation. CT may detect bone abnormalities such as sclerosis and erosion sooner than plain radiography, but its use is limited by radiation exposure and it does not detect changes in soft tissues or bone marrow where early features of sacroiliitis develop. MRI sequences that are used in routine practice to image the SI joint can detect sacroiliitis in 50% of patients with preradiographic SpA. These sequences include fat-saturating techniques, such as short tau inversion recovery (STIR), that are very sensitive in the detection of bone marrow edema, which is a frequent finding in AS-related inflammation of the musculoskeletal system (Figure 75-2).

The T1-weighted sequence allows detection of erosions and fat infiltration, which may occur early in disease and is associated with resolution of inflammation. Diagnostic MRI should include both sequences, while the role of contrast enhancement with gadolinium is still unclear due to its substantial cost and prolonged time for imaging, which patients may find difficult to endure.

Similarly, plain spinal radiographs only show abnormalities once disease is well established, and it plays a minor role in diagnostic or ongoing routine evaluation except to detect other causes of back pain, such as spinal fractures. Characteristic abnormalities include square vertebrae, shiny corners (the Romanus lesion), spondylodiscitis (the Anderson lesion), and syndesmophytes with partial and complete fusion. Spinal inflammation can only be visualized by MRI, where it is typically seen as bone marrow edema in the vertebrae at both anterior and posterior vertebral corners as well as around the intervertebral disk. Lateral and posterior...
elements such as the costovertebral and costotransverse joints, facet joints, pedicles, and spinal ligaments can also show inflammatory lesions. MRI and ultrasonography can be very useful to assess enthesis problems such as Achilles tendinitis and heel pain.

Quantification of spinal inflammation using MRI is highly sensitive to change and is increasingly recognized as an essential component of clinical trials. MRI scores for inflammation correlate with CRP but not with symptoms. Increasing evidence supports an association between MRI inflammation and the future development of structural changes on radiography.

**DIAGNOSIS**

Clinical manifestations of AS usually begin in late adolescence or early adulthood; only rarely do they begin after age 40 years. The diagnosis of AS at an early stage of disease depends primarily on a careful history and physical examination. Two features of the history are critical: (1) the presence of inflammatory low back pain and stiffness and (2) a positive family history for AS.

Low back pain is common in the general population and is frequently due to noninflammatory, nonspecific mechanical causes. However, the low back pain in AS has typical “inflammatory” features (see Table 75-6). A history of inflammatory low back pain can be used as a diagnostic tool. A reassessment of the clinical history for diagnostic purposes among young to middle-aged adults (younger than 50 years) with chronic back pain and an established diagnosis of either AS or mechanical back pain revealed a sensitivity of 37% (95% CI, 28 to 46), a specificity of 84% (95% CI, 76 to 90), a positive likelihood ratio of 2.3 (95% CI, 1.4 to 3.7), and a post-test probability of AS of 11% (given a pretest probability of 5%) if two of the four parameters listed in Table 75-9 were present. If three or four of these items were present, the sensitivity was 34% (95% CI, 25 to 43), the specificity was 97% (95% CI, 92 to 99), the positive likelihood ratio was 12.4 (95% CI, 4.0 to 40), and the post-test probability of AS was 39%. Because the prevalence of AS in many white populations is as low as approximately 0.1% to 0.3%, applying the clinical history as a test for the disease in such low-probability settings provides rather low post-test probability values. However, a positive family history increases the pretest probability of AS from 0.1% for a person belonging to the general population to about 10% for any first-degree relative of an AS proband. The probability of having AS for a first-degree relative with a positive family history of AS increases from 10% to nearly 50% if this relative has inflammatory low back pain. In contrast, the likelihood of having AS increases from 0.1% to only 1% for a person who has inflammatory back pain (without any other inflammatory indications listed in Table 75-9) but has a negative family history for AS.

A definite diagnosis of AS is usually established by radiographic evidence of bilateral sacroiliitis. The plain anteroposterior view of the pelvis is usually adequate for diagnostic purposes. There is, however, considerable intraobserver and interobserver variation in the radiographic diagnosis of sacroiliitis for both conventional pelvic films and CT of the SI joints. Training in reading these films has limited value. Improvement in sensitivity tends to be associated with a decrease in specificity.

In most adult patients, AS can be diagnosed clinically without the HLA-B27 test. This assessment has no additional value in established disease or as a pure screening tool. However, in young patients with inflammatory chronic back pain, a positive HLA-B27 test increases the likelihood of having AS, particularly if imaging of the SI joints shows inflammatory low back pain and equivocal pelvic radiograph demonstrating the following features: A, Bone marrow edema in left iliac and sacral bones. B, Capsular inflammation. C, Joint space inflammation. D, Diffuse erosion of left iliac bone with widening of joint space. E, Fat infiltration in right sacrum and ilium.

**Table 75-9 Proposed Criteria for Inflammatory Back Pain in Young to Middle-aged Adults* with Chronic Back Pain**

<table>
<thead>
<tr>
<th>Morning stiffness of at least 30 min duration</th>
<th>Improvement of back pain with exercise but not with rest</th>
<th>Awakening because of back pain during second half of night only</th>
<th>Alternating buttock pain</th>
</tr>
</thead>
</table>

*Younger than 50 yr.

joints does not provide conclusive results. Usually, however, the contribution of HLA-B27 typing to purely clinical factors in diagnosing axial manifestations of AS among patients with inflammatory back pain of short duration is rather limited, whereas MRI may help in classifying patients as having spondyloarthritis or nonspondyloarthritis.\textsuperscript{119}

Physicians are reluctant to make the diagnosis of AS when radiographic evidence of sacroiliitis is not present. Relatives of AS patients in particular may have signs and symptoms of AS including inflammatory back pain but sometimes do not show radiographic sacroiliitis even after lengthy follow-up. Radiographic sacroiliitis is frequent in AS but is by no means an early or obligate manifestation of the disease. In patients with a clinical diagnosis of possible AS, radiographic sacroiliitis may never become manifest or only after appropriate follow-up. Therefore diagnosing this type of disease, which fits entirely into the concept of axial spondyloarthritis, early, before (conventional) radiographic evidence of sacroiliitis is manifest, constitutes a challenge to the clinician. This is especially true as more effective treatments become increasingly available. In this context, the term (nonradiographic) axial spondyloarthritis is often used. An approach based on pretest probabilities and likelihood ratios has been proposed to diagnose early disease with predominantly axial manifestations before convincing evidence of radiographic sacroiliitis is present.\textsuperscript{81} The majority of patients diagnosed clinically with preradiographic axial spondyloarthritis have MRI evidence of spondyloarthritis,\textsuperscript{120} and the severity of bone edema on MRI has been shown to predict the development of radiographic sacroiliitis on follow-up.\textsuperscript{121} Consequently, MRI is now an acceptable imaging criterion for the diagnosis of spondyloarthritis and is especially useful when the history points to inflammatory back pain, the pelvic radiograph is normal or equivocal, and the patient is positive for B27.

AS rarely develops after age 40; however, late-onset AS does occur. In this case, there may be little or no clinical involvement of the axial skeleton initially, but patients may show moderate oligoarthritis with low cell counts in the synovial fluid and pitting edema of the lower limbs.\textsuperscript{122}

At the other end of the age scale, juvenile-onset AS is not uncommon among young patients with spondyloarthritis. Such patients tend to have enthesitis and peripheral arthritis, which may be severe and disabling.

**ANKYLOSING SPONDYLITIS IN MALES AND FEMALES**

Clinically, AS is more common in males, with a reported male-to-female ratio of about 2 : 1 to 3 : 1. However, extrapolation of studies employing the genetic marker HLA-B27 suggests that, on the basis of radiographs of the SI joints, prevalence rates are about equal in both sexes.\textsuperscript{8}

Disease expression is thought to be different in males and females. A case-control study comparing 35 female patients with 70 male patients as controls showed no differences in spinal symptoms, chest expansion, peripheral arthritis, extra-articular manifestations, or functional outcome. The males with AS more often had radiographic spinal changes and hip joint involvement than their female counterparts. There is still some controversy, but overall, there are no significant clinical or radiographic differences between women and men with AS. However, on average, the disease seems to be more severe in men.\textsuperscript{113,124}

Fertility among female patients with AS is normal.\textsuperscript{125} Most patients (50% to 60%) do not experience major changes in disease activity during pregnancy, but an increase in morning stiffness and low back pain, particularly at night, may occur at about the 20th week of gestation and last for a few days to weeks.\textsuperscript{126} In about 50% of patients, an exacerbation of symptoms is seen within the first half year after delivery. Sacroiliitis including complete ankylosis of the SI joints does not constitute a contraindication for vaginal delivery. Epidural anesthesia is usually possible because most patients have a rather short duration of disease and do not have extensive spinal syndesmophytes. The fetal outcome is not impaired in patients with AS. Every pregnancy in patients with this disease should be considered potential high risk, however, and such pregnancies require close collaboration between rheumatologists and obstetricians.

An uncontrolled study among 612 AS patients (mean age, 50.8 years; 71.6% males) reported a substantial impact on sexual relationships (response rate, 38%). Poor function, depression, disease activity, unemployment, and poor self-efficacy were independently associated with greater impact on patients' sexual relationships.\textsuperscript{127}

**OUTCOME**

The course of AS is highly variable, characterized by spontaneous remissions and exacerbations. Its prognosis has generally been considered rather favorable. The disease may run a relatively mild or self-limited course. However, the disease may also remain active over many years. Life expectancy is somewhat reduced, particularly after 10 years of disease.\textsuperscript{128} A study from Finland indicates that the risk of dying for patients with AS is increased by 50% compared with controls matched for age and gender. Causes of death include complications of the disease such as amyloidosis and spinal fractures, as well as cardiovascular, gastrointestinal, and renal disease.\textsuperscript{129} There is no convincing evidence that the natural history of the disease has essentially changed over the past few decades.\textsuperscript{130,131} No differences exist between familial and sporadic AS in terms of age at onset, age at diagnosis, or prevalence of peripheral arthritis and acute anterior uveitis.\textsuperscript{132}

Functional limitations increase with disease duration. Although structural damage seen on radiographs is clearly associated with physical function and spinal mobility at the group level, individual patients with normal radiographs might exhibit a major reduction in spinal mobility, whereas those with severe radiographic abnormalities might function quite well in everyday tasks.\textsuperscript{133}

Recent data show that the functional prognosis of AS is less favorable than was previously thought. Withdrawal from work in those with paid jobs varies from 10% after 20 years of disease duration to 30% after 10 years, depending on the characteristics of the patients included and the social security system considered.\textsuperscript{134,135} The age- and sex-adjusted withdrawal rate from labor-force participation was 3.1 times higher among Dutch patients compared with the general population.\textsuperscript{137} Older age at disease onset, manual work, lower educational level, and coping strategies characterized
by the limitation and pacing of activities were associated with a higher risk for work disability.\textsuperscript{115-117} Vocational counseling, job training, easy access to the workplace, and support of colleagues and management may reduce the probability of withdrawal from work.\textsuperscript{134,138} Sick leave in those with paid jobs was linked to disease activity and presence of extraspinal disease manifestations.\textsuperscript{134,137,139} Patients with peripheral joint involvement are more likely to take sick leave than are AS patients with axial manifestations only.

Overall, the first 10 years of disease are particularly important with respect to subsequent outcome. Most of the loss of function among patients with AS occurs within this period and is associated with the presence of peripheral arthritis, spinal radiographic changes, and development of a so-called bamboo spine.\textsuperscript{140} In a retrospective study of patients with spondyloarthropathies including AS, of at least 10 years’ duration, seven variables were associated with disease severity if these factors occurred within the first 2 years of follow-up. These factors, expressed as an odds ratio together with its 95% CI, are as follows: arthritis of hip joints (22.9; 4.4 to 118), ESR more than 30 mm/hr (7; 4.8 to 9.5), poor efficacy of nonsteroidal anti-inflammatory drugs (NSAIDs) (8.3; 2.6 to 27.1), limitation of lumbar spine (7; 2 to 25), sausage-like digits (8.5; 1.5 to 9.0), oligoarthritis (4.3; 1.4 to 13.1), and onset before age 16 years (3.5; 1.1 to 12.8).\textsuperscript{141} Although radiographic progression is highly variable among patients, radiographic evidence of spinal involvement, especially the presence of syndesmophytes, appears to be the primary factor that independently predicts further radiographic progression.\textsuperscript{142} The long-term results of total hip replacement in AS are satisfactory. The outcome of 138 total hip replacements and 12 revisions was good or very good in 86%, and 63% of patients had no pain. Mobility was good or very good in 44%. The mean follow-up was 7.5 years (range, 1 to 34 years). Altogether, 69% of the male hip recipients younger than 60 years were at work at the time of the survey.\textsuperscript{143}

### ASSESSMENT AND MONITORING

Signs and symptoms such as spinal pain and limitation of motion might be due to current disease activity or to damage. A plethora of tools is available to assess these dimensions. For example, there are many ways to measure limitation of motion of the lumbar spine. New instruments have been developed to assess various aspects of the disease including the Bath and Edmonton AS metrology indices, Bath AS global index, BASRI, Bath AS disease activity index, and Dougados functional index.\textsuperscript{144,149} However, standardization and validation of many of these instruments are lacking or incomplete. An international Assessment in Ankylosing Spondylitis (ASAS) working group was formed with the aim of selecting, proposing, and testing core sets of measures for different settings.\textsuperscript{150} It was thought that a certain set of variables should be targeted to a specific task. For example, when assessing the efficacy of physical therapy, it would not be realistic to include measures of radiographic changes of the spine. Clearly, the set of measures for a drug’s disease-modifying capabilities will differ from a set that measures analgesic effectiveness only. Four settings have been defined: disease-controlling antirheumatic therapy; symptom-modifying antirheumatic drugs such as NSAIDs; physical therapy; and clinical record keeping in daily practice (Table 75-10).\textsuperscript{150} Also, criteria to assess the response of individual patients have been developed and validated. These ASAS-20 improvement criteria are frequently used in clinical trials.\textsuperscript{151} In addition, more stringent improvement criteria—ASAS-40 and ASAS-5/6—have been proposed,\textsuperscript{152} as well as criteria to define partial remission. The three sets of improvement criteria and the partial remission criteria are presented in Table 75-11.\textsuperscript{151,152} The recently developed Ankylosing Spondylitis Disease Activity Score (ASDAS) enables defining disease activity states and demonstrating improvement by applying cutoff values.\textsuperscript{153}

A needs-based quality-of-life instrument specific for AS has been developed. It is well accepted and easy to perform and, in terms of assessing the impact of interventions, has shown good scaling and psychometric properties and sensitivity to change.\textsuperscript{154,155}

### MANAGEMENT

A systematic review of the literature on the management of AS culminated in a series of treatment propositions developed by ASAS/European League Against Rheumatism (EULAR) that emphasize the key evidence-based components of disease management (Table 75-12; Figure 75-3).\textsuperscript{156,157} For most patients, AS is a relatively mild disease with a good functional prognosis. Most do not experience significant extraskeletal manifestations except for acute anterior

<table>
<thead>
<tr>
<th>Domain</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>BASFI or Functional Index Dougados</td>
</tr>
<tr>
<td>Pain</td>
<td>VAS: last week, spine pain due to AS</td>
</tr>
<tr>
<td>Spinal mobility</td>
<td>VAS: last week, spine pain due to AS; Chest expansion and modified Schober and occiput to wall distance (lateral spinal flexion or BASM1)</td>
</tr>
<tr>
<td>Patient global assessment</td>
<td>VAS: last week</td>
</tr>
<tr>
<td>Stiffness</td>
<td>Duration of morning spine stiffness, last week</td>
</tr>
<tr>
<td>Peripheral joints and entheses</td>
<td>Number of swollen joints (44 joint count); validated enthesis index</td>
</tr>
<tr>
<td>Acute-phase reactants</td>
<td>Erythrocyte sedimentation rate</td>
</tr>
<tr>
<td>Spine radiographs</td>
<td>Lateral view of lumbar spine and lateral view of cervical spine</td>
</tr>
<tr>
<td>Hip radiographs</td>
<td>Pelvic radiograph including sacroiliac joints and hips</td>
</tr>
<tr>
<td>Fatigue</td>
<td>VAS on fatigue from BASDAI</td>
</tr>
</tbody>
</table>

Disease-controlling antirheumatic therapy domains: 1-10; symptom-modifying antirheumatic drug domains: 1-5, 10; physical therapy domains: 1-5, 10; clinical record-keeping domains: 1-7.

AS, ankylosing spondylitis; BASDAI, Bath ankylosing spondylitis disease activity index; BASFI, Bath ankylosing spondylitis functional index; BASMI, Bath ankylosing spondylitis metrology index; VAS, visual analogue scale.

uveitis, which occurs in about 30% of patients. Usually, this eye disease can be well managed with eye drops containing corticosteroids to reduce inflammation and with pupil-dilating, atropine-like agents to prevent or diminish synechiae. At the outset, patients should be warned that acute anterior uveitis may occur at any time during the course of the disease.

The treatment objectives in AS are to relieve pain, stiffness, and fatigue and to maintain good posture and good physical and psychosocial functioning. No drug is currently available that significantly influences the course of spinal disease and retards the process of ossification in particular. Similarly, evidence is lacking that any of the conventional disease-modifying antirheumatic drugs including sulfasalazine and methotrexate alter or inhibit the inflammation seen in the spine and enthesis in AS.

A full explanation of the disease, its course, possible complications (e.g., acute anterior uveitis), and its prognosis is essential to achieve compliance by the patient. Self-help groups provide important information and social support. In addition, patient organizations often provide access to hydrotherapy and group physiotherapy. Exercises are the mainstay of treatment. Preferably, they should be started after a hot shower or a hot bath. Swimming and extension-promoting exercises or sporting activities such as volleyball or cross-country skiing are appropriate. These activities counteract the kyphotic effects of pain and fatigue on posture and reduce stiffness. Patients should avoid vigorous or contact sports if the spine has become fused or osteoporotic because such a spine is susceptible to fracture.

Appliances such as driving mirrors may improve comfort and safety, especially if there is considerable involvement of the cervical spine. In that case, appropriate neck support is also required to reduce the risk of fracturing the vulnerable osteoporotic cervical spine as a consequence of traffic accidents. For the same reason, automobile air bags are strongly recommended.

### Physiotherapy

Evidence indicates that physiotherapy in the form of exercises is effective, at least in the short term (up to 1 year). In a randomized, controlled trial, a program of supervised physiotherapy in groups was superior to individualized programs in improving thoracolumbar mobility and fitness. The program, which consisted of hydrotherapy, exercises, and sporting activities twice weekly for 3 hours per session, resulted in improved overall health and less stiffness, as reported by the patient. An intensive 3-week spondyloarthritis exercise therapy program resulted in marked improvement in both subjective and objective assessments that lasted for up to 9 months. Health resource utilization, in particular NSAID use and sick leave, was significantly reduced during this 9-month follow-up period. The clinical benefits of such treatments can be achieved at acceptable costs.

A Cochrane review concluded that a home exercise program is better than no intervention, supervised group physiotherapy is better than home exercises, and that combined inpatient spa-exercise therapy followed by supervised outpatient weekly group physiotherapy is better than weekly group physiotherapy alone (level A evidence) (Table 75-13). The tendency toward positive effects of physiotherapy in the management of AS calls for further research in this field. New trials should also address other physiotherapy interventions commonly used in clinical practice. Moreover, the impact on physical functioning of physiotherapy in patients who respond well to biologics needs

### Table 75-11  Assessment in Ankylosing Spondylitis (ASAS) International Working Group Improvement Criteria and Partial Remission Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>BASFI</th>
<th>Morning stiffness</th>
<th>Patient global assessment</th>
<th>Pain</th>
<th>Acute-phase reactants</th>
<th>Spinal mobility</th>
<th>Pain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASAS-20 Improvement Criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least 20% improvement and 10 units improvement in 3 of the 4 following domains, without 20% or more worsening and 10 units worsening in the remaining domain:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ASAS-40 Improvement Criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least 40% improvement and 20 units improvement in 3 of the 4 following domains, without any worsening in the remaining domain:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ASAS-5/6 Improvement Criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least 20% improvement in 5 of the 6 following domains:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ASAS Partial Remission Criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A value below 20 units in all 4 domains of the ASAS-20 improvement criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ASAS/EULAR** recommendations for the management of AS

- **Axial disease**
  - NSAIDs
  - Sulfasalazine
  - Local corticosteroids
  - TNF blockers

- **Peripheral disease**
  - Education, exercise, physical therapy, rehabilitation, patient associations, self help groups

- **Analgesics**
  - NSAIDs (nonsteroidal anti-inflammatory drugs)

- **Ankylosing Spondylitis**

**Figure 75-3** Recommended management of ankylosing spondylitis (AS), based on clinical expertise and research evidence. The disease progression with time moves vertically from top to bottom. ASAS/EULAR, Assessment in Ankylosing Spondylitis/European League Against Rheumatism; NSAIDs, nonsteroidal anti-inflammatory drugs; TNF, tumor necrosis factor.
Table 75-12  First Update of the ASAS/EULAR Recommendations for the Management of Ankylosing Spondylitis (AS)

| The overreaching principles of the management of patients with AS are: |
| AS is a potentially severe disease with diverse manifestations, usually requiring multidisciplinary treatment coordinated by the rheumatologist |
| The primary goal of treating the patient with AS is to maximize long term health-related quality of life through control of symptoms and inflammation, prevention of progressive structural damage, preservation/normalization of function, and social participation |
| Treatment of AS should aim at the best care and must be based on a shared decision between the patient and the rheumatologist |
| The optimal management of patients with AS requires a combination of nonpharmacologic and pharmacologic treatment modalities |

1. General Treatment

The treatment of patients with AS should be tailored according to:
- The current manifestations of the disease (axial, peripheral, entheseal, extra-articular symptoms and signs)
- The level of current symptoms, clinical findings, and prognostic indicators
- The general clinical status (age, gender, comorbidity, concomitant medications, psychosocial factors)

2. Disease Monitoring

The disease monitoring of patients with AS should include:
- Patient history (e.g., questionnaires)
- Clinical parameters
- Laboratory tests
- Imaging
  - All according to the clinical presentation as well as the ASAS core set
- The frequency of monitoring should be decided on an individual basis depending on:
  - Course of symptoms
  - Severity
  - Treatment

3. Nonpharmacologic Treatment

The cornerstone of nonpharmacologic treatment of patients with AS is patient education and regular exercise
- Home exercises are effective.
- Physical therapy with supervised exercises, land or water based, individually or in a group, should be preferred as these are more effective than home exercises
- Patient associations and self-help groups may be useful

4. Extra-articular Manifestations and Comorbidities

The frequently observed extra-articular manifestations (e.g., psoriasis, uveitis, and inflammatory bowel disease) should be managed in collaboration with the respective specialists
- Rheumatologists should be aware of the increased risk of cardiovascular disease and osteoporosis

5. Nonsteroidal Anti-inflammatory Drugs (NSAIDs)

NSAIDs, including coxibs, are recommended as first-line drug treatment for AS patients with pain and stiffness
- Continuous treatment with NSAIDs is preferred for patients with persistently active, symptomatic disease
- Cardiovascular, gastrointestinal, and renal risks should be taken into account when prescribing NSAIDs

6. Analgesics

Analgesics, such as paracetamol and opioid (like) drugs, might be considered for residual pain after previously recommended treatments have failed, are contraindicated, and/or are poorly tolerated

7. Glucocorticoids

Corticosteroid injections directed to the local site of musculoskeletal inflammation may be considered
- The use of systemic glucocorticoids for axial disease is not supported by evidence

8. Disease-modifying Antirheumatic Drugs (DMARDs)

There is no evidence for the efficacy of DMARDs, including sulfasalazine and methotrexate, for the treatment of axial disease
- Sulfasalazine may be considered in patients with peripheral arthritis

9. Anti–Tumor Necrosis Factor (TNF) Therapy

- Anti-TNF therapy should be given to patients with persistently high disease activity despite conventional treatments according to the ASAS recommendations
- There is no evidence to support the obligatory use of DMARDs before or concomitant with anti-TNF therapy in patients with axial disease
- There is no evidence to support a difference in efficacy of the various TNF inhibitors on the axial and articular/entheseal disease manifestations, but in the presence of inflammatory bowel disease a difference in gastrointestinal efficacy needs to be taken into account
- Switching to a second TNF blocker might be beneficial, especially in patients with loss of response
- There is no evidence to support the use of biologic agents other than TNF inhibitors in AS

10. Surgery

Total hip arthroplasty should be considered in patients with refractory pain or disability and radiographic evidence of structural damage, independent of age
- Spinal corrective osteotomy may be considered in patients with severe disabling deformity
- In patients with AS and an acute vertebral fracture, a spinal surgeon should be consulted

11. Changes in the Disease Course

If a significant change in the course of the disease occurs, causes other than inflammation, such as a spinal fracture, should be considered and appropriate evaluation, including imaging, should be performed

ASAS, Assessment in Ankylosing Spondylitis; EULAR, European League Against Rheumatism.
to be assessed. Although these patients may experience relief of symptoms, the formation and progression of syndesmophytes may not be influenced by TNF-blocking agents. This reinforces the ongoing need for supplementary therapeutic modalities such as physiotherapy in some patients with AS.

Lying prone for 15 to 30 minutes once or several times a day is useful to reverse the tendency toward kyphosis, which is aggravated by pain and fatigue, as well as flexion contractures of the hip joints. Patients should sleep fully supine on a firm mattress with only a small neck-support pillow.

Medication

Nonsteroidal Anti-inflammatory Drugs

The efficacy and effectiveness of NSAID therapy for the alleviation of symptoms have been well established (level A evidence). When given for prolonged periods of up to a year, there may be improvement in spinal mobility and acute-phase reactants. Many NSAIDs are effective in patients with AS, and no NSAID has documented superiority in terms of efficacy. Selective cyclooxygenase-2 (COX-2) inhibitors have similar efficacy to conventional NSAIDs (level A evidence). A nonselective NSAID is appropriate for most patients with AS, who tend to be relatively young and without comorbidity. A COX-2–selective agent may be used in the presence of risk factors for peptic ulceration, although both categories of NSAIDs may exacerbate inflammatory bowel disease. Once-daily drug regimens may improve patient compliance. Up to 2 weeks may be required to demonstrate maximal symptomatic benefit from an NSAID. If symptomatic relief is inadequate, a switch to another NSAID may be worthwhile; failure of two NSAIDs should prompt an exploration of other management strategies.

Given the gastrointestinal and cardiovascular risks of taking NSAIDs or coxibs, one must address whether this treatment should be on a daily or an “on-demand” basis. A 2-year randomized, prospective, controlled trial in AS patients compared the efficacy of continuous NSAID therapy with that of intermittent on-demand use. The results suggest that continuous therapy retards radiographic disease progression. This study is in line with an older study that also suggested a possible disease-controlling effect of continuous therapy. However, these findings require confirmation by studies less liable to bias.

Second-Line Drugs

Borrowing well-established concepts in the treatment of RA, disease-controlling therapy for AS has been defined as an agent that decreases inflammatory manifestations of disease, sustains or improves function, and prevents or decreases the rate of progression of structural damage. Although most of the second-line agents developed primarily for RA have been studied in AS, none can be considered disease controlling in AS. The greatest amount of data is available for sulfasalazine, which was first proposed as a therapy for AS in 1984, based on the common association between inflammatory bowel disease and spondyloarthritis, the description of inflammatory lesions in the ileum of patients with spondyloarthritis, and its success in the treatment of intestinal inflammation. A total of 11 double-blind, placebo-controlled trials have been published, as well as two meta-analyses. The results of the two largest trials were consistent, demonstrating no significant benefit for sulfasalazine in AS, although subgroup analysis showed that patients with (peripheral) polyarthritis—mostly those with psoriatic arthritis, but also AS patients with peripheral joint involvement—had a significant but modest response. An important limitation in most of these studies was the long disease duration (>10 years) of recruited patients, and it has been suggested that early disease may be more responsive to therapy. However, a 24-week placebo-controlled trial that recruited 230 patients meeting European Spondyloarthritis Study Group (ESSG) criteria for spondyloarthritis and with symptom duration of less than 5 years confirmed that sulfasalazine was ineffective. The most recent meta-analysis, based on 11 trials, concluded that this agent has a significant impact only on the ESR and the severity of spinal stiffness (level A evidence). The primary indication for the use of sulfasalazine in routine practice is a patient who has concomitant peripheral arthritis and has had an inadequate response to NSAIDs and physical modalities.

The evaluation of methotrexate in AS has been limited to case reports and open analyses, mostly reported in abstract form. These studies have included limited numbers of patients for periods of 6 months to 3 years, at doses from 7.5 to 15 mg weekly. The results have been mixed, with some benefit noted in patients with concomitant peripheral arthritis. Two small placebo-controlled trials assessed methotrexate in doses of 10 and 7.5 mg weekly for 24 weeks, with contradictory findings. A meta-analysis concluded that there was no evidence of efficacy and that higher-quality trials, larger sample sizes, longer durations of treatment, and higher dosages of methotrexate were necessary before any definitive conclusions could be drawn (level B evidence).

Corticosteroids may be effective for local intra-articular treatment in AS including the SI joints, if given under fluoroscopic guidance (level B evidence). Systemic steroids are of unproven benefit and are thought to be less effective than in RA (level C evidence). Leflunomide has been studied in AS, and although an open-label study suggested a benefit in patients with peripheral arthritis, a small placebo-controlled study reported no benefit (level B evidence). A controlled dose-response (60 mg vs. 10 mg) evaluation of a bisphosphonate, pamidronate, given intravenously on
a monthly basis for 6 months showed evidence of symptomatic efficacy, primarily in patients with only axial disease (level B evidence). However, this finding needs to be confirmed.

Thalidomide has been used in two open-label studies in AS because it enhances the degradation of TNF messenger RNA. In a Chinese study, improvement was reported in 80% of patients, with deterioration 3 months after treatment discontinuation. Frequent side effects are drowsiness, constipation, and dizziness (level B evidence).

**Biologic Therapies**

A milestone in the treatment of AS is the development of anti-TNF therapies. The rationale is based on the finding of TNF expression in SI joint biopsies of AS patients, the observation that overexpression of TNF leads to sacroiliitis in animal models, and earlier clinical trial data demonstrating the efficacy of one anti-TNF agent, infliximab, in Crohn’s disease. Four anti-TNF agents are of proven benefit in AS according to pivotal phase III trials (level A evidence): infliximab, etanercept, adalimumab, and golimumab. Infliximab is an IgG1 chimeric monoclonal antibody with the Fab portion derived from the mouse. It is given in a dose of 3 to 5 mg/kg every 6 to 8 weeks after loading at 0, 2, and 6 weeks. Etanercept is a recombinant 75-kD TNF receptor IgG1 fusion protein that is self-administered by subcutaneous injection either once (50 mg) or twice (25 mg) weekly. Adalimumab and golimumab are human monoclonal antibodies that are self-administered by subcutaneous injection on alternate weeks (40 mg) or monthly (50 mg), respectively. None require concomitant therapy with methotrexate.

All these agents demonstrate ASAS-20 response rates of 55% to 60% and ASAS-40 response rates of 45% to 50% in phase III trials. Even higher response rates have been observed in patients with preradiographic axial spondyloarthritis and short disease duration who received adalimumab or infliximab in placebo-controlled trials. Improvement is evident by 2 to 4 weeks and is sustained as long as the patient remains on treatment; virtually all patients relapse by 4 months after discontinuation of treatment. Significant improvement is also observed in function, spinal mobility, peripheral synovitis, enthesitis score, and quality of life. Sick leave and work disability are reduced. The number of patients who must be treated to achieve one patient who experiences at least 50% improvement in disease activity is just two (95% CI, 1 to 6). Objective parameters of disease activity that show improvement include acute-phase reactants, synovial histopathology, and MRI features of inflammation in the spine and SI joints. As of now, no evidence indicates that these agents are disease controlling with respect to the prevention of structural damage on plain radiography. Response to treatment appears to be increased in those with high disease activity and worse in those with a long disease duration, impaired function, and no discernible evidence of inflammation on MRI. However, patients with complete spinal ankylosis may benefit from these treatments. Adverse events in AS patients are no different from those reported in RA, and infusion reactions in patients receiving infliximab have been no more frequent than in RA patients on concomitant methotrexate. All anti-TNF agents are effective for psoriasis, and the monoclonal anti-TNF antibodies, infliximab and adalimumab, also have demonstrated efficacy in both uveitis and colitis. Patients who fail to respond to one anti-TNF agent may respond to an alternative anti-TNF agent.

Recommendations for the use of anti-TNF therapies have been developed (Table 75-14). These new therapeutic modalities identify important clinical questions to be answered by further research. All the anti-TNF agents examined to date have symptom-modifying properties, but their long-term safety and disease-controlling effects in terms of preventing structural damage have yet to be demonstrated.

**Surgery**

Involvement of the hip joint may cause serious disability. Ectopic bone formation may occur, but the outcome of total hip replacement is generally favorable.

Vertebral osteotomy may be required in selected cases to correct marked flexion deformity when forward vision is severely impaired. Diaphragmatic herniation may result from the procedure.

**SUMMARY**

Although understanding of AS genetics has improved greatly, knowledge about its cause and pathogenesis is far from complete. Much has been accomplished in terms of classification and assessment of the disease. Treatment with biologics such as anti-TNF is effective, but such therapy has not yet been shown to inhibit radiographic progression of spinal syndesmophytes. The challenge now is to determine how to predict and improve outcomes at the level of individual patients, in particular for those patients who will develop important structural changes and functional decline.

**Future Directions**

The long interval (on average) between the first symptoms of AS and the clinical diagnosis must be shortened.

Research into factors that accurately predict final outcome at the time of diagnosis is essential.

Study of which factors predict response to biologic therapy for individual AS patients is highly desirable. In addition, the long-term effects of treatment with biologics are not yet fully known.

Evidence of the effectiveness of preventing damage in terms of the development of syndesmophytes is still contradictory. It is largely unknown what triggers the process of ankylosing.

A few studies suggest that continuous (versus intermittent) use of NSAIDs or coxibs may slow the progression of axial radiographic manifestations of AS. These findings, however, are controversial. Therefore further studies on this issue are required.
Table 75-14  2010 Update of Recommendations for the Use of Anti–Tumor Necrosis Factor (TNF) Agents in Patients with Axial Spondyloarthritis (SpA)

<table>
<thead>
<tr>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient Selection</strong></td>
</tr>
<tr>
<td>Diagnosis Active disease</td>
</tr>
<tr>
<td>Treatmet failure</td>
</tr>
<tr>
<td><strong>Assessment of Disease</strong></td>
</tr>
<tr>
<td><strong>ASAS Core Set of Daily Practice</strong></td>
</tr>
<tr>
<td>Physical function (BASFI or Dougados functional index)</td>
</tr>
<tr>
<td>Spinal mobility (chest expansion, modified Schober and occiput-to-wall distance, and lateral lumbar flexion)</td>
</tr>
<tr>
<td>Stiffness (duration of morning spine stiffness in the past week)</td>
</tr>
<tr>
<td><strong>BASDAI</strong></td>
</tr>
<tr>
<td><strong>Fatigue (VAS)</strong></td>
</tr>
<tr>
<td><strong>Physical function (BASFI or Dougados functional index)</strong></td>
</tr>
<tr>
<td><strong>Responder criteria</strong></td>
</tr>
<tr>
<td><strong>Assessment of Response</strong></td>
</tr>
</tbody>
</table>

*Modified New York criteria (van der Linden et al, 1984): radiologic criterion (sacroiliitis, grade II bilaterally or grade III to IV unilaterally) and at least two out of three clinical criteria (low back pain and stiffness for more than 3 mo that improves with exercise but is not relieved by rest; limitation of the lumbar spine in both the sagittal and frontal planes; limitation of chest expansion relative to normal values correlated for age and sex).

†BASDAI assessed on a 0-10 VAS or NRS.

‡The expert is a doctor, usually a rheumatologist, with expertise in inflammatory back pain and the use of biologic agents. Experts should be locally defined.

ASAS core set for daily practice: physical function (BASFI); pain (VAS/NRS, last week, spine at night, due to AS and VAS/NRS, last week, spine due to AS); spinal mobility (chest expansion, cervical rotation, occiput-to-wall distance, modified Schober, and lateral lumbar flexion or BASMI); patient’s global assessment (VAS/NRS, last week); stiffness (duration of morning stiffness, spine, VAS/NRS, last week); peripheral joints and entheses (number of swollen joints [44 joints count], enthesitis score such as developed in Maastricht, Berlin, or San Francisco); acute-phase reactants (preferably C-reactive protein or other); fatigue (VAS/NRS). AS, ankylosing spondylitis; ASAS, Assessment in SpondyloArthritis international Society; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; BASMI, Bath Ankylosing Spondylitis Metrology Index; DMARD, disease-modifying antirheumatic drug; NRS, numeric rate scale; NSAIDs, nonsteroidal anti-inflammatory drugs; VAS, visual analogue scale.


**Selected References**


Full references for this chapter can be found on [www.expertconsult.com](http://www.expertconsult.com).
References


major individual variations in a large proportion of patients, J Rheumatol 36(5):997–1002, 2009.


178. Barkham N, Keen HI, Counsell LC, et al: Clinical and imaging efficacy of infliximab in HLA-B27-positive patients with magnetic resonance


KEY POINTS

- Reactive arthritis is a form of spondyloarthritis triggered by particular infections.
- Undifferentiated spondyloarthritis can have both peripheral and axial features.
- Undifferentiated spondyloarthritis is a provisional diagnosis—many cases will evolve into other forms of spondyloarthritis.
- Diagnosis of reactive arthritis rests on symptoms and signs of spondyloarthritis, including extra-articular disease, and evidence of preceding infection.
- Reactive arthritis is often self-limited, but chronic forms require disease-modifying antirheumatic drugs (DMARDs) and even biologics.
- Treatment of undifferentiated spondyloarthritis depends on whether axial or peripheral disease is predominant.

Two forms of spondyloarthritis will be reviewed in this chapter: reactive arthritis and undifferentiated spondyloarthritis. The other members of the spondyloarthritis group—ankylosing spondylitis, psoriatic arthritis, and arthritis associated with inflammatory bowel disease—are described elsewhere, and current thinking on common features that operate in the pathogenesis of all forms of spondyloarthritis can be found in Chapter 74.

DEFINITIONS AND TERMINOLOGY

**Reactive Arthritis**

The term reactive arthritis is sometimes used loosely, and unhelpfully, to mean “any arthritis that comes on after some kind of infection,” that is, as a “reaction” to infection. In this way, diseases such as Lyme disease and rheumatic fever are sometimes termed “reactive,” as are postviral forms of arthritis. However, this is potentially confusing; it is better to have a broad category of postinfectious arthritis and to reserve the term reactive arthritis for the arthritis that follows infection and shares features with other forms of spondyloarthritis. These include clinical features such as frequent evidence of enthesitis, in addition to arthritis, extra-articular features, particularly those involving eyes and skin, and, as required for a spondyloarthritis family member, a clear association with HLA-B27. Using this definition, a relatively small list of bacteria (and no viruses) are common triggers of reactive arthritis (Table 76-1), and a longer “tail” of infections have occasionally been reported as causes. These organisms principally infect the gastrointestinal and genitourinary tracts, although *Chlamydia pneumoniae* (now sometimes termed *Chlamydophila pneumoniae*) is an exception because it causes respiratory infection.

Another unhelpful term is Reiter’s syndrome or Reiter’s triad—consisting of urethritis, conjunctivitis, and arthritis. There are four main reasons for consigning this term to the dustbin of history. First, Hans Reiter was by no means the first to describe reactive arthritis—on this basis, reactive arthritis would be Leroy-Fiessinger-Reiter syndrome, a term that obviously lacks utility. Second, Reiter erroneously attributed the postdysenteric cases he described to spirochetal infection. Third, Reiter had an association with the Third Reich. Fourth, the inclusion of urethritis in Reiter’s triad leads to the mistaken assumption that cases of reactive arthritis with urethritis are likely to be due to sexually acquired infection. This is not the case because urethritis is not uncommon in patients whose reactive arthritis is triggered by enteric infection, especially HLA-B27 patients, and the cases Reiter described in World War I had arthritis associated with dysentery. Therefore, Reiter’s syndrome is at best a synonym for reactive arthritis and is neither useful nor required, but the “triad” does not distinguish a clinically important subgroup of reactive arthritis.

Reactive arthritis secondary to gastrointestinal infection is sometimes bracketed with spondyloarthritis associated with inflammatory bowel disease as enteropathic arthritis. However, although links between the gut and arthritis are very important pathologically, particularly in spondyloarthritis, enteropathic arthritis is really an ill-defined overlap term, which often includes other forms of arthritis that do not have classic features of spondyloarthritis but occur in relation to gastrointestinal disorders. Examples include Whipple’s disease and celiac disease.

**Undifferentiated Spondyloarthritis**

Obvious overlap has been noted between different members of the spondyloarthritis family. By definition, patients with undifferentiated spondyloarthritis have arthritis that fails to satisfy diagnostic or classification criteria for one of the other forms. Thus, it may not be legitimate to regard undifferentiated spondyloarthritis as a separate “disease” because over time, patients often develop new features, which means that their disease is no longer “undifferentiated.” The commonest of these is development of radiographic changes in the sacroiliac joints that allow the patient to satisfy criteria for ankylosing spondylitis. This occurs in approximately 60% of patients in many series, and all ankylosing spondylitis patients will pass through an “undifferentiated” phase.
Thus classification criteria are not, and should not be used to designate homogeneous sets of patients with particular features to be recognized as having a specific condition. Classification criteria are devised to allow for distinguishing clinical entities by means of rather wide classification criteria. It must be noted that patients can first be identified as having spondyloarthritis (often termed axial spondyloarthritis) if seen before developing structural changes on radiographs. Likewise, patients may develop psoriasis at some point after the onset of their spondyloarthritis. In cases where there is a family history of psoriasis or features such as dactylitis, which is very common in psoriatic arthritis, the term psoriatic arthritis sine psoriasis is sometimes used, but if patients fail to meet diagnostic criteria for psoriatic arthritis (such as the recently devised Classification Criteria for Psoriatic Arthritis [CASPAR]), they should be classified as having undifferentiated spondyloarthritis. Note that in the absence of a personal or family history of skin or nail psoriasis, patients fulfill the CASPAR criteria for psoriatic arthritis only if they have dactylitis and juxta-articular new bone formation on hand or foot radiographs, and lack rheumatoid factor. In addition, underlying inflammatory bowel disease may declare itself clinically only at some point after the onset of undifferentiated spondyloarthritis.

Finally, as discussed in the following section, it may be difficult to make a certain diagnosis of reactive arthritis, particularly when evidence of a triggering infection is absent or incomplete, resulting in significant overlap between reactive arthritis and undifferentiated spondyloarthritis—hence their being considered together in this chapter.

### CLASSIFICATION CRITERIA FOR REACTIVE ARTHRITIS AND UNDIFFERENTIATED SPONDYLOARTHRITIS

Because both reactive arthritis and undifferentiated spondyloarthritis are members of the spondyloarthropathy group, patients can first be identified as having spondyloarthritis by means of rather wide classification criteria. It must be recognized that classification criteria are devised to allow homogeneous sets of patients with particular features to be defined. This is critical for research studies, ensuring that different investigators report on the same set of patients. Thus classification criteria are not, and should not be used as, diagnostic criteria; clinicians will certainly make confident clinical diagnoses for patients who fail to fulfill particular classification criteria. Nevertheless, such criteria often serve as a convenient checklist of the features that are usually present in a given disease.

The two best known classification criteria for all forms of spondyloarthritis considered together are those devised by Amor (Table 76-2) and the later European Spondyloarthritis Study Group (ESSG) criteria (Table 76-3). The Amor criteria allocate “points” to 12 features characteristic of spondyloarthritis, and classification is based on reaching a specified total score of 6 points. The ESSG criteria are applied to patients with inflammatory back pain or oligoarthritis, with the presence of additional features required for classification. The sensitivity and specificity achieved by the Amor and ESSG criteria vary in different series and inevitably depend on the population to which they are applied, but these values are usually around 80% and 90%, respectively. A more recent development in the classification of spondyloarthritis has been the separation of axial (spine and sacroiliac) inflammation from peripheral arthritis, with recognition that in different forms of spondyloarthritis, one or another of these may predominate, although with time many patients will develop both. To aid classification of axial spondyloarthritis, which traditionally requires radiographic sacroiliitis, as exemplified by the modified New York criteria for ankylosing spondylitis (AS), magnetic resonance imaging (MRI) changes of sacroiliitis have been included in the recently published Assessment of SpondyloArthritis International Society [ASAS] criteria (Table 76-4). MRI sacroiliitis is carefully defined as bone marrow edema on short tau inversion recovery (STIR) sequences, or as osteitis on T1 images with contrast medium, localized to subchondral or periarticular bone marrow. The

### Table 76-1 Organisms Associated with Reactive Arthritis

<table>
<thead>
<tr>
<th>Common</th>
<th>Gastrointestinal Pathogens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmonella species</td>
<td>Campylobacter jejuni and Campylobacter coli</td>
</tr>
<tr>
<td>Yersinia enterocolitica and Yersinia pseudotuberculosis</td>
<td>Shigella flexneri; less commonly, Shigella sonnei or Shigella dysenteriae</td>
</tr>
<tr>
<td>Clostridium difficile</td>
<td>Genitourinary Pathogens</td>
</tr>
<tr>
<td>Chlamydia trachomatis</td>
<td>Mycoplasma species</td>
</tr>
<tr>
<td>Respiratory Pathogens</td>
<td>Chlamydia pneumoniae</td>
</tr>
<tr>
<td>Reported</td>
<td></td>
</tr>
<tr>
<td>Mycobacterium bovis bacillus Calmette-Guérin</td>
<td>Enterotoxigenic Escherichia coli</td>
</tr>
<tr>
<td>and many others in small numbers of case reports</td>
<td></td>
</tr>
</tbody>
</table>

### Table 76-2 Amor Criteria for Spondyloarthritis

<table>
<thead>
<tr>
<th>Clinical Symptoms or Past History of:</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lumbar or dorsal pain during the night, or morning stiffness of lumbar or dorsal spine</td>
<td>1</td>
</tr>
<tr>
<td>Asymmetric oligoarthritis</td>
<td>2</td>
</tr>
<tr>
<td>Buttock pain</td>
<td>1</td>
</tr>
<tr>
<td>Alternating buttock pain</td>
<td>2</td>
</tr>
<tr>
<td>Dactylitis of finger or toe</td>
<td>2</td>
</tr>
<tr>
<td>Heel pain or other well-defined enthesopathy</td>
<td>2</td>
</tr>
<tr>
<td>Iritis</td>
<td>2</td>
</tr>
<tr>
<td>Nongonococcal urethritis or cervicitis within 1 mo of arthritis onset</td>
<td>1</td>
</tr>
<tr>
<td>Acute diarrhea within 1 mo of arthritis onset</td>
<td>1</td>
</tr>
<tr>
<td>Psoriasis, balanitis, or inflammatory bowel disease</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Radiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sacroiliitis (grade ≥2 if bilateral; grade ≥3 if unilateral)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genetic Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-B27* or family history of ankylosing spondylitis, reactive arthritis, uveitis, psoriasis, or inflammatory bowel disease</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response to Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good response to NSAIDs within 48 hr, or relapse within 48 hr if NSAIDs withdrawn</td>
</tr>
</tbody>
</table>

For a definitive diagnosis of spondyloarthritis, ≥6 points is required; 5 points indicates probable spondyloarthritis.

NSAIDs, nonsteroidal anti-inflammatory drugs.
changes are also required to be multiple or present in at least two consecutive slices. Therefore, patients younger than 45 years of age, with at least 3 months’ history of back pain or sacroiliitis on MRI (or on radiographs), plus at least one feature from the list of spondyloarthritis-associated signs and symptoms shown in Table 76-4, fulfill the criteria. In the absence of sacroiliitis, HLA-B27 individuals can still be classified as having axial spondyloarthritis when two or more of these features are present. These ASAS criteria represent a recent and welcome development, but it remains to be determined what the specificity and sensitivity of MRI changes in sacroiliac joints will be when used in routine practice in nonresearch settings.

The main driver for developing the ASAS criteria for axial spondyloarthritis has been the need to define early ankylosing spondylitis without requiring radiologic changes in the sacroiliac joints. Because these often take years to develop, new criteria were required to identify, investigate, and treat patients with early ankylosing spondylitis. Nevertheless, not all patients who fulfill the criteria for axial spondyloarthritis go on to develop ankylosing spondylitis, that is, early ankylosing spondylitis is a subset of axial spondyloarthritis. Patients who do not progress to radiographic changes or other clinical features of ankylosing spondylitis will continue to be classified as having undifferentiated spondyloarthritis.

The new ASAS criteria for axial disease are clearly useful for identifying patients with undifferentiated spondyloarthritis. In reactive arthritis, acute inflammatory back pain is common, but no studies have established the frequency of MRI-defined sacroiliitis in this disease, or of MRI-defined changes in the spine, although the latter are not used in the ASAS classification. Thus the ASAS criteria for axial spondyloarthritis are not likely to be very useful in reactive arthritis.

More relevant to reactive arthritis are the ASAS criteria for classification of peripheral spondyloarthritis, which have also been published and discussed recently18,19 (Table 76-5). These are applied to patients who present with arthritis, particularly asymmetric oligoarticular lower limb arthritis OR enthesitis OR dactylitis. These patients are classified as having peripheral spondyloarthritis if they have one additional feature from the list shown in Table 76-5. In the absence of any of these features, two features from an additional list are required: arthritis, enthesitis, dactylitis, inflammatory back pain, or family history of spondyloarthritis. Note that because these criteria are applied to patients who present with arthritis or enthesitis or dactylitis, this second list gives weight to the occurrence of these symptoms in combination. These criteria achieved a sensitivity of 78% and a specificity of 83%, representing a compromise between lower sensitivity and greater specificity of the Amor or ESSG criteria. How the ASAS criteria will fare in more general use remains to be seen; they were developed in a population in whom 85% were younger than age 45, and 30% had radiographic changes in sacroiliac joints that met modified New York criteria for ankylosing spondylitis, even though they did not have back pain at recruitment.

Used together, ASAS criteria for axial and peripheral spondyloarthritis should classify all those patients who have some form of spondyloarthritis (again recalling that some patients will be diagnosed with spondyloarthritis even though

### Table 76-3 European Spondyloarthropathy Study Group Criteria for Spondyloarthritis

<table>
<thead>
<tr>
<th>For patients with:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflammatory spinal pain</td>
</tr>
<tr>
<td>or</td>
</tr>
<tr>
<td>Synovitis that is asymmetric or is present predominantly in the lower limbs</td>
</tr>
<tr>
<td>+ ≥1 if:</td>
</tr>
<tr>
<td>Family history of spondyloarthropathy</td>
</tr>
<tr>
<td>Psoriasis</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>Urethritis, cervicitis, or acute diarrhea within 1 mo of onset of arthritis</td>
</tr>
<tr>
<td>Alternating buttoc k pain</td>
</tr>
<tr>
<td>Enthesitis</td>
</tr>
<tr>
<td>Sacroiliitis</td>
</tr>
</tbody>
</table>

### Table 76-4 Assessment of SpondyloArthritis international Society Classification Criteria for Axial Spondyloarthritis (SpA)

<table>
<thead>
<tr>
<th>For patients with back pain for ≥3 mo and aged &lt;45 yr:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sacroiliitis on imaging + ≥1 SpA feature</td>
</tr>
<tr>
<td>or</td>
</tr>
<tr>
<td>HLA-B27 + ≥2 other SpA features</td>
</tr>
<tr>
<td>Sacroiliitis on imaging defined as:</td>
</tr>
<tr>
<td>Active acute inflammation on magnetic resonance imaging highly suggestive of sacroiliitis associated with SpA</td>
</tr>
<tr>
<td>or</td>
</tr>
<tr>
<td>Definitive radiographic sacroiliitis according to the modified New York criteria</td>
</tr>
<tr>
<td>SpA features comprising:</td>
</tr>
<tr>
<td>Arthritis</td>
</tr>
<tr>
<td>Enthesitis (heel)</td>
</tr>
<tr>
<td>Inflammatory back pain</td>
</tr>
<tr>
<td>Dactylitis</td>
</tr>
<tr>
<td>Uveitis</td>
</tr>
<tr>
<td>Psoriasis</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>Good response to nonsteroidal anti-inflammatory drugs</td>
</tr>
<tr>
<td>Family history of SpA</td>
</tr>
<tr>
<td>HLA-B27</td>
</tr>
<tr>
<td>Elevated C-reactive protein</td>
</tr>
</tbody>
</table>

### Table 76-5 Assessment of SpondyloArthritis international Society Classification Criteria for Peripheral Spondyloarthritis

<table>
<thead>
<tr>
<th>For patients with:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peripheral arthritis (usually asymmetric, lower limb)</td>
</tr>
<tr>
<td>or</td>
</tr>
<tr>
<td>Dactylitis</td>
</tr>
<tr>
<td>+ 1 of:</td>
</tr>
<tr>
<td>HLA-B27</td>
</tr>
<tr>
<td>Genitourinary or gastrointestinal infection</td>
</tr>
<tr>
<td>Psoriasis</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>Magnetic resonance imaging sacroiliitis</td>
</tr>
<tr>
<td>or</td>
</tr>
<tr>
<td>+ 2 of:</td>
</tr>
<tr>
<td>Arthritis</td>
</tr>
<tr>
<td>Enthesitis</td>
</tr>
<tr>
<td>Dactylitis</td>
</tr>
<tr>
<td>Inflammatory back pain</td>
</tr>
<tr>
<td>Family history of spondyloarthritis</td>
</tr>
</tbody>
</table>
they do not meet these classification criteria). Of those who meet ASAS criteria, some will have documented psoriasis or inflammatory bowel disease, or radiologic sacroiliitis, and therefore can be classified as having a particular form of spondyloarthritis. The remainder will have undifferentiated spondyloarthritis or reactive arthritis. In relation to diagnosis, checking the features included in the classification criteria will allow the clinician to identify patients in whom some form of spondyloarthritis is likely, and will prompt additional clinical examination and investigations that may clarify the diagnosis or even elicit additional criteria.

In relation to reactive arthritis, there is no wholly satisfactory definition of the disease.\(^\text{22}\) One practical proposal is shown in Table 76-6. In the setting of an outbreak of food poisoning, when all those known to have been infected with *Salmonella* or *Campylobacter* can be followed up, all who develop new joint symptoms can be regarded as having reactive arthritis, but inevitably this also includes patients with only arthralgias.\(^\text{21,22}\) Joint symptoms have a high background prevalence in the community, and it is difficult to be certain that those reported by patients following infection are genuinely related to the infection and not to pre-existing conditions that assume new prominence in the context of an infection, or to questionnaires seeking such symptoms.

In practice therefore it is best to regard only patients with new definite joint swelling, enthesitis, or inflammatory back pain as having reactive arthritis—most of these will fulfill the ASAS criteria for peripheral spondyloarthritis. In specialist rheumatology practice, as compared with community surveys, reactive arthritis patients will have clear evidence of joint inflammation and commonly enthesitis. What is often less certain is the nature of any triggering infection; the degree of certainty runs from patients with culture-proven infection with an organism previously known to cause reactive arthritis, to those who may have only a history suggesting a preceding infection or serologic evidence of infection without clinical history. A “gold standard” for reactive arthritis, requiring spondyloarthritis and unequivocal demonstration of preceding infection by culture (or polymerase chain reaction [PCR]), would allow diagnosis of only a minority of cases, but less strict criteria inevitably result in an overlap with undifferentiated spondyloarthritis. Some surveys of patients given the latter diagnosis have produced immunologic evidence, particularly when synovial fluid T cell responses to reactive arthritis–associated organisms are measured, which suggests that a significant proportion (up to 50%) may actually have reactive arthritis.\(^\text{23}\)

### Incidence of Reactive Arthritis and Undifferentiated Spondyloarthritis

Several community studies have examined reactive arthritis, mainly in Scandinavia, where HLA-B27 prevalence is high. A study in Oslo suggested an incidence of 4 to 5 per 100,000 for both *Chlamydia*–induced and enteric infection–related disease,\(^\text{24}\) and similar results were reported in Finland.\(^\text{25}\) Reactive arthritis following gastroenteritis has been systematically investigated by Leirisalo-Repo and colleagues, who studied populations known to have been infected with *Salmonella*, *Campylobacter*, *Yersinia*, and *Shigella* and provided follow-up by questionnaire and clinical examination to determine the incidence of new inflammatory symptoms in the joints and back, and of clinical reactive arthritis.\(^\text{21,22,26,27}\) Together these studies suggest that 7% to 12% of those infected develop reactive arthritis, and other studies agree with this.\(^\text{28–30}\) An incidence of 4.3 per 100,000 and 0.13 per 100,000, respectively, for *Campylobacter* and *Shigella* was reported in population studies in Finland; the low figure for *Shigella* reflected the finding that this organism was always contracted outside Finland. In other reports of outbreaks of gastroenteritis, widely differing incidences of reactive arthritis have been reported from 0 to 55%.\(^\text{31}\) The reasons for this are not clear but may reflect the particular serotype involved, the dose of bacteria to which subjects were exposed, and the severity of the gastroenteritis. One curious feature is the declining incidence of reactive arthritis due to *Chlamydia* infection set against a large increase in chlamydial infection generally. The explanation is unclear but may relate to the younger age group of those infected and their lack of previous exposure to related organisms.\(^\text{32}\)

The prevalence of undifferentiated spondyloarthritis has been estimated at 0.7% in one study, which identified all those meeting ESSG criteria and then removed those (≈60%) with defined forms of spondyloarthritis.\(^\text{33}\) Because the proportion of those with undifferentiated disease who will eventually proceed to ankylosing spondylitis is approximately 60%, the prevalence of persistent undifferentiated disease is approximately 0.3%; an unknown proportion of these may have undiagnosed reactive arthritis.

### Clinical Features and Diagnosis of Reactive Arthritis

#### History

The arthritis is usually acute in onset and is normally oligoarticular, with a predilection for the lower, weight-bearing,

---

**Table 76-6  Proposed Definition of the Diagnosis of Reactive Arthritis**

<table>
<thead>
<tr>
<th>Patients can be confidently diagnosed with reactive arthritis if they have:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>Classic clinical features:</strong></td>
</tr>
<tr>
<td>Asymmetric oligoarthritis, predominantly lower limbs</td>
</tr>
<tr>
<td>Extra-articular signs</td>
</tr>
<tr>
<td>and</td>
</tr>
<tr>
<td>Proven infection by <em>Salmonella</em>, <em>Campylobacter</em>, <em>Yersinia</em>, <em>Shigella</em>, or <em>Chlamydia</em> (whether symptomatic or not) or</td>
</tr>
<tr>
<td>Proven infection by other organisms previously reported to be associated with reactive arthritis (e.g., <em>Clostridium difficile</em>, <em>Mycobacterium bovis</em> bacillus Calmette-Guérin)</td>
</tr>
<tr>
<td>2. <strong>Any acute inflammatory arthritis, including monoarthritis and/or axial inflammation</strong></td>
</tr>
<tr>
<td>and</td>
</tr>
<tr>
<td>Proven infection by reactive arthritis–associated bacteria</td>
</tr>
<tr>
<td>3. <strong>Classic clinical features (as listed in No. 1)</strong></td>
</tr>
<tr>
<td>and</td>
</tr>
<tr>
<td>Diarrhea or urethritis/cervicitis within the previous 6 wk, infection not proven</td>
</tr>
</tbody>
</table>
limbs. Mild cases picked up in surveys of populations following outbreaks of infection often have polyarthritis or polyarthralgia, but those who present clinically usually have marked synovitis with large effusions, such that septic or crystal-induced arthritis enters the differential diagnosis. A definitive diagnosis of reactive arthritis is greatly aided by two additional aspects of history: extra-articular features and preceding infection. Conjunctivitis is often transient and painless and therefore is discovered only by careful history—sometimes the red eye is noticed only by relatives. Inflammatory back symptoms or enthesitis (heel or plantar fascia pain) may be regarded as minor in comparison with the presenting joint, and a history of these features has to be specifically sought. Specific questioning is needed to elicit a history of preceding infection—the patient will not necessarily spontaneously assume a link between infection and subsequent arthritis, particularly if there is a gap of several weeks between events, although the onset of joint symptoms is rarely more than 4 weeks after infection. Gastrointestinal upset may be minor in some cases (especially with Yersinia infection), and a sexual history will rarely be volunteered spontaneously.

**Signs**

- **Joints**: Careful examination for mild involvement of joints other than the presenting ones is important, particularly in apparent monoarthritis; discovery of even mild synovitis of a proximal interphalangeal (PIP) or wrist joint, in addition to a grossly swollen knee, greatly decreases the likelihood of septic arthritis.

- **Entheses**: Major entheses should be examined for signs of inflammation (Figure 76-1); loss of weight bearing due to ankle or knee arthritis may mean that the patient is unaware of significant Achilles tendinitis or plantar fasciitis. The Maastricht Ankylosing Spondylitis Enthesitis Score (MASES) provides a helpful list of 13 major entheses: first and fourth costochondral joints, L + R; anterior and posterior iliac spines, L + R; iliac crest, L + R; fifth lumbar spinous process; and Achilles tendon insertion, L + R, which should be checked. Missing from this list but readily examined is the calcaneal plantar fascia insertion. Alternatively, the Mander index may be used. Tendinitis may combine with arthritis to produce characteristic dactylitis in fingers or toes (Figure 76-2).

- **Extra-articular disease**: The characteristic skin rash associated with reactive arthritis, keratoderma blenorragica (Figure 76-3), should be sought on soles and palms—again, the former may not be evident to the patient. Histologically, the rash is identical to psoriasis, but skin biopsy is rarely required to establish the nature of the rash. The other typical associated rash is circinate balanitis on the glans penis (Figure 76-4); this also needs to be specifically sought, particularly in uncircumcised males. Erythema nodosum is seen in Yersinia infection but rarely in other reactive arthritis–associated infections. Mouth or palatal ulcers are often painless and may not be reported by the patient. Conjunctivitis sometimes is still present when the patient is first seen by a rheumatologist, but ongoing pain or visual disturbance raises the possibility of uveitis and requires full ophthalmologic assessment.
there is a high pretest probability of reactive arthritis (i.e., cells in the synovial fluid and tissue. Nevertheless, when Chlamydia for example, a patient with active RA and an incidental 
ination, memory T cells are recruited to inflamed joints so that, common to various forms of inflammatory arthritis. In addi 
to ever achieve high specificity because production of pro 
research tests can be used diagnostically. They are unlikely 
arthritis synovial fluid.

Recently, increased numbers of CD4 
forsurface phenotype and intracellular cytokine staining.
to examine cells in synovial fluid using flow cytometry 
cytokine production has been informative. It is also possible 
together with LPS. Hence, diagnosis of reactive arthritis would 
not yet been developed sufficiently. PCR is also helpful in 
establishing a diagnosis of Chlamydia pneumoniae infection 
when arthritis is seen in association with respiratory tract 
infection. The presence of chlamydiae can be detected 
in urine or sputum by means of an enzyme-linked immuno 
sorbent assay (ELISA) to detect chlamydial lipopolysac 
charide (LPS).

Microbiology

Culture and Other Means of Detecting Bacteria. The aim is to gather evidence to implic 
te a reactive arthritis–associated organism. For gastrointestin 
al infection, stool culture should be carried out; organisms such as Salmonella 
and Yersinia may persist in the gut for several weeks after gastroenteritis has resolved. Yersinia has the unusual prop 
erty of growing at 4°C, which allows improved detection 
of the organism by maintaining the stool sample at 4°C for 
24 to 48 hours prior to culture; microbiology laboratories 
should be informed that Yersinia infection is being consid 
ered to allow them to do this. Stool culture for Yersinia is al 
so appropriate in patients without gastrointestinal 
symptoms because symptoms are often mild in infection 
with this organism. For genitourinary infection, swabs 
should be obtained from the urethra or vagina for culture of Chlamydia trachomatis and, if indicated, from the throat 
and rectum. However, in view of the difficulty involved in 
culturing chlamydiae, nucleic acid amplification techniques 
(ligase chain reaction [LCR], PCR) are generally preferred 
and can be performed on urine samples or by self 
administered swabs, both of which are acceptable to patients. Use of PCR to identify Chlamydia in synovial fluid has 
not yet been developed sufficiently. PCR is also helpful in 
serology. Although culture or direct demonstration of 
the organism is preferred, in some circumstances serology 
can provide reliable evidence of a preceding infection. Its 
effectiveness depends very much on the frequency of infec 
tion in the community and therefore on the proportion of 
subjects with pre-existing antibodies. For Salmonella species, 
and increasingly for Campylobacter, antibodies are com 
monly detected in healthy subjects because of a high level 
of community exposure to these organisms. In these cases 
therefore, diagnosis will rest, at best, on an increased titer of pathogen-specific antibodies in acute and convalescent 
sera. Because patients may present to rheumatologists in the 
convalescent phase, as far as the triggering infection is con 
cerned, this substantially decreases the utility of serology in 
diagnosing these infections. In contrast, infection with Yer 
sinia or Shigella is not prevalent in Western societies, so that 
detection of specific antibodies carries greater significance 
even if immunoglobulin (Ig)M antibodies or a fourfold 
increase in titer cannot be demonstrated. For detection of chlamydial infection, serology generally has not proved reli 
able. Micro-immunofluorescence tests can be carried out, in 
which dilutions of sera are applied to multi-well slides 
containing organisms; this is regarded as the “gold standard” 
for Chlamydia serology. However, again the prevalence of antibodies in the population is high, particularly for
Individual recombinant chlamydial antigens have been used to devise additional serologic tests, but none has yet been established for general use.

**HLA-B27 Testing**

Testing for HLA-B27 generally is not required to make the diagnosis of reactive arthritis. Although more severe cases such as those seen in secondary care are strongly associated with HLA-B27 (70% to 80% of cases are positive), the percentage of positivity is much less in milder cases and is highly variable in different published series. However, HLA-B27 testing is worthwhile because it may identify patients with an increased likelihood of persistent disease, in whom early introduction of disease-modifying antirheumatic drugs (DMARDs) would be indicated (see later).

**Imaging**

Imaging techniques that can identify inflammation of affected joints or entheses are what is required; the choice lies between ultrasound (particularly power Doppler studies), scintigraphy, and MRI. As noted previously, MRI may reveal sacroiliitis or other evidence of axial involvement but is not required for joints with clinically obvious synovitis—the norm in reactive arthritis. Power Doppler ultrasonography appears promising for detecting enthesitis but is not without difficulties in interpretation.\(^{43-45}\) This technique may also reveal enthesitis in rheumatoid arthritis, thereby producing difficulties for the ASAS classification criteria, but this enthesitis is usually noted in relation to nearby synovitis rather than as the predominant inflammatory lesion. Likewise, scintigraphy did not perform particularly well in detecting sacroiliitis in a population with chronic back pain;\(^6\) it might perform better in young patients with sacroiliitis due to reactive arthritis, because the background prevalence of abnormalities in healthy young controls is likely to be much lower.

**TREATMENT OF REACTIVE ARTHRITIS**

**Control of Symptoms**

Most patients (80% to 90%) will have self-limiting disease, which does not require disease-modifying drugs. Symptomatic measures include full-dose nonsteroidal anti-inflammatory drugs (NSAIDs), preferably long-acting, with the choice between selective and nonselective drugs dependent on the patient’s risk of gastrointestinal bleeding versus cardiovascular disease; in young patients, the former predominate, so there is a preference for agents such as etoricoxib.\(^{47}\) Additional analgesics may be required. Intra-articular corticosteroids are very useful when infection has been confidently excluded, and injection of enthesitis may also be required with avoidance of the Achilles tendon insertion, in view of the tendency of this tendon to rupture. In patients with severe disease and systemic features, parenteral or short-course oral steroids may be used. Patients require physiotherapy to maintain muscle bulk and range of movement. Informal psychologic support is often crucial because, although the disease is self-limiting, it commonly takes 6 to 12 months to remit. In a young, previously fit patient, this is perceived as a major medical event, curtail many sporting and leisure activities, and sometimes interfering with employment.

**Disease-Modifying Drugs**

Few controlled trials have examined the use of DMARDs in reactive arthritis. Such patients were included in a trial of sulfasalazine in spondyloarthritis,\(^{48}\) and sulfasalazine was also tested in reactive arthritis.\(^{49}\) The natural tendency of spontaneous remission makes it difficult for investigators to conduct trials to determine the effectiveness of DMARDs, but sulfasalazine showed effectiveness in 62% of reactive arthritis patients as compared with 48% in placebo. In undifferentiated spondyloarthritis, sulfasalazine was reported not to be effective, although an unexpected effect on axial symptoms, but not peripheral arthritis, was recorded.\(^{50}\) This result is at variance with other findings.\(^{51,52}\) When should DMARDs be introduced in reactive arthritis? Patients who have severe, persistent disease or recurrent disease, particularly those who are HLA-B27+, are candidates for the early introduction of DMARDs. However, even patients with very severe presentation can settle rapidly with NSAIDs and intra-articular steroids, so only those who are responding slowly or whose disease flares significantly after an initial response would justify introduction of DMARDs within the first 3 months of disease. Those who relapse after having responded well initially are also likely to require DMARDs. The risk-benefit ratio of sulfasalazine is favorable, and the drug is a logical choice in cases triggered by enteric infection. Patients who fail sulfasalazine can progress to methotrexate or leflunomide, or a combination of the two. Note however, that none of these recommendations on the use of DMARDs has been properly tested in controlled trials. In view of the favorable course in most patients, there is reluctance to test early use of drugs such as methotrexate in the way that is now commonplace in rheumatoid arthritis (RA).

**Biologics**

Only in rare cases have biologics such as tumor necrosis factor (TNF)-blocking agents or tocilizumab been used in reactive arthritis, although they appear to be effective. They have not been reported to be associated with recrudescence of infection—an important consideration in view of evidence for persistence of triggering organisms such as *Chlamydia* or *Yersinia*. In light of reports of increased numbers of T helper (Th)17 cells in reactive arthritis joints and genetic evidence implicating the IL-17 axis in spondyloarthritis, rational future use of biologics might suggest the use of inhibitors of IL-23—ustekinumab, which recognizes the shared IL-12/23 p40 subunit and is effective in psoriasis, or a IL-23—specific antibody targeting the p19 unit, currently in trial, or IL-17 itself. IL-6 and IL-1 are also implicated in generating Th17 cells, and the efficacy of tocilizumab (in a single case report) may reflect an effect on the generation of Th17 cells.
**Antibiotics**

By definition, the joint in reactive arthritis is sterile, so there is no prima facie case for treatment of arthritis with antibiotics. Obviously, chlamydial infection in the genitourinary tract requires conventional treatment to avoid damage and scarring, particularly in females; *Chlamydia pneumoniae* respiratory tract infection also needs treatment. In general, food poisoning does not require antibiotics, although ciprofloxacin is often used to reduce symptom duration.63

In relation to arthritis, the use of prolonged courses of antibiotics has been advocated in view of increasing evidence of persistence of the organism, within the joint and elsewhere, despite adequate antibiotic treatment of the primary infection; this applies particularly to chlamydia infection. Several placebo-controlled trials have examined the effects of prolonged courses of antibiotics (3 to 6 months), which are effective with reactive arthritis–associated bacteria (tetracycline, ciprofloxacin, and azithromycin) and with outcome in reactive arthritis.64-68 For disease due to enteric infection, the results are uniformly negative. In an early trial, post hoc subgroup analysis suggested a beneficial effect of prolonged lymecycline in *Chlamydia*-induced disease64; an effect of prolonged antibiotics on *Chlamydia*-induced arthritis was not confirmed in subsequent trials,66,69 although some studies may not have had sufficient *Chlamydia*-induced cases. A recent and controversial study used a combination of antibiotics, rifampicin with doxycycline, and azithromycin, which were chosen to be effective against persistent, slowly dividing organisms; investigators reported effectiveness in *Chlamydia*-induced reactive arthritis, including in patients with long-standing disease (mean duration, 10 to 14 years).70,71 Seventeen of 27 antibiotic-treated patients achieved the trial end point compared with 3 of 15 receiving placebo, and treatment was associated with clearance of *chlamydiae* as detected by PCR. The trial involved relatively small numbers, and the findings require confirmation.

**OUTCOME IN REACTIVE ARTHRITIS**

A vast majority of patients with reactive arthritis make a full recovery with no chronic joint damage, although in more severe cases, 12 to 18 months may be required for complete resolution of symptoms. Remaining patients have progressive spondyloarthritis with relapses and involvement of new joints and entheses in the absence of any evidence of re-infection, and often increasing prominence of axial symptoms and radiographic changes of ankylosing spondylitis.72 HLA-B27 is the main factor predisposing to chronicity or recurrence in reactive arthritis and is worth testing for this reason. It is possible that polymorphisms in other genes,73 including those that influence susceptibility to AS,74 will also affect outcome in reactive arthritis, but this remains to be tested. Of course it may be argued that HLA-B27* patients are predisposed to both reactive arthritis and other forms of spondyloarthritis, and therefore might develop two diseases rather than reactive arthritis evolving into ankylosing spondylitis, but the clinical impression suggests evolution rather than development of a new disease.

Severe arthritis (large effusions, multiple joints, and very high levels of CRP and ESR) does not necessarily correlate with persistent disease, but complete recovery will usually take longer (6 to 12 months) compared with milder disease. Young active patients require counseling to prepare them for this.

**DIAGNOSIS AND TREATMENT OF UNDIFFERENTIATED SPONDYLOARTHRITIS**

Because, as discussed previously, undifferentiated spondyloarthritis by definition is not a “specific” disease, a detailed account of its diagnosis and treatment is somewhat redundant.

**Diagnosis**

Briefly, patients present rheumatologically with arthritis or enthesitis, or dactylitis, which is itself a combination of these conditions. In such cases, the diagnosis will be based on additional features of spondyloarthritis, as detailed in the ASAS classification criteria for axial or peripheral arthritis. An algorithm for diagnosing axial disease on this basis has been published.75 Every attempt should be made to rescue the patient from the undifferentiated diagnosis to a more distinct form of spondyloarthritis, but this will not always be possible, especially in those with early disease and in a percentage of those with long-standing disease.

**Investigations**

These have already been detailed in relation to reactive arthritis. In cases with predominantly axial disease, inflammatory markers may not be elevated.

**Treatment**

Patients with undifferentiated spondyloarthritis will have different combinations of axial and peripheral disease. Where axial disease predominates, management will most resemble that for ankylosing spondylitis. Patients usually will require full-dose NSAIDs (a good response to these reinforces the diagnosis), axial symptoms are generally refractory to conventional DMARDs, and a proportion of patients will require treatment with biologics, particularly anti-TNF, when symptom control with NSAIDs proves inadequate.76,77 An important research question is whether early intervention with anti-TNF will prevent progression of undifferentiated spondyloarthritis to ankylosing spondylitis, that is, whether the “window of opportunity” concept leading to early aggressive treatment of rheumatoid arthritis might also apply to ankylosing spondylitis. Peripheral disease is managed in a similar fashion to psoriatic arthritis, or reactive arthritis in its more chronic, established form, with the use of DMARDs—sulfasalazine, methotrexate, and leflunomide, given individually or in combination. In the absence of psoriasis or a family history of psoriasis, hydroxychloroquine can be used, often as an add-on to another more powerful DMARD. Finally, the antibiotic regimen recently reported to be effective in *Chlamydia*-induced
reactive arthritis was first tested in undifferentiated arthritis on the assumption that it might be related to *Chlamydia*; findings showed some efficacy.  

**UNANSWERED QUESTIONS AND FUTURE RESEARCH**

1. **How Many Patients Currently Classified as Undifferentiated Spondyloarthritis Actually Have Reactive Arthritis?**

Answering this question requires better diagnostic techniques. For *Chlamydia*-induced reactive arthritis, detection of bacteria-specific mRNA has been reported, even in late disease. However, the problems of using PCR with high levels of amplification while avoiding contamination have not been fully resolved. Given that chlamydiae might traffic to any inflamed joint, and therefore may be identified by PCR in, for example, RA joints, it may be that a specific transcript or set of transcripts will be associated with bacterial persistence in the joint, which drives reactive arthritis. Equally, a synovial tissue or fluid gene expression “signature” characteristic of reactive arthritis may exist (better still, a peripheral blood signature) and may be used for diagnosis. This would be particularly useful in reactive arthritis triggered by enteric organisms because bacteria-specific mRNA in the joints of these patients has only occasionally been reported.

2. **Do Antibiotics Have Any Role in the Treatment of Established Reactive Arthritis?**

This question has been re-opened, at least for *Chlamydia*-induced arthritis, by recent data reporting successful treatment with combinations of azithromycin or tetracycline and rifampicin. These observations need to be repeated and expanded. A positive result would lead to exploration of the same combinations in other forms of reactive arthritis.

3. **What Is the Role of HLA-B27 in Inducing Susceptibility to Reactive Arthritis, and What Other Genes Influence Susceptibility?**

These questions are generic for our understanding of the pathogenesis of spondyloarthritis. They have rightly been addressed initially in ankylosing spondylitis, where the incidence of HLA-B27 is very high and case definitions can be well controlled. Mechanisms for the mode of action of HLA-B27, which emerge from these studies, can be re-examined in reactive arthritis, where an interaction with the consequences of intracellular infection might be expected. Although this was initially seen in terms of HLA-B27 presenting “spondylitogenic” peptides from reactive arthritis–associated bacteria, other mechanisms involving homodimeric B27 heavy chains or induction of the unfolded protein response may present opportunities for interactions with bacteria. As other genes associated with ankylosing spondylitis are identified, they need to be tested in reactive arthritis. There is a need therefore to establish cohorts of reactive arthritis patients in which the diagnosis is certain so that accurate genetic work can be done.

**References**

47. van der Heijde D, Baraf HSB, Ramos-Remus C, et al: Evaluation of
43. Kiris A, Kaya A, Ozgocmen S, Kocakoc E: Assessment of enthesitis in
41. Shen H, Goodall JC, Gaston JS: Frequency and phenotype of T helper
37. Shen H, Goodall JC, Gaston JS: Frequency and phenotype of T helper 17 cells in peripheral blood and synovial fluid of patients with reactive arthritis, J Rheumatol 37:2096–2109, 2010.


The references for this chapter can also be found on www.expertconsult.com.
Psoriatic arthritis should be suspected in a patient with an asymmetric joint distribution pattern who may have additional clinical features, such as dactylitis, enthesitis, or inflammatory-type back pain, and who is negative for rheumatoid factor.

New classification criteria, the Classification of Psoriatic Arthritis (CASPAR) criteria, have been validated.

Psoriatic arthritis is a progressive disease, with 47% of patients developing erosions within 2 years of diagnosis. Polyarticular disease and an elevated erythrocyte sedimentation rate are markers of poor outcome.

An essential core set of domains and instruments is now agreed as being necessary for inclusion in clinical trials.

Studies of synovial tissue have highlighted an increase in vascularity and the presence of neutrophils as helping to distinguish spondyloarthropathy from rheumatoid arthritis. Change in synovial CD3\(^+\) T cell infiltration might correlate with clinical response to treatment.

Prominent enthesal involvement with bone marrow edema at enthesal insertions on magnetic resonance imaging has prompted the hypothesis that psoriatic arthritis may originate at the enthesis.

A role for CD8\(^+\) T cells and the innate immune response has been proposed. Although there is a paucity of evidence for efficacy of disease-modifying antirheumatic drugs in psoriatic arthritis, tumor necrosis factor inhibitors have proved effective for skin and joint disease.

Psoriatic arthritis is a member of the spondyloarthropathy family and may be defined as an inflammatory arthropathy associated with psoriasis and usually negative for rheumatoid factor. Until the 1950s, an inflammatory arthritis occurring in the presence of psoriasis was thought to represent rheumatoid arthritis (RA) occurring coincidentally with psoriasis. Based primarily on clinical and radiologic grounds and using the rheumatoid factor, the distinction between RA and psoriatic arthritis became gradually accepted. Wright described the classic clinical features in 1959, and together with his colleague Moll, he published his classification criteria in 1973. These criteria have remained until recently the simplest and the most frequently used in clinical studies. The American Association of Rheumatism included psoriatic arthritis as a distinct clinical entity in the classification of rheumatic diseases for the first time in 1964.

Epidemiologic studies have supported the concept that psoriatic arthritis is a unique disease entity separate from RA. The prevalence of inflammatory arthritis is increased among patients with psoriasis, ranging from 7% to 25% compared with a general population estimate of 2% to 3%. The prevalence of psoriasis among subjects with arthritis also is increased at 2.6% to 7% compared with a general population estimate of 0.1% to 2.8%.

Psoriasis affects about 2% of the population. The prevalence varies, with 5% to 10% of Russians and Norwegians affected, and only 0% to 0.3% of West Africans or Native Americans affected. Onset of psoriasis may occur at any age, but it most frequently peaks in the 20s. Although no gender predilection has been reported, a genetic predisposition has been noted.

Among patients with psoriasis, 7% to 42% develop arthritis. This figure varies so widely in part because of a lack of widely accepted diagnostic criteria; it also varies according to which population is being studied. The exact prevalence and incidence of psoriatic arthritis are unknown. The reported prevalence of psoriatic arthritis ranged from 0.056% to 0.28% in a large population-based study in the United States. Cases were defined as patients who reported a “physician diagnosis” of psoriasis and psoriatic arthritis. Prevalence was calculated at 0.25% (95% confidence interval, 0.18% to 0.31%). Kay and colleagues\(^6\) did a prevalence study in northeast England to evaluate records from six general practices; 81 of 772 psoriasis subjects had an inflammatory arthritis with a prevalence of 0.28%. The reported incidence of psoriatic arthritis has varied from 3 to 23 cases per 100,000. Data from Rochester, New York, have shown an incidence rate of 6.59 per 100,000, whereas in Finland, 16 new cases of psoriatic arthritis were identified in a population of 87,000, yielding a mean incidence rate of 23 per 100,000.\(^8,9\) Using the Icelandic genealogy database, risk ratios (RRs) for developing psoriatic arthritis spanning five generations were estimated. Results confirmed a strong and complex genetic component with a significant risk ratio up to the fourth-degree relatives of psoriatic arthritis patients, as well as an important environmental contribution.

CLINICAL FEATURES

Plaque psoriasis or psoriasis vulgaris is the most common skin phenotype in patients with psoriatic arthritis. Other patterns of skin involvement may be seen (Figure 77-1). Although arthritis usually develops in a setting of an established diagnosis of psoriasis, some patients may be unaware that they have psoriasis, or psoriasis may develop after the onset of arthritis in approximately 15% of cases. In a
clinical study, the incidence of development of arthritis in patients with psoriasis remained constant (74 per 1000 person-years), while the prevalence increased with time since diagnosis of psoriasis, reaching 20.5% after 30 years.\textsuperscript{11} If a patient presents with the classic articular manifestations of psoriatic arthritis but does not volunteer psoriasis or the presence of a rash, it is incumbent on the physician to examine the patient’s skin carefully, including the scalp and nails, because psoriasis frequently lurks in such areas. Examples of nail dystrophic changes are shown in Figure 77-2.

Although a more recent U.S. study suggests that the prevalence of psoriatic arthritis among psoriasis patients increases with psoriasis severity,\textsuperscript{6} in clinical practice, little relationship has been observed between severity of skin

\footnotesize{Figure 77-1  Clinical phenotypes in psoriasis: plaque psoriasis (psoriasis vulgaris). A, At extensor surface of elbow and on scalp (B). Genital psoriasis (C). Infra mammary and umbilical flexural psoriasis (D). Guttate psoriasis in a father and child (E). Erythrodermic psoriasis on the trunk and upper limbs (F).}
Involvement and severity of arthritis. In one prospective study, only 35% of patients reported that their skin and joint components flared at the same time. Other systemic features of joint inflammation are common in patients with psoriatic arthritis, including stiffness after rest and fatigue. In one study of fatigue, a number of factors, including disease activity, physical disability, pain, and psychologic distress, contributed to fatigue, with comorbid fibromyalgia and hypertension further adding to the challenge. Compared with other core outcome measures, fatigue has been found to be an independent outcome measure that is sensitive to change.

Patients with psoriatic arthritis present with symptoms and signs of joint, enthesal, or spinal inflammation. The joints involved at presentation in 129 early psoriatic arthritis patients are shown in Figure 77-3. In one of their seminal papers on psoriatic arthritis, Wright and Moll described five clinical patterns of psoriatic arthritis (Figure 77-4):

1. Asymmetric oligoarthritis
2. Symmetric polyarthritis
3. Axial involvement
4. Distal interphalangeal joint involvement
5. Psoriatic spondylitis
3. Predominant distal interphalangeal (DIP) joint involvement
4. Predominant spondyloarthritis
5. Destructive (mutilans) arthritis

These classification criteria are the most commonly quoted, although many alternative criteria have been proposed. Variability in the definition of terms has led to differences in the reported frequency of psoriatic arthritis in subsets among the different studies. The pattern of joint involvement is not fixed; the patient’s disease may fluctuate and may be influenced by treatment. In a study of 129 patients with early psoriatic arthritis, 53 of 77 initially classified as polyarticular were reassessed at 2 years; 26 of 53 (49%) patients were subsequently classified as oligoarticular, 19 of 53 (36%) remained classified as polyarticular, and 12 of 53 (23%) were in remission.16

The Classification of Psoriatic Arthritis (CASPAR) study has included in the analysis a breakdown of disease pattern subtypes.17 This multicenter study included data on 588 psoriatic arthritis cases and 536 controls. In contrast to the original Moll and Wright paper, but similar to many subsequent publications, approximately 63% of patients had polyarticular joint involvement compared with 13% with oligoarticular disease. The other patterns of joint involvement described by Moll and Wright occurred much less commonly. Predominant DIP disease was found in less than 5%, but DIP involvement can occur in any of the subtypes. Predominant spondyloarthritis is uncommon, although spinal involvement may be found in 40% to 70% of psoriatic arthritis cases, depending on whether or not radiographs are taken.18 Risk factors for spinal involvement include severe peripheral arthritis and HLA-B27.19 Finally, arthritis mutilans, a destructive form of arthritis associated with flail joints, is rare, although more patients may develop this form of joint involvement with time if their disease is not properly controlled.

Features that are typical of psoriatic arthritis, including dactylitis and enthesitis, are helpful in diagnosis. Dactylitis, which is characterized by a sausage-shaped swelling of the fingers or toes (see Figure 77-4B), may be found in 29% to 33.5% of psoriatic arthritis patients at first presentation, and 48% may have an episode of dactylitis during follow-up.16,20 Ultrasound and magnetic resonance imaging (MRI) studies have shown that joint and tenosynovial inflammation are prominent in involved digits.21,22 Enthesitis, inflammation at tendon and ligament insertion into bone, is a feature of all of the spondyloarthropathies and may be a presenting feature in psoriatic arthritis. Overall, enthesitis is found in 38% of patients at presentation.16 The most common enthesal sites involved are the Achilles and plantar fascia.
Certain articular features if present are useful in distinguishing psoriatic arthritis from RA, including dactylitis, DIP involvement, and inflammation at enthesal sites (Table 77-1; see Figure 77-4). In addition, inflammatory-type back pain or sacroiliitis on plain x-ray or MRI should raise the suspicion of psoriatic arthritis because spinal involvement is uncommon in RA. The absence of rheumatoid nodules

**Table 77-1 Clinical Features That Distinguish Psoriatic Arthritis from Rheumatoid Arthritis**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Psoriatic Arthritis</th>
<th>Rheumatoid Arthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psoriasis</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Symmetric</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Asymmetric</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Enthesopathy</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Dactylitis</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Nail dystrophy</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Human immunodeficiency virus</td>
<td>+</td>
<td>–</td>
</tr>
</tbody>
</table>

**Differential Diagnosis**

With the obvious exception of psoriasis and nail dystrophic change, extra-articular disease is less common in psoriatic arthritis than in RA. Iritis or uveitis occurs in 7% to 18%, more bilateral than in ankylosing spondylitis, but is usually found in patients with spinal involvement. Numerous studies have suggested that psoriatic arthritis patients have a higher prevalence of inflammatory bowel disease, sometimes asymptomatic and detected only on biopsy specimen. Whether this inflammatory bowel disease is coincidental or is possibly related to medication effects remains to be clarified. Distal limb edema or lymphedema may occur more commonly in psoriatic arthritis; one case-control study found it in 21% of psoriatic arthritis patients compared with 4.9% of controls (see Figure 77-4A). Finally, amyloid is rare but is described in psoriatic arthritis.

*insertions. Other sites include the insertions of the quadriceps and patellar tendons, the iliac crest, the rotator cuff, and the epicondyles at the elbow. Patients have reported pain at these sites, with tenderness and sometimes swelling noted on examination. Enthesal involvement may be asymptomatic, and ultrasound is more sensitive than clinical palpation. Often spurs are detected on x-ray, although spurs are not always associated with symptoms.*
or other systemic features common to RA can be another useful differentiating feature.

Distinguishing psoriatic arthritis from other spondyloarthopathies is also important. Dactylitis may be a feature in reactive arthritis, where a palmoplantar pustular rash (keratoderma blennorrhagicum) may be clinically and histologically indistinguishable from pustular psoriasis (see Figure 77-1G). In relation to spinal involvement, sacroilitis may be unilateral more frequently, and the spinal changes on plain radiography may be more asymmetric in psoriatic arthritis than with classic ankylosing spondylitis. Finally, crystal-associated arthropathies occasionally can confuse, especially with monoarticular disease, and are best distinguished by synovial fluid crystal analysis. Serum urate levels may be increased in patients with psoriatic arthritis, adding to the confusion. To aid dermatologists and general physicians in identifying those with psoriatic arthritis, a number of screening tools for arthritis in patients with psoriasis have been proposed and are the subject of a current head-to-head comparative study.

LABORATORY FEATURES

No diagnostic laboratory test for psoriatic arthritis is known. Although the absence of rheumatoid factor is considered an important distinguishing feature from RA, low levels of rheumatoid factor may be found in patients (5% to 16%) with typical psoriatic arthritis features. Until a more definitive diagnostic test becomes available, it is difficult to be categorical about diagnosis in these patients. Cyclic citrullinated peptide antibodies initially were thought to be specific to RA, but it is now recognized that cyclic citrullinated peptide antibodies are found in approximately 5% of psoriatic arthritis patients as well.27 Acute phase markers, such as erythrocyte sedimentation rate, C-reactive protein, and serum amyloid A, may be elevated in psoriatic arthritis patients, but less commonly and to a lesser degree than in RA patients. These markers are elevated in particular in patients with polyarticular disease and act as a marker of poor prognosis.28 Finally, as was mentioned previously, hyperuricemia may be found in association with metabolic abnormalities in psoriatic arthritis patients and may not reflect the extent of skin involvement.

RADIOGRAPHIC FEATURES

Although substantial advances have been made in the application, in particular, of musculoskeletal ultrasound (MSUS) and of MRI in patients with arthritis, including psoriatic arthritis, plain radiographic imaging remains the “gold standard” for assessing bony changes in peripheral joints in psoriatic arthritis.

Plain Radiography

Sixty-seven percent of patients with established psoriatic arthritis have radiographic abnormalities,11 and 47% of patients with recent-onset psoriatic arthritis will have developed erosions within 2 years of disease onset.16 Distinctive radiographic features reflect in some cases the clinical phenotype (Figure 77-5). These features include asymmetric joint involvement; involvement of the interphalangeal joints of the fingers and toes, with features of bony erosion and resorption sometimes seen together and resulting in the classic “pencil-in-cup” deformity; joint space narrowing or involvement of entheseal sites, often with bony spurs or periostitis developing; and spinal involvement, frequently less severe and asymmetric than in classic ankylosing spondylitis.

Radiographic progression in psoriatic arthritis is slow in early stages of the disorder, with the mean modified Sharp (to include DIP joints in the hands) erosion score at presentation increasing from 1.2 to 3 at 2 years.16 Larson and Sharp scoring systems have been used in psoriatic arthritis, but neither the Larson nor the Sharp score has been developed specifically for psoriatic arthritis or has been extensively validated.

Musculoskeletal Ultrasound

Many MSUS applications are useful in psoriatic arthritis; these applications are likely to develop further as the technology (in particular, power Doppler) that can be used to allow identification of blood flow is further advanced (Figure 77-6). Already it has been shown that MSUS is more sensitive than clinical examination in detecting knee synovitis in patients with various arthritides, including psoriatic arthritis.29 One study has suggested further that this increased sensitivity may result in reclassification of some patients as polyarticular when they were previously diagnosed as oligoarticular on clinical grounds.30 This reclassification may result in significant changes in prognosis and therapy. Finally, MSUS has been used in objective monitoring of the response of synovitis to therapy.31 MSUS features at the enthesis include enthesal thickening, hypoechoic change, increased vascularity as shown on power Doppler, tenosynovitis, and bony erosions or enthesophyte formation.32,33 MSUS has been shown to be more sensitive than clinical examination in detecting lower limb enthesopathy.33,34 MSUS has been used in studies of dactylitic digits. Together with MRI, MSUS has shown dactylitis to be due to a combination of synovial and tenosynovial inflammation.35 Finally, MSUS guidance for small joint or enthesal aspiration or injection may have particular application in patients with psoriatic arthritis.

Magnetic Resonance Imaging

MRI studies have been particularly useful in offering new insights into disease pathogenesis in psoriatic arthritis. Based on the prominent entheseal-related bone marrow edema seen on MRI, McGonagle and colleagues36 have proposed that psoriatic arthritis, in contrast to RA, is an enthesal-based disease. MRI can be used to study all aspects of joint involvement, including the enthesis, but use of MRI as a routine clinical tool in psoriatic arthritis has not been clarified. Application of MRI to the spine or sacroiliac joints in psoriatic arthritis may prove especially helpful, as has been shown in ankylosing spondylitis, but studies in psoriatic arthritis patients are awaited. Preliminary studies have suggested that MRI can be useful as an outcome measure in the detection of synovitis or of vascularity in patients with psoriatic arthritis undergoing biologic therapies (Figure 77-7). More detailed studies are required. A new MRI
Figure 77-5  Radiologic features in psoriatic arthritis. A, Third left distal interphalangeal joint monoarthritis with prominent new bone formation. B, Bone scan from same patient as in A. C, Asymmetric right-sided sacroiliitis. D, Severe destructive changes (arthritis mutilans) with multiple erosions and "pencil-in-cup" deformities. (Courtesy Dr. Robin Gibney.)

Figure 77-6  Musculoskeletal ultrasound features in psoriatic arthritis. A, Right plantar fascia thickening compared with the left. B, Transverse section through left third finger at the metacarpophalangeal joint showing right tenosynovitis. C, Power Doppler ultrasound through left third finger at the metacarpophalangeal joint confirming increased vascularity (synovitis). D, Right Achilles tendinitis with calcaneal erosion. (Courtesy Dr. Robin Gibney.)
scoring system has been proposed to measure articular inflammation and damage in patients with psoriatic arthritis (PsAMRIS).  

Other Imaging Modalities

The use of other imaging modalities, such as computed tomography (CT) or scintigraphy, has largely been superseded by MRI. CT is now reserved mainly for patients for whom MRI is contraindicated, or for whom MRI is unavailable. Positron emission tomography has been found to be comparable with MSUS and MRI in RA knees; this work needs to be extended to psoriatic arthritis. In a study using high-resolution CT imaging, structural bone changes were compared between RA and psoriatic arthritis. Smaller Ω-shaped and tubule-shaped erosions, as well as large sometimes corona-shaped osteophytes, are typical of psoriatic arthritis.  

DIAGNOSIS

A diagnostic test for psoriatic arthritis is currently unavailable. Nevertheless, in its simplest form, psoriatic arthritis can be considered as an arthritis occurring in the presence of psoriasis, but in the absence of rheumatoid factor. Most psoriatic arthritis patients meet this simple definition. The arthritis can be predominantly spinal, it may involve only enthesal sites, psoriasis may present after arthritis in 15% of cases, and low-titer positive rheumatoid factor may be found. Recognizing these difficulties, the CASPAR group published new classification criteria based on an analysis of 588 psoriatic arthritis cases and 536 controls (Table 77-2). These criteria have yet to be validated in other large patient cohorts, and they should not be used in individual patient diagnosis. In the setting of clinical research, the CASPAR criteria have a specificity of 0.987 and a sensitivity of 0.914. For the individual patient, an algorithm for diagnosis is suggested in Figure 77-8.

CLINICAL COURSE AND OUTCOME

Five early psoriatic arthritis cohorts have been studied. The mean disease duration in these cohorts was 6 to 12 months, the median age at onset of psoriasis was 27 to 31 years, and the median age at onset of arthritis was 38 to 52 years. Overall, little relationship was noted between skin disease severity and onset of psoriatic arthritis; the small joints of hands and feet were most commonly involved; the DIP joints were involved in one-third of patients, usually in association with nail disease, which was present in two-thirds; dactylitis and enthesitis were present in one-third; and spinal involvement only was found in 2% to 4%, but was present in 20% overall. At follow-up, disease had continued to progress in most patients, with 47% developing
Psoriatic Arthritis

<table>
<thead>
<tr>
<th>OUTCOME DOMAINS AND INSTRUMENTS</th>
</tr>
</thead>
</table>

Measuring response to treatment of psoriatic arthritis in clinical trials has been the subject of much interest for members of the Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) and Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT). Many of the data that have been used to date in clinical trials have been adapted from RA and have not been validated. Controversial issues have included the number of joints to count, the usefulness of the acute phase response in psoriatic arthritis, and how important is it to include a measure of function or quality of life. An essential core set of domains that must be included in clinical trials has now been agreed on, with other domains necessary but not mandatory, and yet others requiring considerably more research (Figure 77-9). Instruments for many of these domains have yet to be developed and validated, and some instruments, such as the Psoriasis Assessment Severity Index (PASI), have acknowledged limitations. Table 77-3 lists the currently available instruments for the core domains. Instruments for dactylitis and enthesitis have been proposed and validated.

In the setting of clinical trials, numerous composite scores (e.g., American College of Rheumatology [ACR]20, ACR50, ACR70; EULAR Disease Activity Score [DAS] response criteria) have been used in psoriatic arthritis, most again adapted from RA and not extensively validated in psoriatic arthritis. One scoring system was developed for psoriatic arthritis—the PsARC; although it has been used in numerous studies, it too has not been extensively validated and is considered perhaps insufficiently responsive and discriminant. Much work is required to develop a validated and responsive composite instrument for psoriatic arthritis. More recently, minimal disease activity criteria were developed and validated, and two disease activity indices have been proposed. The Disease Activity Index for Psoriatic Arthritis (DAPSA) appears to work well for...
**PATHOGENESIS**

Many studies have explored key components of disease pathogenesis, including the contribution of genetic factors, the role of infection or trauma, studies of animal models or involved sites of disease, and the importance of components of the immune system such as cytokines.

### Genetic Factors

Familial clustering of psoriasis and psoriatic arthritis is well described. Twin studies in psoriasis have shown a high rate of concordance in monozygotic twins. The genetic basis for this clustering has been the subject of extensive investigations in psoriasis, but has been much less well studied in psoriatic arthritis. Studies of psoriatic arthritis have often included patients as a subset of larger psoriasis cohorts, and the diversity of clinical phenotypes has not often been recognized.

A strong association between psoriasis and the human leukocyte (HLA)-C region of the major histocompatibility complex (MHC) has long been recognized. Whether this was HLA-Cw6 itself, found in approximately 60% of psoriasis cohorts, or a region telomeric to this has been the subject of much controversy. Elder definitively showed that the HLA susceptibility region for psoriasis is HLA-Cw6, which is often in linkage disequilibrium with other HLA-B alleles such as HLA-B57, HLA-B37, and HLA-B13. The presence of HLA-Cw6 is associated with an earlier age of onset of psoriasis (type 1 disease, <40 years old) and with more extensive and severe disease. In individuals with psoriatic arthritis, the association with HLA-Cw6 is slightly weaker, whereas additional associations have been found with HLA-B27, chiefly in patients with predominant spinal disease, and with HLA-B38 and HLA-B39.

These findings have been interpreted to suggest that the MHC association with psoriasis lies close to the HLA-C region, whereas the association with the articular manifestations more likely lies in or close to the HLA-B region. A study of a large cohort of psoriatic arthritis patients in the United Kingdom has found HLA-Cw6 to be in linkage disequilibrium with HLA-DRB1*07, and that possession of both alleles was associated with fewer involved or damaged joints (Pauline Ho, personal communication, 2007).

Other genes within the MHC region have been explored in psoriasis and psoriatic arthritis. Tumor necrosis factor (TNF) promoter polymorphisms or a gene in linkage disequilibrium with TNF may predispose the patient to or increase susceptibility to psoriasis and to psoriatic arthritis. One study has found further an association between the TNF-308 A allele and disease progression in early psoriatic arthritis.

Whole-genome scans in psoriasis have identified additional non-MHC susceptibility regions, known as the PSORS regions on chromosomes 4, 6, and 17. To date, no candidate genes have been identified.

Increasing evidence suggests that an additional or distinct genetic contribution is responsible for the development of psoriatic arthritis. Investigators have pointed to an MHC class I chain-related A (MICA)-A9 polymorphism, which confers additional relative risk, in particular for polyarticular disease, in psoriasis patients who carry Cw*0602. MICA-A9 polymorphism was found in linkage

**Table 77-3** Core Set Instruments Proposed for Use in Clinical Trials

<table>
<thead>
<tr>
<th>Domain</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peripheral joint inflammation</td>
<td>Tender/swollen joint count 68/66</td>
</tr>
<tr>
<td>Patient global assessment</td>
<td>Instrument under study proposed*</td>
</tr>
<tr>
<td>Skin assessment</td>
<td>Psoriasis Assessment Severity Index (if body surface area ≥3%)</td>
</tr>
<tr>
<td></td>
<td>Lesion score (erythema, induration, scale)</td>
</tr>
<tr>
<td>Physical function</td>
<td>HAQ/SF-36 physical function composite</td>
</tr>
<tr>
<td>Health-related quality of life</td>
<td>Generic</td>
</tr>
<tr>
<td></td>
<td>Disease specific (e.g., DLQI, PsAQoL)</td>
</tr>
</tbody>
</table>


DLQI, Dermatology Life Quality Index; HAQ, health assessment questionnaire; PsAQoL, Psoriatic Arthritis Quality of Life; SF-36, Short Form health survey.
dis-equilibrium with HLA-B alleles (B*5701, B*3801). These results suggest that the MICA gene or other nearby genes may be involved in the development of psoriatic arthritis. Additionally, a genome scan identified a paternally influenced locus on chromosome 16—a region not known to be implicated in psoriasis susceptibility. 62

Finally, two genome-wide association studies in psoriatic arthritis have been reported. 62,63 These studies have confirmed associations with HLA-C and interleukin (IL)-12B, and they have identified a new susceptibility region TRAF3IP2 on chromosome 6p, which encodes a protein involved in IL-17 signaling, and which interacts with members of the Rel/ nuclear factor κB (NFκB) transcription factor family.

Environmental Factors

The role of environmental factors in triggering skin or joint disease in patients with psoriasis or psoriatic arthritis has been supported largely by clinical observations, although the mechanism is poorly understood. It has long been recognized that there is a strong association between guttate psoriasis and preceding streptococcal infections in children. 64 That this association might be related to a streptococcal superantigen has been proposed. Some authors have found bacterial antigens in synovial tissue samples from psoriatic arthritis patients, but this may be different from those in noninflammatory control subjects. 65

The Koebner phenomenon (see Figure 77-1H) has been reported to occur in 52% of patients with psoriasis. The Koebner phenomenon is the development of psoriasis along the site of skin trauma. It has been proposed that trauma may play a role in triggering episodes of joint inflammation, and the term deep Koebner phenomenon has been coined. Although the role of trauma has not been proved, in one study, 24.6% of patients reported a traumatic event before the onset of arthritis. 66

Finally, a link between stress and exacerbation of psoriasis has been proposed, supported largely by clinical observational studies. A similar association may exist in psoriatic arthritis, but this has not been systematically examined.

Animal Models

Although spondyloarthropathy has been detected in a variety of primates, rodent models have proved helpful in deciphering pathogenic pathways. In rodents transgenic for HLA-B27 class I molecules, skin, nail, and joint features have been described that mimic some of the features of the human phenotype. 67 When HLA-B27 transgenic rats were raised in a germ-free environment, they seemed to be protected from joint disease. Mice genetically lacking MHC class II have developed skin and joint disease, but confined to the distal phalanges with skin and nail disease also on the affected digits. 68 Involvement of the distal phalanges and nails was reported in aging male DBA/1 mice from different litters that were caged together from 12 weeks. 69 In these animals, dactylitis, periostitis, and ankylosing enthesitis were observed.

Finally, JunB protein was shown to be expressed in normal and in clinically uninvolved psoriatic skin, but expression was considerably reduced in involved psoriatic lesions. 70 Epidermal deletion of JunB and c-Jun in a mouse model resulted in skin and joint disease with 100% penetrance and a clinical and histologic phenotype highly consistent with human psoriasis and psoriatic arthritis. In further experiments, the same authors showed that the joint disease, but not the skin disease, required T and B cells and intact TNF receptor 1 signaling.

Immunopathology

The key pathologic events in psoriatic arthritis occur in the skin, synovium, enthesal sites, and cartilage and bone. Pathobiologic features in the skin and synovium have been well described, but only a few studies have focused on the enthesis. In relation to cartilage and bone, studies have shown the presence of osteoclasts at the cartilage-pannus junction and high numbers of circulating osteoclast precursors in the circulation of psoriatic arthritis patients. Detailed studies similar to those done on RA on the synovial-cartilage-bone interface possibly could yield valuable information regarding joint destruction in psoriatic arthritis.

Psoriasis Skin

Involved psoriasis skin is characterized by epidermal hyperplasia, mononuclear leukocytes in the papillary dermis, neutrophils in the stratum corneum, and an increase in various subsets of dendritic cells. 71 CD8+ T cells are the predominant T cell subset chiefly found in the epidermis, whereas dermal T cells contain a mixture of CD4+ and CD8+. Most T cells in skin lesions express addressin, a cutaneous lymphocyte antigen, in contrast to circulating T cells and T cells found in the inflamed synovium in psoriatic arthritis. 72 Finally, vascular changes are prominent in psoriasis with impressive growth and dilation of superficial blood vessels.

Psoriatic Synovium

Many early studies of synovial pathology in psoriatic arthritis highlighted the presence of prominent and striking vascular changes. In the first study that compared psoriatic arthritis and RA synovial tissue, quantitative immunopathologic analysis confirmed these prominent vascular changes and found that vessel number was significantly increased in psoriatic arthritis. 73 Lining layer hyperplasia was less marked in psoriatic arthritis, and fewer macrophages were seen trafficking into the synovium and out to the lining layer. The numbers of T lymphocytes and their subsets and the number of B cells were similar to the frequency found in RA. Although neutrophil infiltration was not assessed, this study examined adhesion molecule expression further in the two patient subgroups and found E-selectin expression to be considerably reduced in psoriatic arthritis. Many of these observations have been confirmed by other authors.

In a study by Kruithof and co-workers, 74 the synovial immunopathologic features in patients with spondyloarthropathy, including psoriatic arthritis, were compared with the features seen in RA. 74 Using a semi-quantitative scoring system, the authors identified many features characteristic of the spondyloarthropathy group as a whole and in the
psoriatic arthritis subgroup alone. Increased vascularity, higher neutrophil numbers (also seen in involved psoriasis skin), and a higher number of infiltrating CD163+ macrophages, a marker of mature tissue macrophages, reliably distinguished spondyloarthritis from RA. No significant differences were seen between oligoarticular versus polyarticular psoriatic arthritis.

The important role of the vasculature in psoriatic arthritis pathogenesis is perhaps most elegantly shown by the large numbers of tortuous and dilated blood vessels observed through an arthroscopic view of psoriatic joints. An interaction of key growth factors is thought to regulate closely the new vessel formation or angiogenic process. Growth factors, including TNF, transforming growth factor (TGF)-β, platelet-derived growth factor (PDGF), angiopoietins (ANG-1, ANG-2), and vascular endothelial growth factor (VEGF), have been described in skin and synovial tissue. Because this expression is found at an early stage of inflammation, it may represent a primary event in psoriatic arthritis as opposed to a reaction or response to hypoxia. One possibility is that there is a genetic predisposition to endothelial activation, which results in new vessel formation and increased cellular trafficking.

Of interest, two studies have identified a change in synovial CD3+ T cell infiltration as correlating with clinical response in patients commencing biologic therapies (Pontifex, personal communication). Changes in CD3 infiltration also correlated with changes in MRI assessment of synovitis (Pontifex). Changes in CD3 infiltration may well prove to be a useful tissue biomarker of treatment response.

**Enthesal Sites**

Laloux and associates described the immunopathologic features of the enthesis in patients undergoing joint replacement surgery with spondyloarthritis, including psoriatic arthritis, and compared them with RA. The number of patients was small in this study, but a consistent increase in CD8+ T cell expression was observed at the enthesis in patients with psoriatic arthritis compared with RA patients. Ultrasound-guided biopsy of five sites of acute enthesis in early spondyloarthritis confirmed an inflammatory response with increased vascularity and cellular, predominantly macrophage, infiltration. These findings are consistent with the well-described association of psoriatic arthritis with HLA class I antigens. They also are consistent with the previously described dominance of activated and mature CD8+ T cells in psoriatic arthritis synovial fluid samples compared with RA. It is attractive to suggest that entheseally derived antigens might trigger an immune response in adjacent synovial tissue. To date, evidence for this hypothesis has not been found, although this is clearly an area for future study. A search for candidate antigens common to the enthesis and the skin might be informative.

**Cytokines**

Synovial explant tissues obtained from psoriatic arthritis joints have been shown to produce higher levels of the T helper type 1 (Th1) cytokines interleukin (IL)-2 and interferon-γ protein than explants similarly cultured from osteoarthritis and RA patients. This Th1 lymphocyte profile also has been observed in psoriasis plaques. The cytokines IL-1β and TNF were released by psoriasis synovial explants in high concentrations. In contrast, IL-4 and IL-5 were not identified, but IL-10 was highly expressed in psoriatic synovium, although not in skin. A similar pattern of cytokine production in psoriatic arthritis synovium was shown using immunohistochemical and gene expression techniques.

Other innate cytokines, such as IL-18 and IL-15, are present in psoriatic arthritis synovial tissue and are downregulated by methotrexate therapy.

The role of Th17 cells has been explored, and an important immunomodulatory role has been better established in psoriasis than in psoriatic arthritis. Comparison of lesional skin gene expression in responders versus nonresponders to etanercept revealed rapid downmodulation of innate IL-1β and IL-8 sepsis cascade cytokines in both groups, but only responders downregulated IL-17 pathway genes to baseline levels. An increase in circulating Th17 cells has been found in psoriatic arthritis and in an increase in IL-17 has been found in skin, synovial tissue, and synovial fluid of psoriatic arthritis patients. Further supporting a key role for the Th17 pathway, genetic polymorphisms in key cytokines involved in differentiation of T cells to Th1 and Th17 subtypes, IL-12 and IL-23, have been found to be associated with susceptibility to both psoriasis and psoriatic arthritis, and inhibition of the common p40 subunit of IL-12/IL-23 has led to clinical improvement in both skin and joint manifestations.

TNF levels are elevated in psoriatic skin, synovium, and joint fluid of patients with psoriatic arthritis. Several lines of evidence support the concept that TNF is an important cytokine in the psoriatic joint. TNF transgenic mice exhibit extensive bone destruction similar to that observed in some psoriatic arthritis patients. In a study of 129 patients with early psoriatic arthritis, patients with erosions were significantly more likely to have the TNF-308 A allele, an allele associated with high TNF production. As was mentioned earlier, immunohistochemical and gene expression studies have shown marked upregulation of TNF in the psoriatic synovial membrane. Histopathologic analysis of synovial specimens from eight spondyloarthropathy patients, four of whom had psoriatic arthritis, treated with the anti-TNF monoclonal antibody infliximab revealed decreased vascularity, synovial lining thickness, and mononuclear cell infiltration after therapy. In another study, a significant reduction in the quantity of infiltrating macrophages, the CD31+ vascular area, αββ3-positive neovessels/UEX eagletail–positive vessels, VEGF and its receptor KDR/flk-1 (VEGFR-2), and SDF-1α–positive vessels in psoriatic arthritis synovium was noted after 8 weeks (three infusions) of infliximab treatment.

**Matrix Metalloproteinases and Cartilage Destruction**

Radiographs of psoriatic joints often reveal cartilage loss manifested as joint space narrowing. Similar to RA, matrix metalloproteinases (MMPs) and tissue inhibitors of MMPs (TIMPs) were identified in psoriatic arthritis synovial lining and sublining layers. In particular,
immunohistochemical studies revealed that MMP-9 localized to blood vessel walls, whereas MMP-1, MMP-2, MMP-3, TIMP-1, and TIMP-2 showed a cellular and interstitial staining pattern in the synovial lining. Serum levels of MMP-3 exhibited a marked and rapid decrease after successful anti-TNF therapy, raising the possibility that this molecule may serve as a biomarker. In another study, similar levels of MMP-1 and MMP-3 mRNA were detected in RA and psoriatic arthritis synovial tissue despite the fact that RA patients exhibited more erosions on plain radiographs. The elevated ratio of MMPs to TIMP-1 in synovial tissue favored cartilage degradation, although expression of MMPs was not significantly elevated at the cartilage-pannus junction compared with other sites. These reports indicate that MMPs are upregulated in psoriatic arthritis synovium, but their precise functions remain to be defined.

Bone Remodeling

Radiographs of psoriatic arthritis joints can reveal markedly altered bone remodeling in the form of bone resorption (tuft resorption or osteolysis, large eccentric erosions, and pencil-in-cup deformities) and new bone formation (periostitis, spur or enthesophyte formation, bony ankylosis). Important in bone resorption, psoriatic joint biopsy specimens show large multinucleated osteoclasts in deep resorption pits at the bone-pannus junction. Osteoclastogenesis (differentiation of monocytes into osteoclasts) is a contact-dependent process directed by osteoblasts and stromal cells in the bone marrow. These cells release signals necessary for differentiation of an osteoclast precursor derived from the CD14+ monocyte population into an osteoclast.

One of these signals is the receptor activator of nuclear factor κB ligand (RANKL), a member of the TNF superfamily that binds to receptor activator of nuclear factor κB (RANK) on the surface of osteoclast precursors and osteoclasts. This ligand-receptor interaction stimulates proliferation and differentiation of osteoclast precursors and activation of osteoclasts. It has been proposed that the relative expression of RANKL and of its natural antagonist osteoprotegerin (OPG) ultimately controls osteoclastogenesis. In psoriatic arthritis synovial tissues, marked upregulation of RANKL protein and low expression of osteoprotegerin were detected in the adjacent synovial lining. Osteoclasts also were noted in cutting cones traversing the subchondral bone, supporting a bidirectional attack on the bone in psoriatic joints. In addition, osteoclast precursors, derived from circulating CD14+ monocytes, were markedly elevated in the peripheral blood of patients with psoriatic arthritis compared with healthy controls. Treatment of patients with psoriatic arthritis with anti-TNF agents significantly decreased the level of circulating osteoclast precursor, supporting a central role for TNF in the generation of this precursor population. Although abundantly expressed, there appeared to be little relationship between expression of RANKL, OPG, and RANK and both systemic and local inflammation, and also little relationship between the decline in osteoclast precursor numbers and bone marrow edema following etanercept treatment.

The mechanisms responsible for new bone formation in the psoriatic joint are poorly understood. Higher circulating levels of bone formation markers, in particular bone alkaline phosphatase, in psoriatic arthritis patients as compared with RA patients both before and after anti-TNF therapy commenced has been demonstrated (Szentpetery, personal communication). TGF-β and VEGF may be pivotal in this process of new bone formation, given that TGF-β is strongly expressed in synovial tissues isolated from patients with ankylosing spondylitis and synergizes with VEGF to induce bone formation in animal models. De Klerck and colleagues showed that the bone morphogenetic proteins (BMPs) BMP-2 and BMP-7 are upregulated in regions of pathologic new bone formation. The same investigators showed that expression of phosphorylated Smad-1 and Smad-5, important signaling molecules downstream of BMP, was markedly increased in regions of new bone formation taken from the calcaneus in a patient with Achilles tendinitis and periostitis. These studies provide evidence that potential mediators of ankylosis and periostitis in the psoriatic enthesis and joint include BMP molecules and possibly VEGF and TGF-β. Finally, the Wnt pathway and the balance between Wnt and the Wnt antagonist dickkopf-1 (DKK-1), which binds to the Wnt receptor complex on the surfaces of osteoblast lineage cells, may well be of considerable importance in the disordered bone remodeling found in psoriatic arthritis.

Summary

In considering a model for disease pathogenesis in psoriatic arthritis, we have to try to take into account genetic susceptibility; the role of the environment; cellular immunologic mechanisms; and secreted cytokines, chemokines, and other proteins. For some time, the primary hypothesis has been that psoriatic arthritis is an HLA class I–restricted, antigen-driven immune process (Figure 77-10). Considerable evidence has been presented to support this hypothesis; however, despite careful analysis of T cell receptor phenotype, no antigen-driven process other than that driven by Epstein-Barr virus has been identified. The potential role of components of the innate immune response, such as Toll-like receptors or cells bearing natural killer receptors, is currently under active investigation. It is possible that the interaction of environmental factors, such as those derived from pathogens or expressed after trauma, with Toll-like receptors in a genetically susceptible individual may set in train intracellular signaling events leading to cytokine release, immune activation, and release of destructive enzymes such as MMPs (Figure 77-11).

TREATMENT

In considering treatment strategies for psoriatic arthritis, the diverse nature of the clinical phenotype (peripheral arthritis, skin and nail disease, axial disease, dactylitis, and enthesitis) may complicate therapeutic decisions because not all treatments are effective for all features, and patients often display a mixture of all of the features simultaneously. GRAPPA published a systematic review of the evidence for treatment strategies in psoriatic arthritis; this review provided treatment guidelines that are currently being revised. The reader is referred to this GRAPPA publication for a detailed review of the evidence; Figure 77-12 is a proposed preliminary algorithm for treatment choices. Treatment
Although there is little published evidence of a favorable therapeutic effect in psoriatic arthritis, nonsteroidal anti-inflammatory drugs are most often the agents first used in psoriatic arthritis, whatever the clinical phenotype (level 1b, grade A). Expert opinion supports the use of

**Traditional Agents**

Although there is little published evidence of a favorable therapeutic effect in psoriatic arthritis, nonsteroidal anti-inflammatory drugs are most often the agents first used in psoriatic arthritis, whatever the clinical phenotype (level 1b, grade A). Expert opinion supports the use of...
nonsteroidal anti-inflammatory drugs, although occasional exacerbations of psoriasis have been reported. The use of systemic corticosteroids is not evidence based (level 4, grade D), although 24% of patients in one study were taking prednisolone. Concerns have arisen suggesting that exacerbations of psoriasis may follow corticosteroid withdrawal. No randomized controlled trials have examined intra-articular steroids in psoriatic arthritis or the use of local enthesal or dactylitis injections. Expert opinion indicates that intra-articular steroids can be quite effective, especially in oligoarticular disease, or where localized enthesal involvement is present, as in plantar fasciitis (level 4, grade D). Mild skin disease (PASI 10) is usually controlled with topical steroids or vitamin D derivatives; the latter are best used for maintenance therapy.

Systemic therapy is considered in patients with three or more inflamed joints despite conventional therapy, as described previously; persistent or treatment-resistant axial, enthesal, or dactylitic disease, especially where multiple sites are involved; or moderate or severe psoriasis (PASI >10). Randomized controlled trials of disease-modifying antirheumatic drugs (DMARDs) are few and limited by size. Based on evidence and expert opinion, nearly all DMARDs may have small to moderate beneficial effects on peripheral joints, enthesitis, and dactylitis. Axial features and nail disease do not seem to respond. Good or moderate improvements in skin disease have been reported with some of the older systemic agents, such as methotrexate, cyclosporine, sulfasalazine, leflunomide, and acitretin (all level 1b, grade A).

The best evidence for DMARD use comes from studies of peripheral joint disease and of psoriasis. Six randomized controlled trials have studied the use of sulfasalazine in psoriatic arthritis (level 1a, grade A); the largest included 221 patients. Fifty-nine percent of patients achieved a therapeutic response (PsARC), but consistent with other studies, a high therapeutic response (42.7%) was also noted in placebo-treated patients.

Methotrexate remains for many rheumatologists the DMARD of first choice for patients with psoriatic arthritis, but evidence for its use is limited (level 3, grade B). A small, prospective randomized controlled trial concluded that methotrexate was as effective as cyclosporine; another study of 72 patients with active psoriatic arthritis and an incomplete response to methotrexate, in which cyclosporine was added in, showed significant differences only in synovitis as detected by MSUS and PASI score in favor of the combination therapy. Although evidence for methotrexate is lacking, an open study reported significant reductions in synovial cell infiltration and in cytokine gene expression after 3 months of therapy.

Perhaps the best randomized controlled trial in psoriatic arthritis of a DMARD examined the use of leflunomide (level 1b, grade A). This trial included 190 patients who received leflunomide or placebo for 24 weeks. Fifty-nine percent of patients treated with leflunomide compared with 30% of patients given placebo met the primary response criteria (PsARC), with significant, although small, improvements in other individual parameters, including joint scores, health assessment questionnaire results, PASI, and Dermatology Life Quality Index scores. Regarding some older DMARDs, such as gold salts and antimalarials, no evidence of treatment benefit has been found; exacerbation of psoriasis has been reported, and these agents cannot be recommended. One small, randomized controlled trial with azathioprine suggested benefit with a reduction in Ritchie score (level 2b, grade B). Finally, apart from the cyclosporine/methotrexate study referred to earlier, little or no evidence suggests that DMARD combination therapy is beneficial or safe in psoriatic arthritis.

With the exception of psoriasis, there is a paucity of evidence that DMARDs are beneficial for the other features of psoriatic arthritis, including dactylitis, axial disease, or...
enthesis. Absence of evidence does not mean absence of an effect, however. Further randomized controlled trials specifically examining these features in psoriatic arthritis are required. In psoriasis, methotrexate and cyclosporine have been shown to be highly and probably equally effective (level 1b, grade A). Adverse effects, in particular with cyclosporine, may limit usage in some patients.

**Biologics**

The approach to treatment in psoriatic arthritis has changed considerably with the introduction of biologic therapies. As a result of numerous large, well-conducted, randomized controlled trials, accumulating evidence indicates that anti-TNF treatments are effective in controlling peripheral arthritis symptoms and signs, improving quality of life, and preventing radiologic progression (overall level 1b, grade A). Indeed higher rates of Disease Activity Score (DAS) 28 remission are found in psoriatic arthritis patients as compared with RA patients after 1 year of anti-TNF therapy. Patients receiving 25 mg subcutaneously twice weekly of etanercept showed significant improvement in American College of Rheumatology (ACR) 20 responses (39% vs. 15%) at 12 weeks compared with placebo. At 12 months, radiographic disease progression (modified total Sharp score) was inhibited in the etanercept group (−0.03 unit) compared with worsening of +1.00 unit in the placebo group.

Although the anti-TNF therapies have not been compared in any study, their effect on peripheral arthritis seems to be similar. Evidence indicates that anti-TNF therapies are also effective for other disease features, such as nail disease, enthesitis, and dactylitis. These studies are limited by the absence of a validated instrument to measure these features. With psoriasis, the effects of anti-TNF therapies can be quite dramatic. In particular with antibody therapy, highly significant improvements in PASI scores were achieved (e.g., PASI = 75 in 59% of adalimumab-treated patients vs. 1% in placebo-treated patients after 24 weeks; PASI = 90 in 50%). Better skin but not articular responses have been demonstrated in etanercept patients treated with 50 mg twice weekly as compared with once weekly.

Alefacept is a fusion protein of soluble lymphocyte function antigen 3 with Fc fragments of immunoglobulin (Ig) G1. Alefacept was the first biologic agent approved for the treatment of moderate to severe psoriasis (level 1b, grade A). Efficacy was dose dependent and slow, but PASI = 75 was achieved in 33% of patients at some point. Alefacept in combination with methotrexate was evaluated in 185 patients with active psoriatic arthritis. At 6 months, 54% of alefacept-treated patients versus 23% of placebo-treated patients achieved an ACR20 response. Finally, efalizumab, a humanized monoclonal antibody targeting the CD11a component of lymphocyte function antigen 1, has been approved for the treatment of psoriasis, with PASI = 75 achieved in 22% to 39% in randomized controlled trials, but this agent has since been withdrawn owing to safety issues. To date, no evidence for a beneficial effect of efalizumab in other psoriatic arthritis joint manifestations has been observed. Finally, inhibition of the common p40 subunit of IL-12/IL-23 has demonstrated good efficacy in psoriasis, and ustekinumab has now been approved for treatment. A phase II clinical trial in psoriatic arthritis suggests reasonable efficacy for joint disease, but additional studies are required.

**References**


The references for this chapter can also be found on www.expertconsult.com.
This chapter deals with rheumatic conditions associated with gut pathology, which are commonly designated as enteropathic arthritis. People realized centuries ago that dysentery was sometimes followed by arthritis, a condition now known as reactive arthritis. Although it is a well-established example of enteropathic arthritis, reactive arthritis is covered in Chapter 76. This chapter covers musculoskeletal problems associated with inflammatory bowel disease (IBD), celiac disease (CeD), and other less common conditions. New insight into gut physiology, its barrier function, regulation of immune responses, and trophic functions is emerging. Knowledge regarding detailed interactions implicated in the generation of joint disease, however, remains incomplete. The fine-tuning of permeability, interactions between gut contents and gut mucosa, the nature of the gastrointestinal-associated lymphoid tissue, and how gastrointestinal-associated lymphoid tissue generates arthritis are at the core of this chapter.

**GUT BIOLOGY AND THE MICROBIOTA**

**KEY POINTS**

- The gut mucosa is a leaky barrier and is covered by an “unstirred” layer.
- Eighty percent of the gut wall consists of immune cells.
- Gut permeability is delicately regulated and is disturbed in disease.

The gastrointestinal tract has three major biologic functions: serving as a barrier against hostile environmental factors; affecting fluid and food absorption and excretion of waste products; and producing major trophic host functions. The gut mucosa has an estimated surface area of 300 to 400 m², which is 200 times the body’s skin surface area. This mucosa is constantly exposed to potentially harmful environmental agents against which defense is essential, but at the same time tolerance toward a normal microflora must be maintained. In order to fulfill these diverse functions, the gut is an immunologically privileged organ. Disturbances can lead to food allergy or inflammatory bowel disease (Figure 78-1).

The gut epithelium constitutes a so-called leaky barrier. It consists of intestinal epithelial cells including Paneth cells in deep crypts, mucus-secreting goblet cells, M cells, and lymphocytes. It is covered by a thin layer of fluid separated from luminal flow and peristalsis of the intestine, the so-called unstirred layer, which slows down the diffusion of solutes and prevents loss of digestive enzymes. An increased thickness of the unstirred layer may contribute to malabsorption in CeD. Underneath the basal membrane is the lamina propria, which contains immune cells including lymphocytes, dendritic cells, and macrophages (Figure 78-2).

The plasma membrane of the epithelial cells is impermeable to most hydrophilic solutes in the absence of specific transporters. The structure of the interepithelial paracellular space consists of the tight junction, the adherens junction, and the desmosome (see Figure 78-2B). The tight junction is considered rate limiting for permeability. A perijunctional actomyosin ring condensation is regulated by myosin light chain kinase (MLCK), which has been shown to have a central role in regulating tight junction transport and in tumor necrosis factor (TNF)-induced permeability increase. TNF inhibition will influence permeability via inhibition of MLCK transcription. Overexpression of MLCK, on the other hand, results in increased permeability and activation of immune cells in transgenic mice.

Although initially healthy, these mice were prone to accelerated colitis if challenged.

Na⁺ solutes smaller than 5000 daltons can pass through the epithelium freely, whereas bacterial products and dietary antigens are dependent on active transport mechanisms. Altered gut permeability can be observed in several diseases, among them nonalcoholic fatty liver disease. In this condition alterations in the zona occludens 1 (ZO1) expression result in perturbation of permeability. Genetic and exogenous factors such as drugs, nicotine, microorganisms, and cytokines are potential triggers of pathology by influencing paracellular functions in the gut epithelium.
Assessment of mucosal permeability and transport can be achieved by oral feeding of lactalbumin, lactoglobulin, polyethylene glycol particles, $^{13}$Cr-labeled ethylenediamine tetraacetic acid (EDTA), and sugars such as lactulose and mannitol, followed by urinalysis. In addition, intestinal permeability and function can be studied by regional perfusion with the help of endoscopic techniques that close off segments of the gut with inflatable balloons. A study applying enzyme-linked immunosorbent assays to fluid collected by this technique showed marked local immunity to a number of food-related antigens in patients with rheumatoid arthritis. Ethnic differences in gut permeability have also been described.

The healthy gut harbors a mixture of native bacteria acquired at birth or shortly thereafter that retains a relatively constant composition; it also has a smaller population of transient bacteria of varying composition. The former are essential for health and live in symbiosis; the latter contain potential pathogens. Whereas the stomach and duodenum normally contain less than $10^3$ mucosa-adhering bacteria, the number of bacteria increases to $10^8$ in the jejunum and $10^9$ in the ileum. Most of the latter group comprises gram-negative aerobic species. In the colon, the bacterial density is $10^{10}$ or more, consisting mostly of anaerobic bacteria. Transit time is fast in the upper gut and slow in the distal gut, but the immunologic impact of the microflora is higher in the proximal parts of the gut. In other words, the total number of bacterial cells called the human microbiota is 10 times that of cells in the body.

Analysis of the 16S bacterial ribosomal RNA gene sequence has revealed the presence of a bewildering number of phylotypes in the human gastrointestinal tract, and the number of genes in the microbiome is estimated to be 100 times that of the human genome. There is also a high degree of diversity among healthy individuals and some correlation with obesity and other conditions. Most phylotypes have not been cultured, and their functions remain unknown. The potential pathogenicity of individual commensal microbes is demonstrated in a recent paper showing that a single species of segmental filamentous bacteria can trigger autoimmune arthritis in germ-free K/BxN mice by activating Th17 cells in lamina propria.

The normal trophic functions of the gut require this microflora, as demonstrated by host defects in germ-free animals. Bacteria digest food carbohydrates into short-chain fatty acids, which facilitate the absorption of Ca$^{2+}$, Mg$^{2+}$, and Fe$^{2+}$ ions; synthesize amino acids and vitamins; and secrete antibacterial protective substances. Some 300 to 500 different species are represented, and the composition is unique for each individual. Figure 78-3 shows some of the common colonic species and their functions.

**Gastrointestinal-Associated Lymphoid Tissue and Its Interactions**

**KEY POINTS**

- Microbiota are the major regulators of gastrointestinal-associated lymphoid tissue.
- Retinoic acid promotes IgA and lymphocyte homing.
- Flagellin stimulates dendritic cells to induce Th17 cells.
- Vascular adhesion protein-1 is expressed in gut and synovium and is a putative therapeutic target.
- Gut lymphocytes express $\alpha$4$\beta$7, $\alpha$E$\beta$7, and CCR9, which are important for homing.

Gastrointestinal-associated lymphoid tissue (GALT) is a part of the mucosal immune system, a prominent feature of which is the dominating output of secretory immunoglobulin (Ig)A. This was discovered analyzing tears and saliva in the early 1960s. GALT is the largest lymphoid organ of the body, constituting 25% of the mucosal mass. Cellular components of GALT are localized in Peyer’s plaques, gut lymphoid follicles, lamina propria, and intraepithelial T cells. GALT is part of a complex defense and tolerance regulating system involving specialized dendritic cells (DCs) and intestinal epithelial cells (IECs) (Figure 78-4). Commensal bacteria interact with IECs, which send signals to DCs mediated by anti-inflammatory substances (e.g., thymic stromal lymphopoietin [TSLP]) and to lamina propria DCs. Bacteria also interact directly with DCs by means of their polysaccharide A. A subset of lamina propria DCs expresses CD11b$^+$, CD11c$^+$, and Toll-like receptor (TLR5). When exposed to bacterial flagellin, these DCs promote the differentiation of T helper 17 (Th17) cells. These DCs also produce retinoic acid. IEPs also produce retinoic acid from dietary vitamin A, APRIL (a proliferation-inducing}

---

**Figure 78-1** Immune privilege in the gut. CD, Crohn’s disease; IFN, interferon; IL, interleukin; NK-T, natural killer T cell; TNF, tumor necrosis factor; UC, ulcerative colitis. (From Iweala OI, Nagler CR: Immune privilege in the gut: the establishment and maintenance of non-responsiveness to dietary antigens and commensal flora, Immunol Rev 213:82–100, 2006.)

**Figure 78-2** Regulatory adaptive immune cells (CD4$^+$ T cell) and phagocytic innate immune cells (macrophages, neutrophils, epithelial cells) that interact with IECs. (From S. O’Hare and A. O’Garra: Immune regulation by commensal bacteria, Nat Rev Immunol 10:376–387, 2010.)

**Figure 78-3** Tolerance to microbiome and dysregulation of mucosal homeostasis. Th2 and Th1 cells mediate the TH1 and TH2 responses with IL-4, IL-13, and IFN-$\gamma$. Food allergy and inflammatory bowel disease (IBD) are associated with gut homeostasis dysregulation. Gut lymphocytes express $\alpha$4$\beta$7, $\alpha$E$\beta$7, and CCR9, which are important for homing.

**Figure 78-4** Commensal bacteria interact with IECs, which send signals to DCs mediated by anti-inflammatory substances (e.g., thymic stromal lymphopoietin [TSLP]) and to lamina propria DCs. Bacteria also interact directly with DCs by means of their polysaccharide A. A subset of lamina propria DCs expresses CD11b$^+$, CD11c$^+$, and Toll-like receptor (TLR5). When exposed to bacterial flagellin, these DCs promote the differentiation of T helper 17 (Th17) cells. These DCs also produce retinoic acid. IEPs also produce retinoic acid from dietary vitamin A, APRIL (a proliferation-inducing...
Plasma cells in the lamina propria become specifically, which are exported into the lumen and constitute a with IgM to form secretory IgA, SIgA, and SIgM, respecti

The pIgR in IECs combines with pIgA and to a lesser extent polymeric immunoglobulin, pIgA, and IgM. SC/pIgR is produced in IECs. It is a 100-kD transmembrane receptor for receptor that is also called... skews to formation of dimeric IgA, pIgA, in which the monomers are connected by joint, or J, chains.

Cytokines such as interleukin (IL)-10 and transforming growth factor (TGF)-β serve anti-inflammatory functions. Retinoic acid, APRIL, and BAFF contribute to the dominating IgA production in local lymphoid cells. Gut enterochromaffin cells produce 5-hydroxytryptamine, which has anti-inflammatory effects and chromogranins with both proinflammatory and anti-inflammatory effects. Some of these not yet fully understood interactions are depicted in Figure 78-5. Plasma cells in the lamina propria become programmed to produce IgA, but the production is also skewed to formation of dimeric IgA, plgA, in which the monomers are connected by joint, or J, chains.

Intestinal epithelial glycophorin plgR, a polymeric immunoglobulin receptor that is also called secretory component (SC), is produced in IECs. It is a 100-kD transmembrane receptor for polymeric immunoglobulin, plgA, and IgM. SC/plgR is abundantly present in Peyer’s patches in the distal ileum. The plgR in IECs combines with plgA and to a lesser extent with IgM to form secretory IgA, SlgA, and SlgM, respectively, which are exported into the lumen and constitute a noninflammatory, non–complement-binding first line of defense. It is estimated that a healthy adult secretes 3 to 5 g of SlgA into the gut daily (Figures 78-6 and 78-7). Breastfeeding provides the newborn with abundant SlgA and SlgM, which confers passive protection and also regulates much of the child’s immune system (Figure 78-8).

From the Peyer’s patches, primed B lymphocytes disseminate throughout the body’s mucous membranes, notably to other parts of the alimentary tract. Primed T lymphocytes also disseminate into the circulation and lymph nodes and home into target organs such as salivary glands (in Sjögren’s disease), lungs, and synovium. Vascular adhesion protein-1 (VAP-1) expressed on synovial epithelial cells is involved in lymphocyte homing, and P-selectin is a part of macrophage recruitment. VAP-1 is a bifunctional glycoprotein with both adhesive and amino-oxidative properties. Inhibition of this molecule is in development in oncology. A monoclonal antibody is protective against collagen-induced arthritis, and VAP-1 may become a target in the treatment of enteropathic arthritis. Most T lymphocytes in the mucosal lamina propria are CD4+, whereas intraepithelial T cells are mostly CD8+. Gut-associated lymphocytes preferentially express the integrins α4β7 and αEβ7 and the integrin receptor CCR9 on stimulation by intestinal dendritic cells.

Induction of oral tolerance to type II collagen has demonstrated how GALT activation may ameliorate...
A chain of events in the pathogenesis of enteropathic arthritis can begin with gastrointestinal infection with the appropriate microorganism in a genetically predisposed patient. This causes local inflammation in the gut mucosa, formation of secretory IgA, increased permeability, absorption of foreign material, and triggering of T lymphocytes. Circulating immune complexes and memory T cells localize to joints and cause synovitis (Figure 78-9). Figure 78-10 summarizes recent understanding of how colonization by segmented filamentous bacteria triggers arthritis in mice.

In conclusion, the gut wall is a highly diversified immunologic, endocrine, and digestive organ. Its interaction with the microbiota potentially mediates both protective and harmful signals (e.g., triggering enteropathic arthritis).

**KEY POINTS**

The incidence and prevalence of inflammatory bowel disease are higher in developed societies.

Smoking increases the risk in both CD and ulcerative colitis.

Smokeless nicotine (snuff) may be inert or protective.

Ethnic influences are present.

**Epidemiology**

The incidence and prevalence of inflammatory bowel disease (see Figure 78-7).21 A chain of events in the pathogenesis of enteropathic arthritis can begin with gastrointestinal infection with the appropriate microorganism in a genetically predisposed patient. This causes local inflammation in the gut mucosa, formation of secretory IgA, increased permeability, absorption of foreign material, and triggering of T lymphocytes. Circulating immune complexes and memory T cells localize to joints and cause synovitis (Figure 78-9). Figure 78-10 summarizes recent understanding of how colonization by segmented filamentous bacteria triggers arthritis in mice.

In conclusion, the gut wall is a highly diversified immunologic, endocrine, and digestive organ. Its interaction with the microbiota potentially mediates both protective and harmful signals (e.g., triggering enteropathic arthritis).

**INFLAMMATORY BOWEL DISEASE**

**Epidemiology**

**KEY POINTS**

The incidence and prevalence of inflammatory bowel disease are higher in developed societies.

Smoking increases the risk in both CD and ulcerative colitis.

Smokeless nicotine (snuff) may be inert or protective.

Ethnic influences are present.
Figure 78-5 Putative role of chromogranins (Cgs) in immune activation and inflammation. Luminal or internal inflammatory stimuli causes alteration in Cgs or Cgs-derived peptides release. They may act locally on Paneth, goblet, and epithelial cells, as well as on immune cells such as macrophages, dendritic cells, neutrophils, mastocytes, and eosinophils. Endothelial permeability, chemotaxis, rolling, smooth muscle contractility, and the enteric nervous system can also be modulated. IL-16, interleukin-16; (−) inhibition; (+) activation. (From Khan WI, Ghia JE: Gut hormones: emerging role in immune activation and inflammation, Clin Exp Immunol 161:19–27, 2010.)

Figure 78-6 Receptor-mediated export of dimeric IgA and pentameric IgM to provide secretory antibodies (SloG and SlgM) functioning in immune exclusion of antigen (Ag) at the mucosal surface. Polymeric Ig receptor (pIgR) is expressed basolaterally as membrane secretory component (mSC) on secretory epithelial cells and mediates transcytosis of dimeric IgA and pentameric IgM, which are produced with incorporated J chain (IgA + J and IgM + J) by mucosal plasma cells. Although J chain is often produced by mucosal IgG plasma cells (70% to 90%), it does not combine with this isotype and is therefore degraded intracellularly as denoted (±J). Locally produced (and serum-derived) IgG is therefore not subject to pIgR-mediated transport but can be transmitted paracellularly to the lumen together with monomeric IgA as indicated. Free SC (depicted in mucus) is generated when pIgR in its unoccupied state (top basolateral symbol) is cleaved at the apical face of the epithelium-like bound SC in SlgA and SlgM. Commensal bacteria in the right-hand panel are coated in vivo with SlgA, which aids their containment and thereby promotes host-microbial mutualism. (From Brandtzaeg P: Mucosal immunity: induction, dissemination, and effector functions, Scand J Immunol 70:505–515, 2009.)
Figure 78-7  Role of Peyer’s patch (PP) dendritic cells (DCs) in type II collagen oral tolerance. As a result of repeated oral administration of type II collagen, the production of anti-inflammatory cytokines such as interleukin-10 (IL-10) and transforming growth factor-β (TGF-β) by PP cells is enhanced and the populations of CD11c<sup>+</sup>CD11b<sup>+</sup>DCs and IDO<sup>+</sup>DCs are increased in PP. Then, through DC–T cell interaction, type II collagen inducible CD4<sup>+</sup>CD25<sup>+</sup>Foxp3<sup>+</sup> and regulatory T cells are generated in the PP. Regulatory T cells generated in the PP move to the mesenteric lymph node and then enter the systemic circulation, where they suppress the systemic immune response to CII. (From Park KS, Park MJ, Cho ML, et al: Type II collagen oral tolerance; mechanism and role in collagen-induced arthritis and rheumatoid arthritis, Mod Rheumatol 19(6):581–589, 2009.)

Figure 78-8  Integration of mucosal immunity between mother and newborn. Primed B (and probably T) cells from Peyer’s patch migrate via lymph and peripheral blood to the lactating mammary gland, resulting in the presence in breast milk of secretory antibodies (SIgA and SIgM) specific for enteric antigens. GALT, gastrointestinal-associated lymphoid tissue. (From Brandtzaeg P: Mucosal immunity: integration between mother and the breast-fed infant, Vaccine 21:3382–3388, 2003.)
The prevalence and incidence of IBD are higher in developed countries. Demographics in developing countries, however, show an increase of IBD. Prevalence of Crohn’s disease (CD) is not decreasing, although immigration from minor developed countries is accelerating. The prevalence of CD and ulcerative colitis (UC) is about equal, and in the United States one observes between 50 and 100 cases per 100,000 population. In recent years, the incidence of UC has decreased in Western countries, whereas the previously low incidence of IBD in Eastern Europe, South America, and the Pacific has increased. This may be due in part to better reporting. Ethnic affiliation has an influence on the prevalence, although it is still unknown to what extent this is due to genetic or environmental factors. For example, the Jewish population has a higher susceptibility in general but prevalence approaches that of the background population. Romanies have a remarkably lower risk of developing IBD compared with the average Hungarian population. However, ethnic differences in the prevalence of extraintestinal manifestation have not been reported so far.

The incidence of CD and UC in a large Swedish population—based study was 11 per 100,000 and 18 per 100,000 person-years, respectively. In this study it was also shown that “ever smokers” had a relative risk of 1.5 (95% confidence interval [CI], 1.2 to 1.8) and 1.3 (1.1 to 1.5), respectively. It was also shown that “ever users” of moist snuff were not at increased risk of developing IBD. Therefore the nicotine component probably does not contribute to susceptibility. On the basis of the immunosuppressive signal mediated by nicotinic acetylcholine receptors, smokeless nicotine may be innocent or indeed protective. The overall concordance in monozygotic twins is 36%, but it is only 16% in UC. The onset of disease is highest

![Figure 78-9 Immune pathogenesis of enteropathic arthritis.](image)

![Figure 78-10 Effects of colonization of the gut immune system. Segmented filamentous bacteria (SFB) are spore-forming bacteria that are related to the genus *Clostridium*. Inherited from the mother microbiota, SFB develop strong interactions with the ileal mucosa and in immunocompetent mice, the bacteria can largely recapitulate the inducing effects of the whole microbiota on the postnatal maturation of the gut immune system. SFB induce the production of *Reg III*βγ* microbicidal peptides, which protect against colonizing pathogens. Additionally, SFB simultaneously activate strong secretory IgA responses, induce the recruitment and activation of cytotoxic intraepithelial lymphocytes (IELs), and drive various T cell responses including a robust T helper 17 (Th17) cell response. In immunocompetent mice, SFB-induced proinflammatory and regulatory responses balance each other, which results in physiologic inflammation that strengthens the gut barrier. By contrast, colonization by SFB promotes the development of colitis in severe combined immunodeficient (SCID) mice that have been reconstituted with effector T cells. Intestinal colonization by SFB can also promote the development of inflammatory diseases outside of the gut. SFB promote arthritis in autoimmune nonobese diabetic (NOD) mice that express a transgenic T cell receptor (TCR) that is specific for a self-peptide (known as *K*/*BxN* mice), an effect ascribed to the induction of Th17 cells. SFB also enhance the severity of myelin oligodendrocyte glycoprotein (MOG)-induced experimental autoimmune encephalomyelitis (EAE). These aggravating effects may reflect the strong adjuvant properties of SFB. (From Cerf-Bensussan N, Gaboriau-Routhiau V: The immune system and the gut microbiota: friends or foes? Nat Rev Immunol 10(10):735–744, 2010.)
### Table 78-1 Extraintestinal Manifestations of Inflammatory Bowel Disease

<table>
<thead>
<tr>
<th>Feature or Disease</th>
<th>Crohn’s</th>
<th>Ulcerative Colitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peripheral arthritis</td>
<td>15%</td>
<td>10%</td>
</tr>
<tr>
<td>Axial or sacroiliac arthritis</td>
<td>15%-20%</td>
<td>10%-15%</td>
</tr>
<tr>
<td>Septic arthritis</td>
<td>Rare</td>
<td>Not reported</td>
</tr>
<tr>
<td>Skin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erythema nodosum</td>
<td>Up to 15%</td>
<td>&lt;15%</td>
</tr>
<tr>
<td>Erythema multiforme</td>
<td>Rare</td>
<td>?</td>
</tr>
<tr>
<td>Pyoderma gangrenosum</td>
<td>0.5%-2%</td>
<td>0.3%-0.4% in severe disease</td>
</tr>
<tr>
<td>Aphthous ulcers</td>
<td>Rare</td>
<td>1%-8%</td>
</tr>
<tr>
<td>Nephrolithiasis (oxalate)</td>
<td>&lt;15%</td>
<td>?</td>
</tr>
<tr>
<td>Amyloidosis</td>
<td>Very rare</td>
<td>Not reported</td>
</tr>
<tr>
<td>Liver disease</td>
<td>3%-5%</td>
<td>7%</td>
</tr>
<tr>
<td>Uveitis</td>
<td>13%</td>
<td>4%</td>
</tr>
<tr>
<td>Vasculitis</td>
<td>Takayasu’s</td>
<td>&lt;5%</td>
</tr>
<tr>
<td>Clubbing of fingers</td>
<td>Yes</td>
<td>1%-5%</td>
</tr>
<tr>
<td>Increased prevalence of asthma</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Increased prevalence of multiple sclerosis</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Genetics

**KEY POINTS**

- At least 30 genetic associations have been identified in CD, and several occur in UC.
- **NOD2** (previously known as **CARD15**), **IRGM**, and **ATG16L1** are associated with DCs.
- **IL23R** and **MDR1** are associated with both CD and UC.
- **IL23R** deletion protects against colitis in mice.
- CD has features of defective intracellular microbial handling.

Etiopathogenic research regarding IBD is in an active stage of development. New technology including genome-wide association studies (GWAS) has identified at least 31 genes showing replicated associations with CD and a similar number for UC. Approximately half of these associations are shared by both diseases. The strongest associations apart from human leukocyte antigen (HLA) are with the NOD2 (formerly known as **CARD15**) gene on chromosome 16 and the **IL23R** on chromosome 1. NOD2 and NOD1 are cytosolic sensors for the bacterial peptidoglycan muramyl dipeptide and trigger the synthesis of antibacterial α-defensins. Reduced mucosal expression of these defensins is found in patients with CD. IL23 signaling can result in activation of Th17 cells in the gut. The importance of IL23 is supported by the finding that **IL23R**−/− mice are resistant to induction of colitis. Another association is found with the transcription factor **STAT3**, which is involved in Th17 activation. **STAT3** is present on innate and reactive immune cells and is present in increased amounts in the gut mucosa. Another confirmed genetic association is that with **ERAP1**, an endoplasmic reticulum aminopeptidase previously known as **ARTSI**. **ERAP1** functions by trimming the length of peptides before loading into the grove of HLA, and therefore it could play a role in the triggering of immune reactions in various locations such as gut and joint. A recent Spanish study looked at the Fc receptor gene FcRL-3, which is associated with susceptibility to RA, and found an association with peripheral arthritis in CD. Although this needs replication in other studies, it may indicate an example of a gene combination influencing disease phenotype. However, it is clear that no single gene or combination of genes alone can account for manifest IBD and clearly the intestinal microflora remains a major suspect as contributor to the cause of IBD, although final proof is lacking. Experimental models of IBD require presence of gut bacteria. Postoperative therapy with metronidazole has prolonged the time to relapse. Whereas **CARD15** mutations were present in 43% of patients in the initial French study, later population studies found a much lower prevalence in northern European populations and no correlation in Asians. **CARD15** mutations are not related to susceptibility in the United States. High expression of **CARD15** messenger RNA has been found in the small intestine, and it is believed to be a regulator of nuclear factor kB (NFkB) signaling after the engagement of TLRs. **IBD3** on chromosome 6 has shown the most constant association with IBD, and **HLA-DRB1*0103** has been linked to severe UC in several studies. Further, a TNF microsatellite gene factor was associated with CD but not with UC. HLA-DR2 and DR3 associations have been linked to UC but not to CD.

Mutations of the detoxifying ATP-binding cassette, subfamily B, member 1 gene (***ABCB1**), also known as multidrug resistance 1 gene (**MDR1**), are strongly downregulated in unaffected colonic tissue of both CD and UC. **TLR4** and **TLR5** associations have been identified in several populations; this may be of special interest because they act in synergy with **CARD15** and **CARD4** in the induction of proinflammatory cytokines. **TLR** inhibitors are being investigated for therapeutic efficacy. Recently, researchers found a role for a virus as the contributing trigger in genomically manipulated mice, when mice with a disrupted autophagy gene, **ATG16L1**, were infected with a specific strain of murine norovirus and developed CD-like disease.

Pathogenesis

CD and UC are clinically distinct entities with a different pathogenesis, but genetic evidence also shows common features. Both are familial, but hereditary factors are more important in CD, according to twin studies. Whereas the entire gut wall is involved in a patchy way in CD, diffuse mucosal pathology is typical of UC. T lymphocyte proliferation and cytokine generation are also different. In CD, a
Th1 response dominates, but no such dominance has been documented for UC. Increased amounts of proinflammatory cytokines, TNF, interleukin (IL)-1β, IL-6, IL-17, and IL-8, are released locally in both diseases (see Figure 78-1).

The interplay between the intestinal microflora and genetic host factors is disturbed in IBD. The microbial contribution is still largely unclear, but animal work indicates that parts of the normal gut flora may be involved. In addition, pathogenic organisms such as Clostridium difficile have been linked to exacerbations of IBD. As discussed earlier, genetic factors related to both innate and adaptive immunity are involved in susceptibility to IBD. Experimental work with transgenic animals transfected with human HLA-B27 and β2-microglobulin has shown that certain strains of conventional mice and rats develop spondyloarthropathies, whereas identical animals in a germ-free environment are protected.

In human IBD, HLA-B27 remains a strong predisposing factor, but only in those individuals with spinal joint involvement. Jejunal fluid from patients with ankylosing spondylitis (AS) and rheumatoid arthritis collected with the closed-segment endoscopic technique contained antibodies against Klebsiella pneumoniae, Escherichia coli, and Proteus mirabilis. A disturbed and augmented local immune response in parts of the gut against a variety of microorganisms is emerging as a prevalent feature of several chronic joint diseases, but it has not been examined in IBD with this endoscopic technique. The viral contribution cited earlier indicates an important role for Paneth cells and disturbed autophagy in the pathogenesis.

Gene manipulation in mice indicates that IL-2, IL-10, and transforming growth factor (TGF)-β may be protective factors and that HLA-B27 may influence cytokine expression. Altered cytokine balance in the gut mucosa may be an important contributing pathogenic factor.

Increased gut permeability has already been alluded to as an important factor in pathogenesis. Bacteria recovered from the gut lumen in IBD are covered by immunoglobulin, part of which is circulatory IgG. Increased leakage of tissue fluid from the inflamed mucosa allows the egress of complement-binding IgG, which may contribute to inflammation and further augment permeability. The altered immune response to bacteria differs between CD and UC. Increased gut permeability in IBD is under genetic influence. Basal permeability was normal in a study of relatives of patients with CD, but it became abnormally increased after the ingestion of acetylsalicylic acid. Environmental influences on permeability may be partly mediated by bacterial endotoxin. An in vitro perfusion study on rat gut showed that serosal rather than mucosal application of endotoxin impairs the barrier. Absorbed bacterial material could therefore add to an already damaged barrier.

### Clinical Features

**KEY POINTS**

Peripheral arthritis in CD and UC occurs in two distinct forms.

- Enthesitis pain is more pronounced than in AS.
- Sclerosing cholangitis develops in 5% of IBD.

### Table 78-2 Distinct Features of Inflammatory Bowel Disease

<table>
<thead>
<tr>
<th>Feature</th>
<th>Crohn’s</th>
<th>Ulcerative Colitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicated non-HLA genetic associations</td>
<td>CARD15/NOD2, IRGM, ATG16l1, IL23R</td>
<td>IL23R</td>
</tr>
<tr>
<td>Concordance in monozygotic twins</td>
<td>36%</td>
<td>19%</td>
</tr>
<tr>
<td>Non-HLA genes</td>
<td>Several</td>
<td>Several</td>
</tr>
<tr>
<td>Gut permeability sensitive to acetylsalicylic acid on genetic base</td>
<td>Yes</td>
<td>?</td>
</tr>
<tr>
<td>T lymphocyte response in gut</td>
<td>Th1, Th17 (interferon-γ)</td>
<td>No Th1-Th2 imbalance, Th17</td>
</tr>
<tr>
<td>Fas ligand expression</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Effect of smoking</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>Correlation of gut activity to arthritis symptoms</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Intercellular adhesion molecule-1 antisense therapy</td>
<td>Beneficial</td>
<td>No response (?)</td>
</tr>
<tr>
<td>Response to anti-tumor necrosis factor therapy</td>
<td>Well established</td>
<td>Probably effective</td>
</tr>
</tbody>
</table>

Although CD and UC are clinically distinct entities, they share many features (Table 78-2). Spinal involvement occurs in 10% to 20% of cases and may be the only articular manifestation or accompany oligoarthritis. Spinal involvement is often silent, so its prevalence is underestimated; it may precede the onset of IBD or appear later. In contrast to AS, there is an equal sex distribution. In general, the involvement is similar to or identical with that in classic AS, although small differences have been found. Changes in enteropathic disease tended to be milder, squaring was more prevalent, and Romanus lesions were rare. The majority of radiographic features were similar. As noted, spinal symptoms may be mild or absent, but when present, they do not correlate with intestinal symptoms. The issue is complicated by the association of AS with silent CD, as diagnosed by biopsy. In full-blown IBD-related AS, the prevalence of B27 is between 50% and 70%, which is lower than in AS not associated with IBD.

Between 5% and 15% of patients in most studies develop peripheral arthritis, slightly more often in CD than in UC (Table 78-3). CD patients with large bowel disease (=70%)...
PART 10 | SPONDYLOARTHROPATHIES

Peripheral arthritis commonly occur together in IBD and have been linked to HLA-DRB1*0103 and TNF gene polymorphism.

Uveitis is also a feature of other spondyloarthropathies such as AS and reactive arthritis. In IBD, however, uveitis is more often bilateral, and the tendency toward chronicity is more pronounced.

Up to 5% of the patients with UC and CD can develop primary sclerosing cholangitis (PSC), a progressive liver disease that can lead to end-stage liver fibrosis.

The risk of developing colorectal cancer is increased in IBD and related to the extent of the colonic involvement and disease duration. Colorectal cancer mortality has decreased significantly in the past few decades, however, due to better surveillance and earlier diagnosis.

Diagnosis

Because of the lack of specific tests to confirm a suspicion of IBD-related arthritis, a careful history and clinical examination, supplemented by imaging, are the principal
diagnostic tools. Arthopathy can precede intestinal symptoms in a subgroup of patients, and thus colonscopy with histologic exploration can be informative regarding the origin of occurring joint symptoms. Fecal calprotectin has emerged as a sensitive screening test for IBD and helps to select cases for endoscopic exploration. As mentioned earlier, genetic mapping has shown interesting clinical correlates, but genotyping is not part of the routine clinical workup at present, except perhaps for HLA-B27. Stool cultures should be performed when infection with special pathogens is suspected. In patients with apparent IBD and monoarthritis, joint aspiration is important to exclude septic arthritis, especially before starting immunosuppressive therapy.

### Treatment

**KEY POINTS**

- Symptomatic treatment of joint disease in IBD is often adequate.
- In disease-modifying antirheumatic drug–unresponsive disease, TNF inhibitors may work.
- Interferon-γ, IL-23, and IL-6 are potential new targets.
- Probiotic efficacy is unproved.

Joint manifestations are considered secondary to active IBD. Current dogma states that treating the latter will benefit the former, but there is no rigorous proof. Placebo effects account for perhaps 20% of the treatment response in IBD; therefore only placebo-controlled evidence can be trusted. To bring into focus the arthritic discomforts, symptomatic relief is still the main target because in general, IBD-associated arthritis is nondestructive. Pain control with nonsteroidal anti-inflammatory drugs (NSAIDs) is a possible problem owing to their potential induction of flares. However, they are widely used, often well tolerated, and remain an important tool, especially in patients with primary involvement of the spine or enthesitis. Experience with the selective cyclooxygenase-2 inhibitor celecoxib, although limited, has failed to show improved gastrointestinal tolerance in comparison with other NSAIDs.

Sulfasalazine and its derivative 5-acetylsalicylic acid (5-ASA) inhibit the function of NFκB, and several studies have shown the efficacy of these drugs compared with placebo in UC but not in CD. Glucocorticoids are effective in both forms of IBD, although the response of uveitis to topical therapy with glucocorticoids may be less prompt than in other forms of uveitis. In patients with oligoarthritis, intra-articular glucocorticoid injection can be effective and safer than oral administration. Azathioprine has been widely used to maintain remission in IBD. It has proven long-term efficacy in both UC colitis and CD, according to a large European study. It should not be combined with 5-ASA owing to a pharmacokinetic interaction. If azathioprine is not tolerated, methotrexate is recommended under frequent monitoring for signs of hepatotoxicity.

TNF inhibition with infliximab results in remission of gastrointestinal manifestations in close to 60% of patients with CD, as confirmed in several placebo-controlled studies. Similar data are available for adalimumab, even in patients with previous infliximab exposure and certolizumab pegol. Interestingly, etanercept, which is not beneficial for intestinal manifestation, seems to be effective for spinal and peripheral involvement. An overview of approved anti-TNF therapy in CD is given in Table 78-4.

Infliximab was found to be superior to placebo in UC patients resistant to conventional drug therapy, although the evidence is less robust if compared with glucocorticoid therapy. However, for UC, infliximab remains the only currently approved TNF-inhibiting treatment. A recent Swedish study shows higher survival on drug therapy among patients of male sex and with peripheral arthritis.

Natalizumab, targeting α4 integrin, has shown promising effects in both multiple sclerosis and CD. However, enthusiasm was reduced after the occurrence of three cases of lethal progressive leukoencephalitis. The drug is still in limited use. ABT-874/J695, an anti-interleukin 12/23 antibody and ustekinumab, directed against the shared subunit p40, are more effective compared with placebo in first studies with patients suffering from CD.

Currently anti–interferon (IFN)-γ antibodies such as fotolizumab are being investigated for their efficacy in CD, and early data suggest a significant response compared with placebo. In a pilot study, tocilizumab, an anti-interleukin-6 receptor monoclonal antibody, showed significant but not complete response in patients with active CD. Probiotics, although commercially promoted, have not been proved effective in IBD. Metronidazole, ciprofloxacin, and other poorly absorbed broad-spectrum antibiotics have been widely tested in controlled studies, but there is no convincing evidence that they are better than placebo and they are usually inferior to glucocorticoids.

### Outcome

No prospective studies have addressed the outcome of arthritis complicating IBD. Central and peripheral arthritis shares most features with spondyloarthropathies not associated with or with silent IBD.

### BRUCELLA ARTHRITIS

#### Epidemiology

Brucellosis has been eradicated in Western Europe and North America but is still a major zoonosis in areas of South America, the Middle East, India, and other places where goat and sheep farming is practiced and poverty is prevalent. With increased global travel, sporadic cases can be expected.

---

**Table 78-4** Approved Biologics for Crohn’s Disease (as of 2012)

<table>
<thead>
<tr>
<th>United States</th>
<th>European Union</th>
<th>Switzerland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infliximab</td>
<td>Infliximab</td>
<td>Infliximab</td>
</tr>
<tr>
<td>Adalimumab</td>
<td>Adalimumab</td>
<td>Adalimumab</td>
</tr>
<tr>
<td>Certolizumab pegol (limited)</td>
<td>Certolizumab pegol (limited)</td>
<td>Certolizumab pegol (limited)</td>
</tr>
</tbody>
</table>

---
in Europe and the United States. In endemic areas, the reported incidence is between 1 and 200 cases per 100,000.\textsuperscript{52}

**Cause and Pathogenesis**

*Brucella* are small gram-negative bacteria that infect macrophages and are harbored in the liver, spleen, and bone marrow; from there, they can spread to joints. The four species causing human disease are *Brucella melitensis*, *Brucella abortus*, *Brucella suis*, and *Brucella canis*. *Brucella* arthritis is thought to be reactive on the basis of the failure to grow microorganisms from joint fluid and the poor response to antibiotic therapy, but this has never been proved.\textsuperscript{6}

**Clinical Features and Diagnosis**

Brucellosis causes arthritis in about one-third of cases. The main locations are the spine in adults and the peripheral joints in children and adolescents. Knees, hips, and ankles are the dominant peripheral locations. Sacroiliitis can be extremely acute and painful. Rising titers of serum antibodies and a confirmatory culture solidify the diagnosis. Case reports describe septic prepatellar bursitis\textsuperscript{84} and olecranon bursitis\textsuperscript{84} and indicate that fluid from these lesions can be diagnostic.

**Treatment and Outcome**

Rifampicin 600 to 900 mg and doxycycline 200 mg daily for at least 6 weeks are recommended by the World Health Organization, but other combinations have been tried.\textsuperscript{82} The arthritis can become destructive unless treated early. Spinal stenosis may be a complication.

**CELIAC DISEASE**

**Epidemiology**

CeD is a common condition with a global distribution. It used to be considered most prevalent in children, but new evidence shows that it is even more common in adults. Intestinal symptoms may be minimal or absent; consequently, published prevalence figures of 1% may be too low.\textsuperscript{6,85,86} In juvenile idiopathic arthritis CeD has been found in 6% to 7% of cases.\textsuperscript{86}

**Cause and Pathogenesis**

*CeD* is caused by an immune reaction to partly digested wheat gluten by T lymphocytes in the gut of genetically HLA-DQ2–positive or HLA-DQ8–positive individuals. A seminal observation was the finding that tissue transglutaminase is the major autoantigen in CeD.\textsuperscript{85} It was shown in 2002 that dietary gluten is partly digested by gastric enzymes to generate a stable 33–amino acid peptide that is deamidated by tissue transglutaminase.\textsuperscript{87} The peptide is then presented in the context of HLA-DQ2 or HLA-DQ8 to CD4\textsuperscript{+} T cells, resulting in IFN-\(\gamma\) release and inflammation, altered gut permeability, and eventually villus atrophy. Autoantibodies against tissue transglutaminase are also formed.\textsuperscript{85} CeD can now be considered as a systemic disease that may involve GI-related pathology, as well as endocrine, skin, locomotor, and neural abnormalities (Figure 78-14).\textsuperscript{86}

**Clinical Features and Diagnosis**

Only two-thirds of patients present with diarrhea or irri-tated bowel symptoms. Nonspecific signs such as fatigue, headache, and arthralgias may occur and delay a correct
WHIPPLE’S DISEASE

Epidemiology

Whipple’s disease is a rare condition. Incidence and prevalence figures are unknown. A retrospective French study identified 52 patients, 73% of whom were men.\(^6\)

Genetics and Pathogenesis

Whipple’s disease, or intestinal lipodystrophy, as it was initially called in 1907,\(^6\) is an intestinal infection with a unique microorganism called *Tropheryma whipplei*, belonging to the Actinomycetes family. The organism has been found in sewage, but the source of infection in humans is not known.\(^9\)

Six different genotypes have been confirmed in culture from diseased tissue.\(^91\)

The organism lives in macrophages, and these elicit a skewed lymphocyte response, with suppressed Th1 dominating Th2 cells. Expression of the cytokine IL-16 stimulates growth of the pathogen.\(^92\) In vitro monocytes from patients with Whipple’s disease express CD163, possibly explaining a reduced oxidative impact against *T. whipplei*.\(^93\)

A correlation between different clinical manifestations and variable genotypes of the bacterium has not been identified, and host factors seem to be crucial for the kind of diagnosis.\(^86\) CeD is a systemic disease that can involve type I diabetes, anemia, osteoporosis, neuropathies, and joint symptoms in up to 25% of patients.\(^88\) This can be an asymmetric oligoarthritis or polyarthritis, and axial involvement is common. Arthritis may be the presenting symptom of the disease.\(^89\)

In addition to small bowel biopsy, the diagnosis can be established by assaying for the presence of serum for IgA antitissue transglutaminase antibodies and IgA antiendomysial antibodies.\(^90\)

Treatment and Outcome

Elimination of gluten from the diet is the rational therapy and is often the only one required. In addition, various experimental approaches have been discussed but are not yet supported by data. These include the administration of IL-10 to boost regulatory T cells, the induction of tolerance by nasal application of gluten peptides, and gene therapy. No specific therapy has been established for the joint problems.

Children with verified CeD still have abnormal mucosa in adulthood and must continue the dietary restriction. There are no outcome reports dealing with joint involvement.
Clinical Features

The disease can have many faces and may remain undiagnosed for many years. Recurrent fever; malaise; hematologic, pulmonary, and cardiac disturbances; and neurologic and ophthalmic symptoms are sometimes present and misinterpreted. Articular symptoms, however, are the presenting feature in 67% of cases, compared with intestinal symptoms in only 15%. Eventually, 83% of patients develop diarrhea, abdominal pain, and malnutrition. Arthralgias and arthritis are most commonly seen in knee joints but can localize in any peripheral joint, as well as in spinal joints and disks. Sacroiliitis has been described.

Diagnosis

The diagnostic test of choice rests on immune histology, with the occurrence of periodic acid–Schiff–positive material, an abundance of CD68+ macrophages, and staining with antisera specific for *T. whipplei*. Quantitative PCR of saliva and stool has been suggested as a noninvasive screening method.94 PCR of cerebrospinal fluid can detect central nervous system involvement.95 The organism can grow out in culture, which takes an average of 30 days. In one study, only 2 of 10 small bowel specimens were culture positive; the yield is higher using sterile cardiac or nerve tissue.96 Culture therefore remains a research tool.

Treatment

No randomized, controlled studies are available. Initial treatment should be ceftriaxone for 2 weeks to ensure entrance into the central nervous system. Then oral trimethoprim-sulfamethoxazole is administered for a prolonged or indefinite period. In case of intolerance or lack of efficacy, tetracycline can be used. A prospective comparison between ceftriaxone and meropenem treatment for 2 weeks followed by trimethoprim-sulfamethoxazole for 1 year showed good 3-year remission in both groups.96 However, immune histology still shows some evidence of remaining pathology,97 indicating that recurrence could be anticipated after discontinued therapy.

Use of penicillin, streptomycin, and chloramphenicol has been abandoned.98

Outcome

Without treatment, Whipple’s disease is chronic or relapsing, usually progressive, and ultimately fatal. With adequate antibiotic therapy, clinical remission is usually complete or near complete.

MICROSCOPIC COLITIS

Microscopic colitis (MC) is a name given to two conditions presenting with profuse diarrhea. The first, collagenous colitis (CC), was described in 1976, and the second, lymphocytic colitis (LC), was published in 1989.99 They are now joined by a rare third condition called collagenous gastritis (CG).100 Apart from gastrointestinal symptoms, they all may be associated with extraintestinal autoimmune manifestations.

Epidemiology

Although initially considered rare, it is now evident that MC is a common cause of watery diarrhea in several populations. The minimal overall annual incidence is 1 to 5 cases per 100,000. It is 5 to 10 times more common in individuals older than the age of 65 and distinctly more common in women.100 The female-to-male ratio is 7:1 in collagenous colitis and 2:1 in lymphocytic colitis. The peak incidence is seen among those 60 to 80 years old. For some years the incidence seemed to be increasing, but this has now leveled off.101

Cause

The cause is unknown. Although no strong genetic factor has been identified, an association with HLA-DQ2 and with a polymorphism in the TNF gene has been described.100 Exposure to antirheumatic therapy has been suspected in MC patients with rheumatoid arthritis on the basis of the observation that arthritis usually precedes the onset of CC. Reaction against some luminal factor is deduced from the observation that histology is normalized when ileostomy is performed but recurs after closure. Drugs that are prime suspects include NSAIDs and acetylsalicylic acid, Lansoprazole, ranitidine, sertraline, and ticlopidine.102 CC, LC, and CG all seem to be associated with CD, autoimmune thyroid disease, and arthritis, but no causal relationship is established.

Pathogenesis

Infection is suspected, but no agent has been identified. Luminal factors are strongly suggested by almost complete histologic normalization after performing ileostomy and recurrence of pathology and symptoms after its closure.102 Connective tissue growth factor was markedly increased in the subepithelial zone of biopsies from CC patients analyzed by reverse transcription PCR, indicating a pathogenic role for connective tissue growth factor (CTGF).103 If confirmed, this may lead to targeted therapy of CC.

Clinical Features

Chronic, intermittent (or sometimes chronic, persistent) painful watery stools, weight loss, and fatigue are the cardinal intestinal symptoms. There is no difference between collagenous and lymphocytic colitis in this regard. The course is acute or chronic but usually benign.

CC and LC are associated with a variety of rheumatic syndromes in 10% to 20% of cases. These include Sjögren’s syndrome, nondestructive oligoarthritis, migratory arthralgias, sacroiliitis, and rheumatoid arthritis. A survey of 63 consecutive cases in one Swedish center identified 8 cases of rheumatoid arthritis and 3 cases of AS, clearly
monotherapy or in combination with cholestyramine 4 g/day. Budesonide 9 mg/day has good immediate effect but gut relapse is common.

**PONCET’S DISEASE AND BACILLE CALMETTE-GUÉRIN–INDUCED ARTHRITIS**

Tuberculous arthritis, or Poncet’s disease, is a rare aseptic form of insidious fever, weakness, and arthritis described mostly in young adults suffering from extrapulmonary tuberculosis. It responds slowly to antituberculous therapy, and in the absence of pulmonary changes, the intestine is assumed to be the port of entry. The attenuated Mycobacterium strain bacille Calmette-Guérin (BCG) is used intradermally as an adjuvant in cancer therapy to stimulate T cell–mediated immunity; it is also instilled into the urinary bladder to treat superficial cancer.

**Epidemiology**

No epidemiologic data are available. A recent review identified 50 bona fide cases of Poncet’s disease. Aseptic arthritis occurs in 0.4% to 0.8% of patients treated with the instillation of BCG for bladder malignancy, and anecdotal evidence indicates an increased prevalence of HLA-B27 among them. This finding was associated with sacroiliitis in 20% and oligoarthritis with predominant localization to the lower limbs; it occurred more often in men. In cases of reactive arthritis occurring after the intradermal administration of BCG, 6 of 10 patients were women and symmetric hand arthritis dominated. In view of the global increase of tuberculosis incidence, one should exercise vigilance regarding possible new cases of Poncet’s arthritis.

**Cause and Pathogenesis**

An aseptic complication of active tuberculosis or the administration of BCG precipitates the process. By definition, it is a reactive arthritis, which means that the infectious agent triggers an immune reaction; however, the enteropathic nature is not firmly established. HLA-B27 may be a susceptibility factor in post-BCG arthritis. Mycobacterium heat shock protein 65 has been incriminated in both these sterile forms of arthritis, as well as in others. Mycobacterial and human heat shock proteins are 50% homologous, and one hypothesis is that both the therapeutic efficacy and the arthritis are caused by cross-reactive T lymphocytes. Heat shock protein also has homologies with proteoglycan and HLA-DR. The pathogenesis of arthritis after intravesical instillation of BCG might be different and related to antigen persistence, setting the stage for a kind of reactive arthritis.

**Clinical Features and Diagnosis**

Insidious fever, weakness, and arthritis are described mostly in young adults suffering from extrapulmonary tuberculosis. The arthritis consists of oligoarthritis or polyarthritis of large or small joints, or both. The onset is not as acute as
in regular enteric reactive arthritis. Most peripheral joints can be affected. An interesting case was identified in Finland with the help of a sensitive ELISPOT IFN-γ release test. The patient was a female 55-year-old hospital attendant presenting migratory arthritides of an ankle and toes. Arthritis developing in the presence of active tuberculosis or after recent exposure to BCG and proven to be aseptic is sufficient for diagnosis.

**Treatment and Outcome**
There is no established treatment. Post-BCG cases usually heal within 3 months. However, despite negative culture results, 4 to 6 months of conventional antituberculous therapy is often practiced.

**ENTEROVIRAL AND HEPATITIS VIRUS–ASSOCIATED ARTHRITIS**

**Epidemiology**
Enterovirus infections, like other common virus infections, often give rise to arthralgias. The precise prevalence is not known. Hepatitis B and hepatitis C disease cause joint manifestations in 10% to 25% of cases.

**Genetics and Pathogenesis**
No genetic factors have been identified. Enterovirus species can invade joints and may rarely be isolated from joint fluid. Immune complex–mediated activation may result in viral hepatitis. A preferential deposition of such immune complexes in joints results in arthralgias and arthritis. In hepatitis C disease, precipitation of cryoglobulins is often observed. Molecular mimicry has been proposed but not proved. Viral persistence in host cells with continuous production of viral protein expressed on cell surfaces has been observed. Immunosuppressive therapy may reactivate latent virus growth and may be misinterpreted as increased disease activity in patients with other rheumatologic conditions.

**Clinical Features**
Hepatitis B may be associated with a variety of general symptoms including fever, abdominal pain, nausea, and vomiting after a prodromal period of up to 6 months. Among these patients, 10% to 25% also develop articular symptoms, which may present during the prodromal phase. Preferentially, locations are hands and knees. In addition to morning stiffness, migratory arthritis is common. Interestingly, a coincident appearance of urticaria-like rash, predominantly of the legs, is common. The arthritis characteristically disappears with the onset of jaundice. Less than 5% of hepatitis B–infected individuals develop a chronic disease with persistent or remitting arthritic manifestations. The disease is self-limited and only requires symptomatic treatment.

Hepatitis C virus is associated with joint symptoms in 20% of cases. Other frequent symptoms are myalgia, glomerulonephritis, and vasculitis or essential mixed cryoglobulinemia, which may become severe. Oligoarthritis occurs in one-third of the cases, whereas two-thirds develop polyarticular disease.

Enteroviruses, coxsackieviruses, and echoviruses are only associated with arthropathy in less than 1% of cases, but due to the high prevalence of the primary disease the number of arthritis cases is not trivial. After an incubation period of 3 to 5 days the patients develop fever, fatigue, headache, pharyngitis, myalgia, rash, pleuritic pain, and conjunctivitis or, alternatively, nausea, abdominal pain, enteric cramps, vomiting, and diarrhea. Joint symptoms are self-limited and disappear within days, although recurrence can occur. Any peripheral joint may be affected.

**References**
5. Iqbal TH, Lewis KO, Geary JC, Cooper BT: Small intestinal permeability to maninol and lactulose in the three ethnic groups resident in west Birmingham, *Gut* 39:199, 1996.
CHAPTER 78 | ENTEROPATHIC ARTHRITIS


1268 PART 10 | SPONDYLOARTHROPATHIES


87. Shan L, Molberg O, Parrot I, et al: Structural basis for gluten intoler...
Systemic lupus erythematosus (SLE) represents one of the most significant diseases in all of medicine. Predominantly targeting young women in their childbearing years and with the potential to cause significant physical disfigurement, morbidity, and occasionally mortality, lupus is the focus of strong advocacy to support research that will generate insights into disease pathogenesis. In fact, characterization of the immunologic contributors to lupus, the prototype systemic autoimmune disease, has been the focus of particularly intense study since the flowering of the discipline of immunology in the 1950s and 1960s. Recent efforts to define the genetic variations that underlie susceptibility to lupus have supported the central role of the immune system in disease pathogenesis but have extended the view of lupus pathology beyond the important role of autoantibodies to include a significant contribution of the innate immune system to disease. An underlying role for the vasculature as a target of the immune system and its products is gaining renewed interest as an important component of lupus pathogenesis. Together, these recent advances provide important insights into how the intersection of genetic variations with environmental triggers amplifies immune system activation and target organ vulnerability to generate the classic manifestations of lupus and its clinically significant comorbidities. Figure 79-1 provides a schematic overview of many of the contributors to SLE and how they interact to drive autoimmunity and tissue damage.

**HISTORICAL VIEW OF LUPUS PATHOGENESIS**

Insights into the immunopathogenic mechanisms that account for development of autoimmunity in patients with SLE and the tissue damage that results in disease have come in fits and starts over many decades. Progress has been informed by the scientific tools available at the time along with the acute observations of clinicians and research scientists. Descriptions of the clinical manifestations of disease suggested a multisystem disease that typically involves skin and joints, with renal, cardiac, and neurologic pathology suspected and then documented once histologic studies were performed. Alterations of blood vessels in many organs were recognized as an important component of the disease process in early studies, with the pathognomonic “onion
Contributors to systemic lupus erythematosus (SLE) pathogenesis. Mechanisms that promote development of SLE are related to the underlying genetic profile of the individual. Many of the disease-associated genetic variants (examples are illustrated) contribute to excessive production or impaired clearance of stimulatory nucleic acids; increased generation of products of the innate immune response, particularly type I interferon (IFN); or an altered threshold for activation or efficiency of signaling of cells of the adaptive immune response. In most cases multiple genetic risk variants are required to establish a state of immune activation that is receptive to environmental triggers that promote development of autoimmunity. In rare cases, a mutation of a critical regulator of immune activation might be sufficient to initiate the altered immune state that can lead to disease. Type I interferon is a product of plasmacytoid dendritic cells (pDCs), and activation of those cells by intracellular nucleic acids or exogenous triggers such as a virus or debris derived from damaged or dying cells might represent mechanisms of initiation of disease. Once IFN-α is produced, it mediates numerous effects on immune system cells that mimic the response to a viral infection. The antigen-presenting capacity of myeloid dendritic cells can be augmented, promoting activation of self-reactive T cells and differentiation of B cells toward production of pathogenic antibodies. Activated T cells express CD154 (CD40 ligand) and produce interleukin-21 (IL-21), providing effective help for B cells to generate antibody-producing plasma cells. IFN-α also supports the production of B cell activating factor (BAFF), a survival and differentiation factor for B cells. Once autoantibodies are produced, immune complexes amplify immune activation by accessing endosomal Toll-like receptors in pDCs and B cells and deposit directly in the vicinity of blood vessels, inducing complement activation, inflammation, and tissue damage. Reactive oxygen species (ROS) and proinflammatory cytokines produced by monocytes and macrophages contribute to tissue damage, as does IFN-α, which stimulates endothelial cells and is associated with poor vascular repair and sclerosis. BCR, B cell receptor; DC, dendritic cell; MHC, major histocompatibility complex; MØ, macrophage; PMN, polymorphonuclear neutrophil; RNA, ribonucleic acid; TCR, T cell receptor.

skinning” of splenic arterioles, cellular infiltration and damage to renal glomeruli, and vasculopathy in skin and brain demonstrated clinically by manifestations such as nephritis, livedo reticularis, and stroke. The lupus erythematosus (LE) cell, described by Hargraves in 1948, suggested that engulfment of cellular debris, a mechanism that remains an important concept in considerations of lupus pathogenesis, was active at sites of inflammation.

The recognition that antibodies directed at cellular components, particularly cell nuclei, were present in the sera of patients with SLE directed the attention of the investigator community toward the immune system and the conclusion that lupus reflected an autoimmune process. Together with pathologic studies of kidneys from lupus nephritis that showed deposition of immunoglobulin and complement components in kidney glomeruli, the elution of anti-deoxyribonucleic acid (DNA) autoantibodies from lupus kidneys contributed to the concept that autoantibodies, particularly anti-DNA antibodies, were pathogenic. Isolation from lupus sera of complexes of antibody with DNA or DNA-binding proteins such as histones was an important factor in classifying SLE as predominantly an immune complex–mediated autoimmune disease. Although lupus pathogenesis does depend on those immune complexes and
the inflammatory responses that they induce, subsequent investigation identified virtually all cell components and many soluble immune system products as contributors to the immune system dysfunction that ultimately accounts for disease in SLE. The development of the field of cellular immunity in the 1970s and the later identification of the T cell antigen receptor, along with families of co-stimulatory molecules mediating interactions between T cells and antigen-presenting cells or B cells, led to the implication of self-reactive T cells as important regulators of immune responses, as well as helpers for B cell differentiation to autoantibody-producing cells in the case of patients with SLE. Numerous deficiencies in T cell function including altered production of typical T cell cytokines such as interleukin-2 (IL-2) were described, and studies of murine lupus models strongly supported the essential role that T cells play in lupus pathogenesis.

The discovery of Toll-like receptors (TLR) in the 1990s, the recognition that those receptors recognize common determinants expressed by microbes, and the demonstration that TLRs mediate immune system activation was a major advance, perhaps the most significant milestone in many decades, that has led to gains in understanding of lupus pathogenesis. Characterization of the typical exogenous ligands of each of the TLRs has been followed by identification of endogenous ligands for their receptors, with nucleic acids being the most relevant for amplification of immune system activation and autoimmunity in SLE. It is not yet clear if the TLRs act to augment immunologic activity that is initiated by other molecular pathways or if they are also essential as sensors for the primary initiators of autoimmune disease.

In spite of these considerable gains in elucidating the pathogenesis of SLE, the environmental triggers and genetic susceptibility factors that lead to the initiation of autoimmunity in some individuals but not in others remain largely undefined. Regarding triggers of disease, clinical observations have pointed to exposure to sunlight, microbial infection, and certain drugs as factors that can lead to initiation or exacerbation of lupus. Whether there is a common thread among these environmental triggers is not clear, but ultraviolet light–mediated DNA damage and modification of DNA methylation are among the processes that might render self–nucleic acid stimulatory to the immune system.

Significant technologic advances in recent years, together with development of consortia of investigators who have cooperated to share biologic samples from patient cohorts and control subjects, have supported important progress in defining the genetic variants that show a statistical association with a diagnosis of SLE. These recent genome-wide association studies (GWASs) have built on long-standing observations of the high risk of disease in patients with complement deficiencies to identify lupus-associated genes that encode important components of molecular pathways that promote production of type I interferon (IFN), alter the threshold for activation of lymphocytes, or generate immune stimuli.

Together, these scientific advances have led to the recent, long-awaited activity in drug development programs targeting mechanisms likely to impact lupus disease activity or clinical progression. This chapter reviews those recent studies that have illuminated the immunopathogenesis of SLE.

**GENETIC CONTRIBUTIONS TO LUPUS PATHOGENESIS**

Current concepts in genetic studies of complex human diseases including SLE identify both common variants with a small impact on risk of disease, as well as low-frequency mutations that can have a great impact on the risk of developing lupus or a lupus-like syndrome.

Observations of several individuals with SLE within a family, along with a high frequency of concordance of SLE in identical twins, have pointed to a strong genetic contribution to SLE. Data suggest that concordance of clinical lupus disease in twins is 10 times more frequent in monozygotic than in dizygotic twins, although the highest reported concordance rate is still only 57%. Moreover, the suggestion that multiple autoimmune diseases are associated with common genetic susceptibility factors is supported by the aggregation of several distinct autoimmune diseases within a family. The pace of discovery of common genetic variants, typically identified by statistical analysis of single nucleotide polymorphisms (SNPs) in relation to a diagnosis of SLE, has markedly accelerated in recent years, as the cost of genotyping thousands of patient and control samples has become feasible. Publication of two important collaborative genome-wide association studies, one by the SLE Genetics (SLEGEN) consortium and the other organized by Genentech, identified at least nine new genes or genomic loci associated with SLE. Since presentation of those studies, additional lupus-associated genetic variants have been confirmed or at least strongly supported. The common theme among the lupus-associated genes is that the vast majority encode proteins implicated in immune system function. Although this result is not surprising, it confirms the essential contribution of the immune system to the autoimmunity, inflammation, and tissue damage that characterize lupus disease and points to the most significant molecular pathways that mediate altered immune function.

The lupus-associated genes that play likely roles in lupus pathogenesis can be grouped on the basis of their roles in immune function (Table 79-1). In parallel to the requirements for immune system activation stimulated by foreign antigens, SLE-associated genes are involved in generation of self-antigen, activation of the innate immune response, and activation of the adaptive immune response. Although the specific functional alterations that are conferred by the lupus-associated variant compared with the more common variant have yet to be defined in detail, in most cases there is sufficient information regarding function to allow formulation of hypotheses that can be tested in functional genetic studies.

The rare but high-risk deficiencies in complement pathway gene products, including C2, C4, and C1q, are thought to contribute to lupus pathogenesis by impairing clearance of cellular debris, a function that is typically supported by those complement components. Increased availability of nuclear debris can provide sufficient self-antigen for induction of self-reactive T cells or serve as an
endogenous adjuvant for activation of the innate immune response. The ancestral major histocompatibility complex (MHC) 8.1 haplotype block, HLA-B8/DR3/DQw2/C4AQ0, that is associated with lupus susceptibility, encodes the alleles B8 and DR3 and bears a short C4B gene and no C4A gene, whereas other nonrisk haplotypes carry either a longer C4B segment and/or one or more copies of C4A. In fact, the relative risk related to the C4A null allele is twice that of either HLA-B8 or DR3, pointing to the significance of the C4 genes in disease risk. It should be noted that this risk haplotype is also associated with accelerated disease in patients infected with human immunodeficiency virus (HIV), insulin-dependent diabetes mellitus, and several other autoimmune diseases. Deficiency of C1q, the recognition protein for the classical complement pathway, might contribute to disease on the basis of its important role in promoting clearance of apoptotic cell debris by mononuclear phagocytes. C1q plays an additional role in inhibition of IFN-α production by directing stimulatory immune complexes to monocytes rather than IFN-producing plasmacytoid dendritic cells. Through this mechanism C1q deficiency can augment IFN-α and promote broad immune dysregulation. C-reactive protein (CRP), a member of the pentraxin family, also contributes to clearance of apoptotic debris. Polymorphisms in CRP have been associated with SLE and with decreased levels of CRP, but it remains unclear how much of the genetic contribution to basal CRP levels is due to variations within the gene itself versus variations in other genes.

Recent studies of families with a lupus-like disease, the Aicardi-Goutières syndrome, suggest that mutations in genes encoding nucleases that cleave either DNA or ribonucleic acid (RNA) might result in excess stimulatory nucleic acid and innate immune system activation. Aicardi-Goutières and several related conditions are characterized by skin lesions, autoantibodies, central nervous system disease, and high levels of type I IFN. Mutations in the TREX1 gene, encoding DNAse III, a 3′-5′ exonuclease, as well as two additional genes, RNASEH2 and SAMHD1, are associated with this syndrome. A functional connection between TREX1 and control of type I IFN production was demonstrated in a murine model in which TREX1 deficiency resulted in increased levels of IFN-β. A recent analysis of more than 8000 lupus patients found TREX1 mutations in approximately 0.5% of patients. In addition, a common variant in the TREX1 gene conferred an odds ratio for diagnosis of lupus of 1.73 compared with healthy controls. Taken together, these recent demonstrations of association of rare genetic variants of enzymes that regulate degradation of nucleic acids with SLE point to the central role of those nucleic acids as triggers for immune system activation and disease.

A large number of lupus-associated single nucleotide polymorphisms (SNPs) are found in genes that encode proteins involved in induction of type I IFN or response to that family of cytokines. IFN regulatory factor 5 (IRF5) and IRF7 are cytoplasmic proteins that translocate to the nucleus after effective activation of the endosomal TLRs by DNA or RNA. IRF5 and IRF7 then act as transcription factors to initiate transcription of IFN-α and other proinflammatory mediators. TNFAIP3 encodes A20, a protein that regulates several proinflammatory cellular mechanisms including signaling through endosomal TLRs and activation of nuclear factor κB (NFκB). Genetic variants in IRF5, IRF7, TNFAIP3, and numerous other lupus-associated genetic variants can be mapped to molecular pathways responsible for induction of innate immune system activation or responsiveness to its products, particularly IFN-α.

A third set of lupus-associated gene variants contributes to altered thresholds for lymphocyte activation or efficiency.
of cell activation. The ancestral MHC 8.1 haplotype that has been shown to be strongly associated with a diagnosis of SLE, along with other autoimmune diseases, appears to influence the early stages of immune system activation. Although more efficient presentation of self-antigens by disease-associated MHC class II alleles would seem to be the most likely mechanism by which the MHC risk haplotype confers predisposition to autoimmunity, available data point to a number of immune alterations, many focused on the T cell, in healthy individuals bearing the ancestral haplotype. The impact of those alleles is best defined clinically as determining immune responses that generate particular autoantibody specificities, and the alleles seem to be important in determining whether anti-DNA autoantibodies, antibodies specific for RNA-associated proteins, or both types of autoantibodies can be induced and produced through T cell–dependent B cell differentiation. Additional lupus-associated variants that alter adaptive immune system activation are involved in cytokine signaling such as STAT4 or efficiency of signaling downstream of the T and B cell surface antigen receptors such as PTPN22 in the case of both T and B cells and LYN, BANK, BLK, TNFAIP3, and others in the case of B cells.

A fourth category of lupus-associated genetic variants defines determinants of target organ damage. The understanding of genetic determinants of target organ vulnerability to immune mediated or oxidative damage is less well developed than is the role of genetic variants in altered immune system function in SLE, but identification of polymorphisms in members of the kallikrein gene family are associated with SLE and suggest areas for future investigation. Table 79-1 lists many of the genetic variants documented to have a statistical association with a diagnosis of SLE.

Although impressive progress based on GWAS has identified statistical associations of sequence variations in genes or genetic loci with a diagnosis of SLE, the functional consequences of those variations have not been extensively characterized. The best developed insights regarding the impact of lupus-associated variants on immune function have derived from studies of gene products involved in production of type I IFN. The risk alleles for IRF5 and IRF7 are associated with increased serum type I IFN activity in those patients who demonstrate autoantibodies targeting DNA or RNA-associated proteins. Those studies support the hypothesis that nucleic acid–containing immune complexes are important stimuli that act through endosomal TLRs such as TLR7 and TLR9, those TLRs that mediate cell signals through the IRAK and IRF7 transcription factors. The impressive collaborative efforts that have successfully collected sufficient patient and control samples and performed the required genotyping and statistical analyses have defined or at least confirmed those molecular pathways that are central to the immunopathogenesis of SLE. The strong support for activation and altered regulation of the TLR pathway, along with variations in lymphocyte signaling, point to potential therapeutic targets for future study. In addition, the recent identification of less common mutations in enzymes required for degradation of intracellular nucleic acids suggests that further study of the TLR-independent pathways of innate immune system activation, mediated by nucleic acid sensors and their associated kinases and adaptor molecules, might be another fruitful area for further investigation.

Subphenotype analysis will continue to amplify the information that can be gleaned from patient and control genotyping. A recent study of anti-dsDNA+ and anti-dsDNA− SLE patients, along with healthy controls, identified HLA-DR3, STAT4, and ITGAM as significantly associated with the presence of anti-dsDNA antibodies. A case-only analysis additionally associated PTPN22, IRF5, and PTTG1 with anti-dsDNA antibodies. In contrast, a distinct group of lupus-associated genes showed comparable association between anti-dsDNA positive and negative patients, including FCGR2A, OX40L, IL10, PTK, UHRF1BP1, PRDM1, BLK, and IRAK1. Although it is possible that some of these variants are associated with distinct autoantibody specificities, it is more likely that they represent risks for other aspects of lupus pathogenesis, perhaps contributing to inflammation and tissue damage.

Whether the insights regarding specific genetic susceptibility factors can be used to predict development of lupus or particular manifestations of disease is as yet unclear. Recent studies have attempted to determine the predictive value of accumulated genetic risk variants. Although increased numbers of lupus-associated alleles are associated with some increased risk of disease and the number of lupus-associated genetic alleles is significantly higher in patients with anti-DNA antibodies than in those without those autoantibodies, at this time, it would not appear that genotyping for lupus risk is sufficiently informative to warrant practical application in patient management.

**FEMALE PREDOMINANCE OF SYSTEMIC LUPUS ERYTHEMATOSUS**

A discussion of genetic contributions to SLE pathogenesis cannot ignore the dramatic 9:1 female predominance of the disease. Of all of the characteristic clinical features of lupus, it is the extreme sex skewing that remains least understood. Hormonal contributions to immune system activation are likely to represent a component of the female predominance of the disease; estrogen can modulate lymphocyte activation, and prolactin has been shown to be expressed at increased levels in lupus serum.

Granting a contribution of hormones to increased immune activation, it would appear that additional concepts should be entertained to understand why 9 or 10 females are diagnosed with SLE for every male lupus patient. It is intriguing that Klinefelter’s syndrome, characterized by a 47,XXY genotype, is increased 14-fold among men with SLE compared with men without SLE. These data are proposed to support an X chromosome gene dose effect as an important contributor to SLE pathogenesis. Identifying the X chromosome as a possible risk factor for SLE provides a basis for new hypotheses regarding disease susceptibility, but the nature of that risk remains undefined. A possible role for altered regulation of epigenetic processes such as DNA methylation has been raised, and murine studies have supported the impact of duplications of portions of the X chromosome that encode the TLR7 gene in activation of the innate immune system, production of type I IFN, and generation of autoimmunity. Investigation of
altered gene dosage in human patients has not confirmed a similar duplication of the TLR7 gene, and studies of DNA methylation are most suggestive of altered DNA methylation reflecting generalized immune activation rather than a primary etiologic event. Further studies of epigenetic regulation of X chromosome structure and gene expression are warranted.

A potential pathogenic role for the distinct events that occur in the ovary compared with those in the testes also deserves further study. Onset of SLE most typically occurs in the childbearing years, after menarche and before menopause. A positive association between early menarche and SLE has been observed, and breastfeeding confers a protective effect. Both observations might be consistent with molecular and cellular events related to ovulation as factors contributing to lupus pathogenesis. It is intriguing to consider that the biology of germ cell maturation, with female germ cells undergoing a second meiotic division before ovulation each month, might involve mechanisms and mediators that affect immune recognition. Although not yet fully understood, the carefully orchestrated demethylation and remethylation of DNA, along with the production of regulatory RNAs and RNA-associated proteins such as so-called PIWI proteins, in the germ cells and associated somatic cells might provide a source of stimulatory nucleic acid–containing complexes that could access TLR-dependent or TLR-independent pathways and result in immune activation. The challenge in pursuing such concepts is the obvious limitation on access to ovarian tissue for study, although murine models might be helpful in that regard.

ENVIRONMENTAL TRIGGERS OF LUPUS

Concordance rates for development of SLE in identical twins range from 24% to 57%, and it is understood that both environmental factors and stochastic events (i.e., chance) contribute to development of SLE in an individual. Although some environmental triggers of disease have been identified on the basis of clinical observation and epidemiologic studies (Table 79-2), in general there is incomplete understanding of the range of factors that can induce disease and the mechanisms by which they do so. Socioeconomic factors have been demonstrated to contribute to poor outcomes in lupus patients, likely related to poor access to care. The distinct contributions of low socioeconomic status versus genetic factors that impact disease severity in certain ethnic groups are difficult to disentangle.

Clinical manifestations that are often present at the time of diagnosis including fatigue and arthralgias have suggested that a virus might initiate the disease. In fact, epidemiologic studies demonstrating higher prevalence of antibodies specific for Epstein-Barr virus (EBV) antigen in children with SLE compared with the general population, as well as higher frequency of anti-EBV antibodies in a military cohort before disease pathogenesis. The molecular basis of this apparent cross-reactivity is not known, although DNA can associate with EBNA-1 and might account for the described antibody reactivity. T cell responses specific for EBV have been studied in patients with lupus and found to be deficient, perhaps contributing to the increased numbers of EBV-infected mononuclear cells and increased copy number of EBV DNA found in SLE patients.

Environmental toxins are of interest, yet their potential role in lupus pathogenesis has not been comprehensively explored. Data support current smoking as a risk factor for SLE, with a dose response relating pack years of smoking to risk. As is currently proposed with regard to pathogenesis of rheumatoid arthritis, smoking might provide an inflammatory stimulus to epithelial or mononuclear cells in the lungs, promoting protein modification or nonspecific

### Table 79-2  Environmental Factors That Might Play a Role in the Pathogenesis of Systemic Lupus Erythematosus (SLE)

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite</td>
<td>Ultraviolet B light</td>
</tr>
<tr>
<td>Probable</td>
<td>Estrogen and prolactin—In humans, female-to-male ratio is 9:1 between menarche and menopause, 3:1 in young and old EBV</td>
</tr>
<tr>
<td></td>
<td>Lupus-inducing medications*</td>
</tr>
<tr>
<td></td>
<td>Hydralazine</td>
</tr>
<tr>
<td></td>
<td>Procainamide</td>
</tr>
<tr>
<td></td>
<td>Isoniazid</td>
</tr>
<tr>
<td></td>
<td>Hydantoins</td>
</tr>
<tr>
<td></td>
<td>Chlorpromazine</td>
</tr>
<tr>
<td></td>
<td>Methyldopa</td>
</tr>
<tr>
<td></td>
<td>Penicillamine</td>
</tr>
<tr>
<td></td>
<td>Minocycline</td>
</tr>
<tr>
<td></td>
<td>Tumor necrosis factor inhibitors</td>
</tr>
<tr>
<td></td>
<td>Interferon-α</td>
</tr>
<tr>
<td>Possible</td>
<td>Dietary factors</td>
</tr>
<tr>
<td></td>
<td>Alfalfa sprouts and related sprouting foods containing canavanine</td>
</tr>
<tr>
<td></td>
<td>Pristane and other hydrocarbons</td>
</tr>
<tr>
<td></td>
<td>Infectious agents other than EBV</td>
</tr>
<tr>
<td></td>
<td>Bacterial DNA</td>
</tr>
<tr>
<td></td>
<td>Smoking</td>
</tr>
<tr>
<td></td>
<td>Human retroviruses or endogenous retroelements</td>
</tr>
<tr>
<td></td>
<td>Endotoxins, bacterial lipopolysaccharides</td>
</tr>
<tr>
<td></td>
<td>Vitamin D deficiency</td>
</tr>
</tbody>
</table>

*Although each drug listed and many others can induce lupus-like symptoms in predisposed individuals, there is little evidence that they can induce true SLE or even activate disease in individuals with spontaneous, established SLE. If the clinical care of a patient with SLE would benefit from the use of one of these drugs, the drug should not be withheld.

DNA, deoxyribonucleic acid; EBV, Epstein-Barr virus.
inflammation. Silica has also been proposed as a potential pathogenic factor in SLE on the basis of its known capacity to serve as an adjuvant.13

Two well-described triggers of lupus, ultraviolet light and certain drugs, are likely to promote lupus pathogenesis through their effects on DNA. Ultraviolet (UV) light has many effects on skin cells including induction of DNA breaks that might alter gene expression or lead to apoptotic or necrotic cell death. Even in the absence of cell death, DNA breaks or prolonged maintenance of DNA-protein cross-links might provide either an adjuvant or antigenic stimulus to the immune system. Altered DNA methylation has been proposed as a likely mechanism of drug-induced lupus.14 In the case of hydralazine, the drug inhibits extra-cellular signal-regulated kinase (ERK) pathway signaling, resulting in decreased expression of DNA methyltransferase 1 (DNMT1) and DNMT3a, two enzymes that mediate DNA methylation. Altered DNA methylation modifies gene expression and might also expose potential ligands for TLR-mediated immune system activation.

**INNATE IMMUNE SYSTEM ACTIVATION IN SYSTEMIC LUPUS ERYTHEMATOSUS**

Arguably the most significant recent advance that has improved understanding of the pathogenesis of SLE and other autoimmune and inflammatory diseases is the discovery of the TLRs and, more generally, the elucidation of the central role of innate immune system activation in regulation of the adaptive immune response, inflammation, and tissue repair. The TLRs are composed of an ectodomain with leucine-rich repeats, a transmembrane domain, and a cytoplasmic domain that associates with adaptor molecules that initiate signaling, resulting in activation of members of the IFN-regulatory factor family, NFκB, and members of the MAP kinase family. TLR-independent innate immune activation is initiated by intracytoplasmic nucleic acid sensors such as RIG-I and MDA5, which use distinct adaptors to trigger signal transduction through IRFs and the NFκB pathway. Over the past decade studies of nucleic acid-responsive TLRs and TLR-independent nucleic acid sensors that result in production of type I IFNs and other proinflammatory mediators have identified a direct pathogenic role for nucleic acids and nucleic acid–containing immune complexes as mediators of immune dysfunction in lupus. These studies complement the demonstration of broad expression of type I IFN-inducible genes in peripheral blood cells of lupus patients, referred to as the “IFN signature.” Together, the data point to the innate immune response and its products as significant factors in lupus pathogenesis. Table 79-3 describes many of the key components of the TLR-dependent and TLR-independent immune system pathways that are important in activation and control of lupus immune responses.

The endosomal TLRs, particularly TLR7, reactive with single-stranded RNA, and TLR9, reactive with unmethylated CpG-rich DNA, have been associated with lupus pathogenesis on the basis of in vitro studies that demonstrate activation of those TLRs by immune complexes that gain access to their intracytoplasmic compartment with the help of Fc receptors that bind the Fc fragment of immunoglobulin in the complex.16-40 Clinical data demonstrating association of autoantibodies with specificity for RNA-binding proteins (such as Ro, La, Sm, and RNP) with high expression of IFN-induced genes in patient peripheral blood cells point to a significant role for RNA-containing
immune complexes in innate immune activation and IFN production. Associated neutrophil-derived proteins including high mobility group box 1 (HMGB1) and the cathelicidin protein LL37 can facilitate access of those immune complexes to the TLR-containing endosome. Lupus mice made deficient in one or another TLR have supported the role of TLR7 in generating type I IFN, autoantibodies specific for RNA-binding proteins and disease. TLR9 activation can also generate IFN-α, but the findings from TLR9-deficient MRL/lpr mice have been somewhat confusing because those mice show more severe disease, suggesting that the activation of TLR9 might actually be protective. It appears that TLR9 activation can regulate the TLR7 pathway, resulting in reduced production of RNA-related autoantibodies. In addition to demonstrating the important role of the endosomal TLRs in lupus pathogenesis, the data link TLR pathway activation with production of particular autoantibody specificities. These significant advances in characterization of the role of TLRs and the innate immune response in lupus disease have suggested new therapeutic approaches that are currently being pursued.

The elucidation of the mechanisms of induction of type I IFN, with plasmacytoid dendritic cells (pDCs) the major producers, have promoted studies that identified roles for additional cell types in amplification of this pathway. Platelets, largely ignored in studies of SLE pathogenesis, were shown to promote IFN production by pDCs through signals mediated by CD40 ligand (CD154) on activated platelets and CD40 on the pDCs. One of the

**Figure 79-2** Endogenous stimuli for innate immune system activation. Endogenous molecules released by dying cells such as high mobility group box 1 (HMGB1) and β-defensins are recognized by cell surface Toll-like receptors (TLRs) and induce production of inflammatory mediators. Self-DNA or self-RNA, either in association with HMGB1 or LL37 or in the form of immune complexes that bind Fc receptors, is internalized into endosomal TLRs and transducer signals that result in transcription of type I interferon or proinflammatory cytokines. Self-nucleic acids can also be recognized by cytoplasmic sensors and trigger gene expression that contributes to immune activation, autoimmunity, and inflammation. DNA, deoxyribonucleic acid; ECM, extracellular matrix; Hsp, heat shock protein; LDL, low-density lipoprotein; RAGE, receptor for advanced glycation end products; RNA, ribonucleic acid; TBK1, TANK-binding kinase 1. (From Kawai T, Akira S: The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors, Nat Immunol 11:373–384, 2010.)
initial microarray data sets demonstrating an IFN gene expression signature in peripheral blood cells of pediatric lupus patients also detected expression of genes that are typically expressed in granulocytes. Recent studies from several groups have extended this observation and have documented low-density granulocytes in the circulation of lupus patients.

Neutrophil extracellular traps (NETs) have gained attention as potential mediators of innate immune system activation. The NETs derive from a discrete cellular process, termed NETosis, in which aggregates of chromatin including DNA and its associated histones, HMGB1, LL37, elastase, and myeloperoxidase, are systematically extruded from nuclei into the extracellular environment. Generation of the NETs can be induced by interaction of neutrophils with vascular endothelial cells, activated platelets, or various cytokines. Recent data also implicate nucleosomes and RNA-containing immune complexes in the induction of NETosis. Although NETs have only recently been linked to SLE, they present an intriguing and possibly significant mechanism that might account for observed alterations in immune function. They have the capacity to induce production of type I IFN by pDCs, serve as a source of relevant self-antigens for presentation to T lymphocytes, and mediate vascular damage and thrombosis.

As described earlier, the elucidation of a significant role for the nucleic acid–sensing TLRs in recent years has drawn attention to the important role of autoantigen in driving disease in lupus. The contribution of apoptotic debris, either through increased apoptosis and/or impaired clearance of apoptotic cells, has been assumed for a number of years to be at least one likely mechanism that drives immune system activation and focuses its specificity on nucleic acids and their associated proteins. Apoptotic blebs are enriched in the typical targets of autoantibodies such as Ro and Sm. Recent attention on so-called microparticles, small membrane-enclosed particles released by activated and dying cells, has generated data demonstrating binding of those particles by some autoantibodies in lupus plasma. A combination of increased generation of nuclear material, perhaps modified, and impaired clearance may be important mechanisms of lupus immunopathogenesis. The role of C1q and complement components in the clearance of apoptotic debris has been suggested as an explanation for the strong association of C1q, C2, and C4 deficiencies with SLE. A second role for complement components that likely reduces the pathogenicity of immune complexes is the capacity of complement to solubilize immune complexes, reducing the likelihood that those complexes would deposit in tissue and cause damage.

ADAPTIVE IMMUNE SYSTEM ALTERATIONS IN SYSTEMIC LUPUS ERYTHEMATOSUS

T cell function has been carefully studied in SLE patients for several decades, and deficiencies or alterations in signaling pathways, production of cytokines, cell proliferation, and regulatory functions have been documented. CD4+ T cells are viewed as a requirement for development of lupus on the basis of their essential role in providing helper signals that drive differentiation of B cells to autoantibody-producing cells. Although some in vitro experiments have supported the capacity of cytokines such as IL-21 and B cell activating factor (BAFF)/B lymphocyte stimulator (BlyS) and TLR ligands to mediate antibody production by B cells, T cells are recognized as the most efficient drivers of B cell differentiation.

The activation status and signaling mechanisms of lupus T cells are distinct from those of T cells from healthy individuals. Among the alterations in function is decreased proliferation in response to self-non–T cells or allogeneic-non–T cells or in response to presentation of soluble antigens. In contrast, T cell proliferation induced by direct activation of the T cell antigen receptor is at least as strong as observed in control cells. Lupus T cells readily express CD40 ligand (CD154) after activation and maintain expression of that important co-stimulatory molecule longer than control T cells, leading to augmented help for activation and differentiation of B cells exposed to those T cells. Another long-standing observation is that lupus T cells produce less IL-2 than control T cells, perhaps one factor that contributes to impaired generation of IL-2-dependent T regulatory cells. The molecular basis of the altered T cell activation in lupus patients is complex, but at least one factor might be the observed substitution of the T cell receptor zeta chain with the common gamma chain that is a component of the Fc receptor signaling machinery. Correction of this defect can normalize T cell signaling and IL-2 production. Augmented calcium responses following T cell receptor ligation have been observed, and hyperpolarization of mitochondria in T cells has been associated with altered T cell activation and function.

An interesting observation from studies of lupus T cells might reflect more global epigenetic alterations to the lupus genome that impact autoimmunity. Treatment of mouse and human T cells with 5-azacytidine resulted in increased expression of the lymphocyte function antigen-1 (LFA1; CD11a) adhesion molecule and increased proliferative responses to self-non–T cells. Lupus T cells studied ex vivo show hypomethylation of CG-rich areas of the genome, and a recent report of genome methylation in identical twins discordant for lupus demonstrated relatively less methylation of the twin with active lupus. Multiple mechanisms have been postulated to contribute to DNA demethylation in lupus lymphocytes including increased expression of growth arrest and DNA damage-induced 45alpha (GADD45alpha), a protein that removes methyl groups from DNA, decreased expression of DNAMT1, and ERK pathway signaling.

Generalized lymphopenia is typical of SLE, but expansions of specific T cell populations have been described. The recently documented T follicular helper population, characterized by expression of ICOS, CXCR5, and Bcl6 and by production of IL-21, mediates important signals that promote differentiation of autoantigen-specific B cells. A population of CD8+ cells with a memory phenotype is associated with poor prognosis, possibly based on tissue damage mediated by those cells.

The regulation and function of T regulatory cells (Tregs), with capacity to suppress immune responses, and Th17 cells, which promote inflammation by production of IL-17, have been intensively studied in recent years. Some studies of
lupus patients do show a relative depletion of Tregs and increased Th17 cells and IL-17.\textsuperscript{63} The functional impact of those alterations in human lupus is still not clear.

Cytokine production is altered in SLE, with decreased production of IL-2 a characteristic feature of patient T cells, as it is of individuals carrying the HLA 8.1 haplotype.\textsuperscript{4} Although this deficiency in IL-2 was initially linked to the often poor proliferative responses of lupus T cells stimulated with autologous or allogeneic T cells or soluble antigen, the recognition that IL-2 is important for maintenance of T regulatory cells suggests another mechanism through which impaired production of IL-2 might contribute to immune system activation and autoimmunity.\textsuperscript{62}

B cell regulation is also impaired in SLE, contributing to differentiation to production of autoantibodies and cytokines. The activated phenotype of SLE B cells has the potential to promote efficient presentation of specific self-antigens to T cells. It has not been entirely clear whether altered B cell function is strictly secondary to increased availability of T cell help, B cell survival, proliferation and differentiation factors, and signaling through TLRs or whether primary B cell dysfunction also contributes to autoimmunity. Recent studies characterizing the B cell repertoire using single-cell polymerase chain reaction and cloning of individual immunoglobulin transcripts will permit improved understanding of the role of altered B cell tolerance mechanisms in the bone marrow versus secondary effects of T cell and cytokine help in generating self-reactive antibody specificities. BAFF/BllyS, IL-10, and IL-21 are among the candidate therapeutic targets that could modify the B cell differentiation program, a suggestion supported by clinical studies of BAFF/BllyS blockade. An additional approach toward improved regulation of B cell function is suggested by recent data demonstrating an association between vitamin D deficiency and presence of antinuclear antibodies in healthy individuals and increased B cell activation and serum type I IFN activity in vitamin D–deficient SLE patients.\textsuperscript{63}

The baseline activation status of lupus B cells, as well as their response to antigen stimulation, is altered when compared with B cells from healthy individuals, and current studies are exploring the functional implications of lupus-associated genetic variants in several kinases, phosphatases, and adaptor molecules such as BLK, BANK, and PTPN22 for B cell function. Studies of the risk variant of PTPN22 in B cells from healthy subjects point to the influence of that lupus-associated variant on impaired counterselection of self-reactive B cells.\textsuperscript{64} Lyn deficiency in mice is associated with a lupus-like phenotype, and similar deficiencies in the Lyn kinase have been described in lupus B cells.\textsuperscript{65} Polymorphisms in the gene encoding Lyn studied in European-descent individuals could potentially impact B cell signaling and autoimmunity.\textsuperscript{66} The gene encoding the inhibitory Fc receptor, FCGR2B, is polymorphic, and several variants have been associated with SLE.\textsuperscript{67,68} SLE B cells show decreased expression of this FcR and altered cytokine production when engaged by immune complexes. Interestingly, the inhibitory capacity of FcγRIIB can be exploited using an engineered bifunctional antibody that coligates CD19 and FcγRIIB.\textsuperscript{69} This antibody suppressed B cell functions in a mouse engrafted with human B cells.

Current investigations of antibody-producing cells are defining several categories of B lineage cells that contribute to autoimmunity and disease. Long-lived plasma cells that are maintained by chemokines and stromal cell products in protective niches in the bone marrow are proposed as sources of those lupus autoantibodies such as anti-Sm and anti-Ro that are chronically maintained at fairly constant levels. Those antibodies have been refractory to modulation by immunosuppressive or B cell depletion therapy.\textsuperscript{70} In contrast, circulating preplasma cells or plasmablasts are sources of anti-dsDNA antibodies that fluctuate in some patients in association with variations in disease activity and might be more amenable to anti–B cell therapy.\textsuperscript{71}

The T and B cell alterations that have been described in studies of lupus immune function are undoubtedly a reflection of multiple genetic variations that play out in altering threshold for cell activation and efficiency of cell signaling. They are also the result of interactions among those lymphocytes, along with antigen-presenting cells and their products, that over time amplify the overall level of immune activation. Increased production of type I IFN and presence of antinuclear antibodies in healthy family members of lupus patients likely reflect the consequences of genetic variations that are among the factors that predispose to SLE. When the circuits of immune activation are amplified by environmental triggers, conditions are in place to generate the characteristic T and B cell autoimmunity that is required for tissue damage and disease.

**AUTOIMMUNITY IN SYSTEMIC LUPUS ERYTHEMATOSUS**

Autoantibodies are traditionally viewed as essential mediators of pathology in SLE, particularly when they are in the form of immune complexes. Virtually all lupus patients demonstrate a positive antinuclear antibody test, and the majority of patients have one or more of the characteristic autoantibody specificities that have been associated with lupus (see also Chapter 55). Among those, anti-dsDNA and anti-Smith (Sm) are most specific for SLE. Interestingly, those specificities have been linked functionally in murine studies that demonstrate that dsDNA can bind to a peptide comprising amino acids 83-119 of the SmD1 protein, and T cells specific for that peptide provide help for production of anti-dsDNA antibody.\textsuperscript{72,73} Anti-Ro, anti-La, and anti-RNP antibodies are characteristic of SLE but are also seen in other systemic autoimmune diseases. These characteristic autoantibodies in SLE can be categorized in relation to their targets including DNA and DNA-binding proteins, typically aggregated in nucleosomes that contain histones; RNA and RNA-associated proteins, typically aggregated in cytoplasmic or nuclear ribonucleoprotein particles; phospholipids exposed in plasma membranes and associated proteins such as β2-glycoprotein I; and cell membrane proteins, typically those expressed on blood cells. Some of these self-antigens typically targeted by lupus autoantibodies are most likely accessed by antigen-presenting cells in the form of cell membrane–enveloped blebs derived from apoptotic cells or microparticles, small aggregates of cellular material derived from cells and generated after
flipping of phosphatidylserine to the outer aspect of the cell membrane in the setting of cell activation or apoptosis. Although some patients who present with a clinical picture characteristic of SLE do not have significant titers of these autoantibodies, it is likely that those patients have undefined autoimmunity targeted at unspecified antigens.

The pathogenic antibodies in SLE tend to be those produced by cells that are far along in the B cell differentiation process, either preplasma cells or plasma cells, and have undergone immunoglobulin class switching, a process that is driven by CD4⁺ T helper cells and in some cases by TLR ligands together with B cell differentiation factors such as IL-21 and BlyS/BAFF. It is not possible to define the earliest step in the generation of pathogenic autoantibodies, and in fact it is probable that the generation of autoimmunity can begin in any number of ways. Presentation of self-antigen by an excessively activated antigen-presenting cell, stochastic activation of low-avidity self-reactive T cells that are present in healthy individuals but have the potential to expand and drive differentiation of B cells with broad specificity for self-antigens, or direct activation of self-reactive B cells in the presence of activating TLR ligands such as those provided by demethylated CpG-rich DNA from bacteria or viruses or self-derived nuclear debris might all be starting points for development of pathogenic autoimmunity. The process might begin with production of type I IFN by a virus-infected cell or a cell with impaired degradation of intracellular nucleic acids, altering the threshold for activation of the immune system by subsequent exposures to self-antigens. Less important than the precise starting point is the observation that autoimmunity develops and builds over time, with presence of autoantibodies observed more than 5 years before clinical manifestations of disease appear and the range of specificities of antigens targeted expanding over time. The data from analysis of prevalence of several lupus autoantibodies among sera collected from members of the military who later developed SLE demonstrate an intriguing sequence, with anti-Ro antibodies typically presenting earliest, anti-dsDNA antibodies appearing several years later, and anti-Sm, the antibody specificity most specific for a diagnosis of SLE, appearing approximately at the time of clinical diagnosis. This pattern, along with studies from murine models demonstrating determinant spreading of autoantibody specificities not only to distinct proteins within a nucleic acid–protein particle but also to other self-antigen targets, provides clues to lupus pathogenesis that are not yet understood but hold promise for future breakthroughs in elucidating important pathogenic mechanisms. Additional autoantibody specificities that are associated with lupus activity and with proliferative lupus nephritis and are thought to be pathogenic include those reactive with C1q. Anti-C1q antibodies appear to recognize neoeptopes of C1q when it is bound to early apoptotic cells.

A shift from a predominant polyclonal IgM picture toward polyclonal IgG, driven by T cell help and cytokines, occurs over time in most patients with lupus and with development of tissue pathology and damage. In fact, some IgM antibodies that have self-reactivity are viewed as protective, with the switch from IgM to IgG or IgA representing an important point of altered immune regulation that contributes to the immunopathogenesis of SLE. Some of these IgM natural antibodies react with apoptotic cells and actually inhibit their activation through TLRs.

In addition to immunoglobulin class, with class-switched antibodies better able to access extravascular spaces compared with IgM antibodies, amino acid sequence and charge of the antigen-binding site can influence pathogenicity. Arginines in the CDR3 region of anti-dsDNA antibodies are characteristic and influence binding to their DNA target. Molecular mimicry is a concept that has been applied to reactivity of antibodies specific for a microbial protein that also react with self-antigens, but the concept can also be applied to those antibodies that unexpectedly bind to two distinct self-antigens. For example, some anti-dsDNA antibodies were found to bind to a peptide that is a feature of glutamate receptors on central nervous system neurons. Beyond features of Ig class and antigen-binding site that influence access to tissue and affinity of binding, glycosylation and complement-fixing capacity are important determinants of the antibody’s capacity to bind to Fc receptors and promote complement activation, resulting in target cell death or inflammation.

The role of SLE autoantibodies in the pathogenesis of the disease has traditionally focused on the deposition of immune complexes in skin, renal glomerulus, and other sites of tissue injury, along with a potential contribution of direct targeting of antibodies to local or “planted” antigens. But in recent years, with the recognition that nucleic acid–containing immune complexes can directly induce cell signaling and new gene transcription after accessing endosomal TLRs, an additional important pathogenic role for autoantibodies as immune modulators has been defined.

Particular lupus autoantibody specificities have been associated with specific clinical manifestations of disease (Table 79-4). Perhaps the best developed characterization of the mechanisms by which an autoantibody mediates disease is the role of maternal-derived anti-Ro antibody in the neonatal lupus syndrome. After transplacental transfer of anti-Ro antibody, RNA-containing immune complexes are proposed to form, activating the production of cytokines that contribute to fibrosis of the cardiac conduction system. This scenario, similar to that which mediates production of type I IFN by pDCs, identifies nucleic acid–autoantibody immune complexes as regulators of cytokine production that results in tissue damage. Anti-Ro and anti-La antibodies, although common to several autoimmune diseases, are also characteristic in association with sicca symptoms, subacute cutaneous lupus, and, as noted, in neonatal lupus. Anti-RNP antibodies are seen in both SLE and mixed connective tissue disease.

Antiphospholipid antibodies, with lupus anticoagulant the most informative antibody profile, contribute to placental damage and vascular thrombosis, as well as thrombocytopenia in the antiphospholipid syndrome and in some patients with lupus. Recent studies in a murine model indicate that tissue damage and thrombosis depend on activation of the complement system by the antiphospholipid antibodies. Similar mechanisms might be relevant in forms of lupus renal disease characterized by microangiopathy.
MECHANISMS OF TARGET ORGAN DAMAGE

Clinical disease is ultimately a reflection of tissue damage mediated by the inflammatory sequelae of the described autoimmunity and immune system activation, along with an exaggerated or aberrant repair response. The classic view of the mechanisms that result in tissue damage involves activation of the complement system by immune complexes deposited in tissue, as well as release of products of phagocytes including enzymes released from neutrophil granules and reactive oxygen intermediates from macrophages. Recent studies in mouse and human systems involving extensive analysis of renal infiltrating cells at various points in the disease process have identified a monocyte population that has undergone differentiation to generate a functional phenotype that mediates what is apparently uncontrolled tissue repair, contributing to sclerosis and organ dysfunction. Studies in a murine model have shown that IFN-\(\alpha\) can also contribute to development of crescents in lupus kidneys.

The role of the vasculature as a significant target organ contributing to disease in lupus has once again come to the fore, after several decades that have focused predominantly on the immune system and the contributions of autoantibodies to pathology. Altered structure and function of both venous and arterial vessels have been noted for many years including the periarteriolar concentric onion skinning seen in spleen, microangiopathy and associated microthrombi seen in some organs, and the endothelial dysfunction that has been associated with premature atherosclerosis in lupus patients. Recent studies have focused on the potential role of type I IFN on endothelial cells and endothelial cell progenitor cells and have postulated that increased production of IFN is at least one contributor to impaired vascular repair. A role for granulocytes and proinflammatory lipids has also been proposed. Characterization of the micropathology of the vasculature in Degos disease, a syndrome with some common pathologic features with SLE, has drawn attention to the association of activation of the IFN pathway with vascular sclerosis and endothelial damage. A possible direct contribution of type I IFN to vascular injury supplements the previously described role of complement activation products C3a and C5a and expression of endothelial cell surface adhesion molecules E-selectin, vascular cell adhesion molecule 1 (VCAM-1), and intercellular adhesion molecule 1 (ICAM-1) in lupus flares.

SUMMARY

Considerable gains in understanding mechanisms of lupus pathogenesis have established the knowledge base that will underlie future advances in development of targeted therapies that will improve patient outcomes. The recognition of a central role for innate immune system activation including the role of type I IFN in lupus pathogenesis has been a major advance in recent years. Rapid progress in characterizing the genetic variants that are associated with a diagnosis of lupus has provided strong support for alterations in molecular pathways that regulate nucleic acid degradation, TLR signaling and lymphocyte activation thresholds, and signaling efficiency as important mechanisms that determine disease susceptibility. The products of the immune system including autoantibodies and their immune complexes, cytokines, and complement components, along with proinflammatory mediators and reactive oxygen products released from neutrophils and macrophages, remain key mediators of tissue damage. But an additional pathogenic role for nucleic acid–containing immune complexes as immune modulators through their important capacity to access and activate endosomal TLRs is an important new

### Table 79-4 Correlation among Clinical Manifestations of Systemic Lupus Erythematosus and Autoantibodies, Immune Complexes, and T Cells

<table>
<thead>
<tr>
<th>Manifestation</th>
<th>Autoantibodies</th>
<th>Immune Complexes</th>
<th>T Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nephritis</td>
<td>Anti-dsDNA, Anti-Ro, Anti-C1q, Ids 16/6, 3I and GN2</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Arthritis</td>
<td>Anti-Ro, Anti-dsDNA, Id 16/6</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dermatitis</td>
<td>Anti-Ro</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Vasculitis</td>
<td>Anti-Ro</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Central nervous system</td>
<td>Antiribosomal P, Antineuronal</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Hematologic</td>
<td>Antilymphocyte</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Lymphopenia</td>
<td>Antierythrocyte, Antiplatelet</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Hemolysis</td>
<td>Antiphospholipid</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Thrombocytopenia</td>
<td>Antiphospholipid</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Clotting</td>
<td>Anti-Ro</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Fetal loss</td>
<td>Anti-Ro</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Neonatal lupus</td>
<td>Anti-Ro</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Sicca syndrome</td>
<td>Anti-RNP without other autoantibody except antinuclear antibody</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Mild disease</td>
<td></td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>
concept that can guide future therapeutic approaches. Targeting the components of innate immune system activation that affect the TLR pathway along with modulation of T cell help for B cell differentiation to autoantibody-producing cells would appear to be a rational therapeutic strategy in light of our current understanding of pathogenic mechanisms.

References


The references for this chapter can also be found on www.expertconsult.com.
Clinical Features of Systemic Lupus Erythematosus

MARIA DALL’ERA • DAVID WOFSY

KEY POINTS
Systemic lupus erythematosus (SLE) is a multisystem autoimmune disease characterized by a relapsing-remitting course and a highly variable prognosis.

It is characterized by the production of a broad array of autoantibodies, with antinuclear antibodies having the greatest sensitivity for the diagnosis and anti–double-stranded DNA and anti-Smith antibodies having the greatest specificity.

Women of childbearing age and African-American, Asian, and Hispanic populations have the highest prevalence of disease.

Constitutional symptoms, rash, mucosal ulcers, inflammatory polyarthritis, photosensitivity, and serositis are the most common clinical features of the disease.

Lupus nephritis is the most common of the potentially life-threatening manifestations.

Atherosclerosis, a complication of long-standing SLE, requires aggressive risk factor modification.

CLASSIFICATION CRITERIA
Systemic lupus erythematosus (SLE) is the prototypic systemic autoimmune disease characterized by diverse multisystem involvement and the production of an array of autoantibodies. Clinical features in individual patients can be quite variable, ranging from mild joint and skin involvement to severe life-threatening internal organ disease. Criteria for the classification of SLE were initially developed by the American College of Rheumatology (ACR) in 1971, revised in 1982, and revised for a second time in 1997 (Table 80-1). A person must fulfill 4 of 11 criteria to be classified as SLE; all other reasonable diagnoses having been excluded. A patient does not have to manifest all 4 criteria simultaneously; the required 4 of 11 criteria can be fulfilled over a period of weeks or years. The ACR criteria were developed as a means of classifying patients with SLE for the purpose of inclusion in clinical and epidemiologic studies. In clinical practice, these criteria are often cited to support a diagnosis of SLE. However, it should be emphasized that fulfillment of these classification criteria is not an absolute requirement for diagnosis. Rather, diagnosis typically rests on the judgment of an experienced clinician who recognizes a characteristic constellation of symptoms and signs in the setting of supportive serologic studies, after exclusion of alternative differential diagnostic possibilities. Recently, a concerted effort has been made to further revise the classification criteria, for example, to make lupus nephritis a “stand-alone” criterion, to increase the weight of neurologic manifestations, and/or to add a low-complement criterion. These proposed changes to the classification criteria are currently in development.

EPIDEMIOLOGY
Prevalence and incidence rates of SLE vary widely in the literature. Reported prevalence frequencies range from 20 to 240 per 100,000 persons, and reported incidence rates range from 1 to 10 per 100,000 person-years. This variation is partly due to methodological differences between studies (e.g., different case definitions of SLE and methods of case ascertainment). One study from Rochester, Minnesota, determined that the incidence of SLE increased almost fourfold between the time periods of 1950 to 1979 and 1980 to 1992. This increase in reported incidence may reflect a combination of factors, including an actual increase in disease, changes in population demographics, more widespread case-finding efforts, and detection of milder cases.

Prevalence and incidence of SLE vary across gender, geographic regions, and racial/ethnic groups. SLE demonstrates a striking female predominance with a peak incidence during reproductive years. The degree of female predominance varies with age. The female-to-male ratio is 10 to 15:1 in adults, 3:1 in older-onset SLE, and 8:1 in children. The prevalence of SLE is believed to be approximately three- to fourfold higher in African-American, Asian, and Hispanic populations compared with white populations. SLE is rare among blacks in Africa.

Most SLE patients present with their disease between 15 and 64 years of age. Patients with pediatric-onset SLE (≤16 years old) are more likely to be African-American than those with later-onset SLE. SLE tends to be more severe in men and in pediatric patients. Late-onset SLE (>50 years old) is characterized by a more insidious onset with a higher occurrence of serositis and pulmonary involvement and a lower incidence of malar rash, photosensitivity, alopecia, Raynaud’s phenomenon, neuropsychiatric disease, and nephritis.

CLINICAL FEATURES
Systemic lupus erythematosus has protean clinical manifestations that may differ dramatically from patient to patient. Just as the signs and symptoms of SLE vary widely among patients, so too does the severity of disease. Some patients with lupus will have relatively mild disease that never threatens a life-sustaining organ; other patients will progress rapidly to life-threatening disease.

The great variability in the expression and severity of SLE constitutes a major challenge to accurate diagnosis.
Table 80-1  1997 Update of the 1982 Revised American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus*

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malar rash</td>
<td>Fixed erythema, flat or raised, over the malar eminences, sparing the nasolabial folds</td>
</tr>
<tr>
<td>Discoid rash</td>
<td>Erythematous raised patches with adherent keratotic scale and follicular plugging; atrophic scarring may occur in older lesions</td>
</tr>
<tr>
<td>Photosensitivity</td>
<td>Skin rash as a result of unusual reaction to sunlight, by patient history or physician observation</td>
</tr>
<tr>
<td>Oral ulcers</td>
<td>Oral or nasopharyngeal ulceration, usually painless, observed by a physician</td>
</tr>
<tr>
<td>Arthritis</td>
<td>Nonerosive arthritis involving two or more peripheral joints, characterized by tenderness, swelling, or effusion</td>
</tr>
<tr>
<td>Serositis</td>
<td>a. Pleuritis-convincing history of pleuritic chest pain or rub heard by a physician or evidence of pleural effusions or</td>
</tr>
<tr>
<td></td>
<td>b. Pericarditis-documented by electrocardiogram or rub or evidence of pericardial effusion</td>
</tr>
<tr>
<td>Renal disorder</td>
<td>a. Persistent proteinuria &gt;0.5 g/day, &gt;3+ if quantification not performed or</td>
</tr>
<tr>
<td></td>
<td>b. Cellular casts: may be red blood cell, hemoglobin, granular tubular, or mixed</td>
</tr>
<tr>
<td>Neurologic disorder</td>
<td>a. Seizures: in the absence of offending drugs or known metabolic derangements (e.g., uremia, acidosis, electrolyte imbalance) or</td>
</tr>
<tr>
<td></td>
<td>b. Psychosis: in the absence of offending drugs or known metabolic derangements (e.g., uremia, acidosis, electrolyte imbalance)</td>
</tr>
<tr>
<td>Hematologic disorder</td>
<td>a. Hemolytic anemia with reticulocytosis or</td>
</tr>
<tr>
<td></td>
<td>b. Leukopenia &lt;4000/mm³ or</td>
</tr>
<tr>
<td></td>
<td>c. Lymphopenia &lt;1500/mm³ or</td>
</tr>
<tr>
<td>Immunologic disorder</td>
<td>a. Anti-DNA: antibody to native DNA in abnormal titer or</td>
</tr>
<tr>
<td></td>
<td>b. Anti-Smith: presence of antibody to Sm nuclear antigen or</td>
</tr>
<tr>
<td></td>
<td>c. Positive finding of antiphospholipid antibodies based on (1) abnormal serum concentration of IgG or IgM</td>
</tr>
<tr>
<td></td>
<td>(2) positive test result for lupus anticoagulant using a standard method, or (3) false-positive serologic test for syphilis known to</td>
</tr>
<tr>
<td></td>
<td>be positive for at least 6 mo and confirmed by Treponema pallidum immobilization or fluorescent treponemal antibody absorption test</td>
</tr>
<tr>
<td>Positive antinuclear antibody</td>
<td>An abnormal titer of antinuclear antibody by immunofluorescence or an equivalent assay at any point in time</td>
</tr>
<tr>
<td></td>
<td>and in the absence of drugs known to be associated with drug-induced lupus syndromes</td>
</tr>
</tbody>
</table>

*The presence of four or more criteria is required for systemic lupus erythematosus classification. Exclude all other reasonable diagnoses.

Table 80-2  Frequencies of Various Manifestations of Systemic Lupus Erythematosus*

<table>
<thead>
<tr>
<th>Manifestation</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constitutional symptoms (fatigue, fever, weight loss)</td>
<td>90%-95%</td>
</tr>
<tr>
<td>Mucocutaneous involvement (malar rash, alopecia, mucosal ulcers, discoid lesions, etc.)</td>
<td>80%-90%</td>
</tr>
<tr>
<td>Musculoskeletal involvement (arthritis/arthritis, avascular necrosis, myositis, etc.)</td>
<td>80%-90%</td>
</tr>
<tr>
<td>Serositis (pleuritis, pericarditis, peritonitis)</td>
<td>50%-70%</td>
</tr>
<tr>
<td>Glomerulonephritis</td>
<td>40%-60%</td>
</tr>
<tr>
<td>Neuropsychiatric involvement (cognitive impairment, depression, psychosis, seizures, stroke, demyelinating syndromes, peripheral neuropathy, etc.)</td>
<td>40%-60%</td>
</tr>
<tr>
<td>Autoimmune cytopenia (anemia, thrombocytopenia)</td>
<td>20%-30%</td>
</tr>
</tbody>
</table>

*Systemic lupus erythematosus is a heterogeneous disease that can affect virtually any organ system in variable ways.

Although it is difficult to pinpoint their precise frequencies, it is clear that the most common presenting manifestations are constitutional symptoms (fever, fatigue, and/or weight loss), cutaneous manifestations (e.g., malar rash), and articular manifestations (arthritis and/or arthralgia). Each of these manifestations appears to be present in at least 50% of lupus patients at the time of diagnosis. The other clinical features of SLE are much less likely to be presenting manifestations, although virtually any of them might be the first clue to the correct diagnosis. More commonly, these manifestations appear over time as the disease evolves. Taken together, various descriptive studies in the literature1,8-12 support a cumulative descriptive frequency of symptoms and signs that is summarized in Table 80-2.

Mucocutaneous Involvement

Mucocutaneous involvement is very common in SLE. Gilliam and colleagues have categorized cutaneous lupus erythematosus (LE) lesions as “lupus specific” or “lupus non-specific” based on the histopathologic finding of interface dermatitis13,14 (Table 80-3). The presence of lupus-specific lesions confirms the diagnosis of cutaneous LE; lupus non-specific lesions can occur in diseases other than lupus. Lupus-specific lesions are further subdivided into acute lupus erythematosus (ACLE), subacute lupus erythematosus (S克莱), and chronic cutaneous lupus erythematosus (CCLE) lesions, based on additional clinical and histopathologic information. Discoid lupus is the most common subtype of CCLE. S克莱 and CCLE can occur as distinct isolated entities or as one of several manifestations of SLE. The risk of SLE varies with each cutaneous subset. One study of 161 patients with lupus-specific lesions showed that the classification criteria for SLE were present in 72% of patients with ACLE, 58% with S克莱, 28% with any form of discoid lupus, and 6% with localized discoid lupus confined to head and neck. Patients commonly displayed more than one type of cutaneous lesion.15

Acute Cutaneous Lupus Erythematosus

Acute cutaneous lupus erythematosus (ACLE) lesions can be localized or generalized. The hallmark feature of ACLE is localized to the malar region (“butterfly rash”) and is characterized by confluent, macular or papular erythema lasting days to weeks that occurs symmetrically on the cheeks and bridge of the nose, sparing the nasolabial folds (Figure 80-1). Induration and scaling may occur. The malar rash of SLE can be mimicked by a variety of other facial
Table 80-3  Gilliam Classification of Skin Lesions Associated with Lupus

A. Lupus erythematosus (LE)-specific skin lesions
1. Acute cutaneous LE (ACLE)
   Localized ACLE
   Generalized ACLE
2. Subacute cutaneous LE (SCLE)
   Annular SCLE
   Papulosquamous SCLE
3. Chronic cutaneous LE (CCLE)
   Classical discoid LE (DLE): (a) localized; (b) generalized
   Hypertrophic DLE/verrucous DLE
   Lupus panniculitis/profundus
   Mucosal DLE
   LE tumidus
   Chilblain LE
   Lichenoid DLE (DLE-lichen planus overlap)

B. LE-nonspecific skin lesions
1. Cutaneous vascular disease
   Vasculitis
   Leukocytoclastic vasculitis
   Polyarteritis nodosa–like
   Vasculopathy
   Degos disease–like
   Atrophie blanche–like
   Periungual telangiectasia
   Livedo reticularis
   Thrombophlebitis
   Raynaud’s phenomenon
   Atrophe blanche–like
   Periungal telangiectasia
   Livedo reticularis
   Thrombophlebitis
   Raynaud’s phenomenon
   Erythromelalgia
2. Nonscarring alopecia
   "Lupus hair"
   Telogen effluvium
   Alopecia areata
3. Sclerodactyly
4. Rheumatoid nodules
5. Calcinosis cutis
6. LE-nonspecific bullous lesions
7. Urticaria
8. Papulonodular mucinosis
9. Cutis laxa/anetoderma
10. Acanthosis nigricans (type B insulin resistance)
11. Erythema multiforme (Rowell’s syndrome)
12. Leg ulcers
13. Lichen planus

Subacute Cutaneous Lupus Erythematosus

Subacute cutaneous lupus erythematosus (SCLE) is characterized by the presence of nonscarring, photosensitive lesions that can take one of two distinct forms: (1) papulosquamous lesions that resemble psoriasis, or (2) annular-polycyclic lesions with peripheral scale and central clearing (Figure 80-2). These two forms can occur concurrently in the same patient. SCLE has a predilection for the back, neck, shoulders, and extensor surfaces of the arms and usually spares the face. The lesions typically last for weeks to months and heal without scarring. Uncommonly, a severe TEN-like eruption can evolve from SCLE lesions after sun exposure. SCLE, particularly the annular subtype, is strongly associated with the presence of anti-SSA/Ro antibody. Several drugs are known to induce SCLE; angiotensin-converting enzyme inhibitors, terbinafine, hydrochlorothiazide, and calcium channel blockers are common culprits. Finally, SCLE has been implicated as a paraneoplastic syndrome.
Chronic Cutaneous Lupus Erythematosus

Chronic cutaneous lupus erythematosus (CCLE) refers to a variety of subtypes of photosensitive lesions that can lead to skin atrophy and scar and that may persist for several months. Discoid lupus (DLE) is the most common subtype of CCLE and is subdivided into localized discoid lupus (limited to head and neck) and generalized discoid (occurring above and below the neck) (Figures 80-3 and 80-4). The term “discoid” refers to the sharply demarcated disk-shaped appearance of the lesions. The lesions are raised, erythematous plaques with adherent scale that commonly occur on the scalp, face, and neck. The cheeks, nose, ears, and upper lip are classic locations. Typically, a raised, erythematous border denotes the actively expanding component. Follicular plugging is a characteristic finding. Left untreated, DLE can result in permanent alopecia and disfigurement. Squamous cell carcinoma has been described as a sequela of long-standing DLE; thus, active surveillance of known lesions and evaluation of changing lesions are critical. Other subtypes of DLE include mucosal DLE (described later) and hypertrophic LE. Hypertrophic LE consists of chronic, indurated lesions that are covered by hyperkeratotic, multilayered scales. These lesions can be a source of diagnostic confusion because they may visually and histologically resemble squamous cell carcinoma.

Other forms of CCLE include lupus panniculitis/profundus and chilblain lupus. Lupus panniculitis is a lobular panniculitis that has a predilection for the scalp, face, arms, buttocks, and thighs. When a cutaneous discoid lesion overlies the panniculitis, the entity is referred to as lupus profundus. Biopsy is often necessary to secure the diagnosis because reports have described T cell lymphoma mimicking panniculitis. However, biopsy should be performed carefully because the lesions have a tendency to break down. Lupus panniculitis is one of the few panniculitides that can occur above the waist. Lupus panniculitis is associated with low risk of concomitant SLE. Chilblain lupus manifests as tender, erythematous, or violaceous papules occurring on acral areas, especially fingers, toes, heels, nose, and ears. The lesions are brought on by cold, damp air.

Other Systemic Lupus Erythematosus Skin Lesions

SLE patients can develop lupus-nonspecific skin lesions such as cutaneous leukocytoclastic vasculitis, bullous lesions, periungual erythema, and livedo reticularis. Cutaneous leukocytoclastic vasculitis most commonly presents as palpable purpura on the lower extremities. Bullous lupus erythematosus is a rare cutaneous manifestation characterized by subepidermal vesiculobullous skin changes. It is manifested by a nonscarring bullous eruption (Figure 80-5). SLE may be associated with other bullous disorders such as bullous pemphigoid and dermatitis herpetiformis. The physical examination finding of periungual erythema represents dilation of the capillaries at the base of the nail. These capillaries can be visualized at the bedside with a dermatoscope or ophthalmoscope. Other disorders associated with periungal erythema include scleroderma and mixed connective tissue disease (MCTD). Unlike scleroderma and MCTD, SLE is not associated with capillary dropout. Livedo reticularis is characterized by an erythematous to violaceous reticular or net-like pattern of the skin. It is highly associated with the antiphospholipid antibody syndrome.

Photosensitivity

Photosensitivity occurs frequently in SLE. In one study, photoprovocation testing caused an abnormal skin reaction to ultraviolet A, ultraviolet B, or visible light in greater than 90% of lupus patients. Most abnormal skin reactions
Although a Bullous lupus erythematosus. These lesions are a rare and vermilion border, common locations being the hard palate, buccal mucosa, and can occur anywhere on the oral mucosa, the most frequent those present in lupus skin lesions include basal layer vasculopathic degeneration of keratinocytes, perivascular and periadnexal inflammation, follicular plugging, mucin deposition, and hyperkeratosis. These findings occur to different degrees in various lupus-specific skin lesions, with discoid lesions showing the most profound changes. In contrast, early ACLE lesions may have minimal histopathologic findings and only a sparse lymphocytic infiltrate.

Immunofluorescence demonstrates granular deposition of immunoglobulin and complement components along the dermal-epidermal junction. Immunoglobulin (Ig)G and IgM are the most common immunoglobulin subtypes deposited. Various complement components including C3, C1q, and the membrane attack complex have also been identified. Direct immunofluorescence (DIF) of nonlesional, sun-exposed skin is called the lupus band test. Although a positive lupus band test is often seen in patients with SLE, this may also be seen in patients with other rheumatic diseases, as well as in healthy people. In one study, 20% of healthy young adults had a positive DIF in sun-exposed skin. It is important to note that serologic testing with antinuclear antibody (ANA), anti–double-stranded DNA (anti-dsDNA), and anti-Smith (anti-SM) has largely supplanted use of the lupus band test in confirming the diagnosis of SLE. DIF of nonlesional, sun-protected skin is believed to be more specific for SLE.

Alopecia

Scarring alopecia is a common complication of discoid lupus. Scalp discoid lesions most frequently develop on the vertex and parietal areas. Nonscarring alopecia in SLE patients can take several forms. “Lupus hair” is characterized by short, irregularly sized hair at the frontal hairline and is associated with active systemic disease. Telogen effluvium manifests as diffuse hair thinning. Lastly, the incidence of alopecia areata (discrete areas of hair loss) is believed to be increased in SLE.

Mucosal Ulcers

SLE patients commonly develop nasal or oral lesions that represent the mucosal counterparts of cutaneous lupus. Acute oral lupus lesions present as red macules, palatal erythema or petechiae, erosions, or ulcerations. These lesions are usually painless. Subacute oral lesions are rare, and are characterized by well-demarcated, round, red patches. Oral discoid lesions present as painful, well-demarcated, round, red lesions with white radiating hyperkeratotic striae. When the lesions evolve, they may take on a honeycomb appearance. Oral discoid lupus frequently involves the lip and spreads from the vermilion border to the skin of the lip. Lupus oral ulcers have a gradual onset and can occur anywhere on the oral mucosa, the most common locations being the hard palate, buccal mucosa, and vermilion border. These lesions are most commonly unilateral or asymmetric. The relationship between the presence of oral lesions and systemic disease activity remains unclear. Note that oral candidiasis and oral lichen planus can take on a similar appearance to SLE oral ulcers. The histopathology and immunopathology of mucosal lesions are similar to the alterations seen in the skin. Vasculitis is absent.

MUSCULOSKELETAL

Arthritis

Arthritis and arthralgias are very common manifestations of SLE, present in up to 90% of patients at some point during the course of their disease. Severity of involvement can range from mild joint pain to deforming arthritis. Although any joint can be involved, lupus arthritis is characterized by a symmetric, inflammatory arthritis predominantly affecting the knees, wrists, and small joints of the hands. Synovial effusions are typically small and not as inflammatory as those present in rheumatoid arthritis. Hand deformities can occur as a result of ligamental and/or joint capsule laxity and joint subluxation. This manifestation is called “Jaccoud’s-like arthropathy” because it resembles the arthropathy that develops in patients with a history of rheumatic fever (Figure 80-6). Lupus hand deformities are reducible.
Jaccoud’s-like arthropathy sometimes occurs in the foot as well.\(^{34}\)

Although lupus arthritis is not classically associated with erosions on plain radiography, erosive disease has been described in a small subset of patients.\(^{35}\) In addition, MRI studies have shown occasional erosions in some patients with lupus arthritis.\(^{36}\) Erosive arthritis is more commonly a feature of MCTD. Some studies have demonstrated an association between lupus erosive arthritis and the presence of anticyclic citrullinated protein (anti-CCP) antibodies.\(^{37}\) In addition to arthritis, tendinitis or tenosynovitis is frequently observed in SLE patients.\(^{37}\) Tendon rupture is a very uncommon occurrence.

Synovial biopsies from lupus arthritis patients have shown a variety of abnormalities, including deposition of fibrin-like material, focal or diffuse synovial lining cell proliferation, vascular congestion, perivascular mononuclear cell infiltration, vasculitis, and obliteration of vessel lumina.\(^{38}\) Radiographic studies of patients with SLE have revealed changes such as cystic bone lesions, periarticular soft tissue swelling, demineralization, acral sclerosis, joint subluxations, and erosions.\(^{39,40}\)

**Avascular Necrosis**

Avascular necrosis (AVN), also referred to as aseptic necrosis and ischemic necrosis, is a painful and disabling condition that occurs in some SLE patients.\(^{41}\) AVN is the end result of interruption of the blood supply to bone, leading to reactive hyperemia of adjacent bone, demineralization, and then collapse. The most commonly affected sites include the femoral heads, tibial plateaus, and femoral condyles, but smaller joints can be involved as well. AVN is often bilateral, and joint effusions may occur. AVN of the femoral head should be suspected in an SLE patient with groin pain that is worsened with weight bearing and movement of the hip. The pain may radiate down the side of the thigh, and a limp might be evident. Although both plain radiographs and magnetic resonance imaging (MRI) can be helpful in the diagnosis of AVN, MRI is the more sensitive test. One prospective MRI study of 45 SLE patients on glucocorticoid therapy demonstrated that 34% of patients developed silent osteonecrosis of the femoral head.\(^{42}\) However, MRI studies may be too sensitive an indicator, in that some lupus patients with suggestive MRI findings never progress to clinical symptoms of AVN. Therefore, MRI findings should always be interpreted in the context of the clinical setting. The use of high doses of glucocorticoids is a well-known risk factor for the development of AVN, but AVN has also been described in SLE patients who have never used glucocorticoids. An MRI study of 72 newly diagnosed SLE patients demonstrated that AVN typically developed within the first 3 months of initiation of high-dose glucocorticoids.\(^{43}\) Epidemiologic studies have shown that high disease activity as measured by the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) and the use of cytotoxic medications are also associated with AVN.\(^{44,45}\)

**Myositis**

Although myalgias occur commonly in SLE, true myositis is relatively rare. One study of SLE patients at the National Institutes of Health (NIH) found a prevalence of myositis of 8%.\(^{46}\) In most of those patients, myositis was one of the presenting features of SLE. Myositis usually involves the proximal upper and lower extremities. Histologic findings in SLE myositis are often less pronounced than those observed in polymyositis.

A biopsy study of 55 unselected SLE patients demonstrated that several pathologic changes, including type II fiber atrophy, lymphocytic vasculitis, and myositis, were increased in SLE patients compared with control patients.\(^{47}\) It is important to distinguish myositis secondary to SLE from myopathy caused by glucocorticoids, antimalarial agents, or statins because the treatment is very different. Muscle enzymes such as creatine phosphokinase (CPK) and aldolase are typically normal in patients with both glucocorticoid- and hydroxychloroquine-induced myopathy. Biopsy specimens usually reveal characteristic findings, including vacuolar changes in hydroxychloroquine myopathy and type II fiber atrophy in glucocorticoid myopathy in the absence of inflammation. Finally, it is important to think broadly about other potential causes of myopathy and myositis in SLE patients, including thyroid disease, electrolyte abnormalities, and infectious myositis. One must also consider the possibility of MCTD because myositis can be a prominent feature of that disorder.

**RENAL INVOLVEMENT**

**General Considerations**

Renal involvement is common in SLE and is a significant cause of morbidity and mortality.\(^{48}\) It is estimated that up to 90% of SLE patients will have pathologic evidence of renal involvement on biopsy, but only 50% will develop clinically significant nephritis. The clinical presentation of lupus nephritis is highly variable, ranging from asymptomatic hematuria and/or proteinuria to frank nephrotic syndrome to rapidly progressive glomerulonephritis with loss of renal function. Lupus nephritis typically develops within the first
36 months of the disease, although there are exceptions. Thus, periodic screening for the presence of nephritis is a critical component of the ongoing evaluation and management of SLE patients. Routine screening procedures include inquiring about new-onset polyuria, nocturia, or foamy urine and looking for the presence of hypertension or lower extremity edema. It is important to screen at regular intervals for the presence of proteinuria and/or hematuria and a change in serum creatinine; in active SLE patients, screening at 3-month intervals is prudent.

**Types of Renal Involvement in Systemic Lupus Erythematosus**

Several forms of renal involvement have been noted in SLE, including immune complex–mediated glomerulonephritis (most common form), tubulointerstitial disease, and vascular disease. Glomerulonephritis is characterized by immune complex deposition and inflammatory cell infiltration into the glomerulus. The pattern of glomerular injury is primarily related to the site of immune complex deposition. Tubulointerstitial and vascular disease can occur with or without immune complex–mediated glomerulonephritis. Tubulointerstitial disease has been observed in up to 66% of SLE renal biopsy specimens and is characterized by inflammatory cell infiltrates, tubular damage, and interstitial fibrosis. The presence of tubulointerstitial disease is a strong predictor of poor long-term renal outcome.

Renal vascular lesions in SLE include “lupus vasculopathy,” thrombotic microangiopathy (TMA), vasculitis, and nonspecific vascular sclerosis. Lupus vasculopathy is defined as the presence of immunoglobulin and complement-containing hyaline thrombi within the glomerular capillary or arteriolar lumina. Inflammatory changes to the vascular wall are absent. TMA is characterized by the presence of fibrin thrombi within the glomerular capillary or arteriolar lumina and may be associated with the presence of antiphospholipid antibodies. The finding of TMA should prompt consideration of antiphospholipid antibody syndrome nephropathy (APS/N). Although exceedingly rare, true vasculitis characterized by leukocyte infiltration and fibrinoid necrosis of arterial walls can occur. Nonspecific sclerotic vascular lesions characterized by fibrous intimal thickening are commonly observed. The presence of such vascular lesions is associated with decreased renal survival. In addition to the lupus-related renal lesions described previously, SLE patients may develop renal abnormalities that are unrelated to their underlying SLE. Such pathologic lesions include focal segmental glomerulosclerosis (FSGS), hypertensive nephrosclerosis, and thin basement membrane disease. In an SLE patient in whom renal disease is suspected, renal biopsy is critical in distinguishing between these potential causes and in guiding appropriate management decisions.

**Laboratory Evaluation**

**Urinalysis**

Performance of a urinalysis with microscopy is essential in the screening and monitoring of lupus nephritis. Hematuria, pyuria, dysmorphic red blood cells, red blood cell casts, and white blood cell casts may all be present. Red blood cell casts are very specific, but not sensitive, for the diagnosis of glomerulonephritis. Early morning urine specimens, which tend to be concentrated and acidic, are ideal for the detection of red blood cell casts. White blood cells, red blood cells, and white blood cell casts may indicate the presence of tubulointerstitial involvement. Hematuria in the absence of proteinuria might be due to urolithiasis, menstrual contamination, or bladder pathology, particularly transitional cell carcinoma in a patient with previous cyclophosphamide exposure.

Accurate measurement of proteinuria is critical because proteinuria is a very sensitive indicator of glomerular damage. In addition, studies of chronic kidney disease have shown that the magnitude of proteinuria is a strong predictor of glomerular filtration rate decline. Normal daily protein excretion is less than 150 mg. Although the gold standard tool is an accurately collected 24-hour urine protein, this test can be cumbersome for patients and prone to errors in undercollection and overcollection. Thus, many clinicians are currently using the random spot urine protein-to-creatinine ratio out of convenience. Use of the spot ratio is controversial because data suggest that the spot ratio often is not representative of the findings in a timed collection, especially in the range of 0.5 to 3.0 (the range of most lupus nephritis flares). However, a spot ratio can be a helpful screening test for the presence of proteinuria and is useful in differentiating nephrotic from non-nephrotic range proteinuria. Urine dipstick should not be used for the quantification of proteinuria because it reflects protein concentrations and varies depending on the volume of the sample. Many experts currently recommend calculation of the protein:creatinine ratio from a 12- or 24-hour urine collection as the gold standard of proteinuria assessment.

**Measurement of Renal Function**

Although easy to measure, serum creatinine is a fairly insensitive indicator of early decline in glomerular filtration rate (GFR). Creatinine is freely filtered across the glomerulus and is also secreted by the proximal tubule. As GFR falls, the rise in serum creatinine is counteracted by increased tubular creatinine secretion. In addition, hemodynamic changes such as those caused by treatment with angiotensin-converting enzyme inhibitors or nonsteroidal anti-inflammatory drugs are a common cause of changes in serum creatinine levels in the absence of progression of underlying renal disease. However, trending serum creatinine over time is a reasonable method by which to follow a patient’s renal function. Some clinicians prefer to utilize equations that estimate GFR, such as the Cockcroft-Gault and Modification of Diet in Renal Disease (MDRD) study equations. Whichever method is chosen, the detection of changes in renal function over time is more important than the absolute level when following lupus nephritis patients in clinical practice.

**Renal Biopsy**

When an SLE patient has clinical or laboratory features that suggest the presence of nephritis, a renal biopsy should be
performed to confirm the diagnosis, evaluate the degree of disease activity, and determine an appropriate course of treatment. Before renal biopsy, ultrasonography is recommended to assess kidney size and structure and to rule out renal vein thrombosis. Kidney size of less than 75% of normal is a relative contraindication to biopsy. SLE glomerulonephritis is classified by the International Society of Nephrology/Renal Pathology Society (ISN/RPS) into six categories based on light microscopic, immunofluorescent, and electron micrographic findings (Table 80-4 and Figure 80-7).

An individual biopsy might exhibit just one of the ISN/RPS pathologic classes or a combination of classes. Class I is characterized by normal appearing glomeruli on light microscopy and mesangial immune deposits on immunofluorescence. Class II is characterized by mesangial proliferation on light microscopy and mesangial deposits on immunofluorescence. Class III and IV lupus nephritis lesions are highly inflammatory and are characterized by immune complex deposition in the subendothelial space. They have traditionally been described as “proliferative” because of the presence of proliferating endocapillary cells within the glomeruli. They are believed to be interrelated lesions that differ in the distribution of endocapillary immune complex deposition. Class III denotes that less than 50% of glomeruli are involved, and class IV denotes that 50% or more of glomeruli are involved. Class IV lesions are subcategorized according to whether most glomeruli show focal (≤50% of the glomerular tuft) or global (≥50% of the glomerular tuft) involvement. These lesions are further described as active (A), chronic (C), or a mixture of the two (A/C). Thick subendothelial immune deposits form classic “wire loop” lesions. Class V lupus nephritis is characterized by immune complex deposition in the subepithelial space, resulting in widespread thickened capillary loops. These findings are similar to those observed in idiopathic membranous nephritis. However, the presence of concomitant mesangial deposits plus or minus tubuloreticular inclusion bodies would favor the diagnosis of lupus. This lesion is commonly manifested clinically as nephrotic range proteinuria. Class V nephritis may occur in a pure histopathologic form or in combination with features of class III or class IV nephritis. Class VI nephritis is defined by the presence of more than 90% globally sclerotic glomeruli. In addition to the type of glomerular pathology, the ISN/RPS classification system dictates that tubulointerstitial disease and/or vascular disease should be noted on the diagnostic line. Immunofluorescence studies are an important supplement to the findings on light microscopy. Immunofluorescence reveals the type and pattern of immune complex deposition. Lupus nephritis is characterized by a granular pattern of immunofluorescence along the glomerular basement membrane, mesangium, and/or tubular basement membranes. The characteristic findings of lupus nephritis are sometimes referred to as the “full-house” pattern, because IgG, IgM, IgA, C3, and C1q are all found in the deposits. Electron microscopy is useful in more precisely localizing the sites of immune complex deposition. The finding of tubuloreticular inclusion bodies within endothelial cells is strongly suggestive of the diagnosis of lupus nephritis. However, because tubuloreticular inclusion bodies are

### Table 80-4 International Society of Nephrology/Renal Pathology Society Classification of Lupus Nephritis

<table>
<thead>
<tr>
<th>WHO Type</th>
<th>Description</th>
</tr>
</thead>
</table>
| Class I  | Minimal Mesangial Lupus Nephritis  
Normal glomeruli by light microscopy, but mesangial immune deposits by immunofluorescence |
| Class II | Mesangial Proliferative Nephritis  
Purely mesangial hypercellularity of any degree or mesangial matrix expansion by light microscopy, with mesangial immune deposits.  
Few isolated subepithelial or subendothelial deposits may be visible by immunofluorescence or electron microscopy, but not by light microscopy |
| Class III | Focal Lupus Nephritis  
Active or inactive focal, segmental, or global endocapillary or extracapillary glomerulonephritis involving ≤50% of all glomeruli, typically with focal subendothelial immune deposits, with or without mesangial alterations |
| Class IV | Diffuse Lupus Nephritis  
Active or inactive diffuse, segmental, or global endocapillary or extracapillary glomerulonephritis involving ≥50% of all glomeruli, typically with diffuse subendothelial immune deposits, with or without mesangial alterations. This class is subdivided into diffuse segmental (IV-S) lupus nephritis when ≥50% of the involved glomeruli have segmental lesions, and diffuse global (IV-G) lupus nephritis when ≥50% of the involved glomeruli have global lesions. Segmental is defined as a glomerular lesion that involves less than half of the glomerular tuft. This class includes cases with diffuse wire loop deposits, but with little or no glomerular proliferation |
| Class V  | Membranous Lupus Nephritis  
Global or segmental subepithelial immune deposits or their morphologic sequelae by light microscopy and by immunofluorescence or electron microscopy, with or without mesangial alterations  
Class V nephritis may occur in combination with class III or class IV, in which case both are diagnosed  
Class V nephritis may show advanced sclerotic lesions |
| Class VI | Advanced Sclerotic Lupus Nephritis  
≥90% of glomeruli globally sclerosed without residual activity |

WHO, World Health Organization.  
Hematuria might be absent in patients with severe class IV nephritis, and proteinuria can be modest in patients with class V nephritis. A repeat renal biopsy may be indicated in certain clinical settings (e.g., if a patient is not responding appropriately to therapy, if a patient unexpectedly worsens after having achieved a good response to therapy). Repeat renal biopsy is especially important because urinary parameters such as hematuria and the degree of proteinuria imperfectly predict the underlying renal pathology. Additionally, situations associated with increased levels of interferon alpha, chronic viral infections such as hepatitis B/C and the human immunodeficiency virus (HIV) must be ruled out.

Renal biopsy is especially important because urinary parameters such as hematuria and the degree of proteinuria imperfectly predict the underlying renal pathology. Hematuria might be absent in patients with severe class IV nephritis, and proteinuria can be modest in patients with class V nephritis. A repeat renal biopsy may be indicated in certain clinical settings (e.g., if a patient is not responding appropriately to therapy, if a patient unexpectedly worsens after having achieved a good response to therapy). Repeat
renal biopsy can be useful in detecting class transformation that occurs in 15% to 50% of lupus nephritis patients during the course of their disease. Class transformation can occur spontaneously or as a result of treatment.

Outcome

Each ISN/RPS histopathologic class portends a distinct renal prognosis. Patients with class I and class II nephritis have an excellent renal prognosis and do not require any specific therapy. In contrast, the long-term renal prognosis of class III or class IV nephritis is believed to be poor in the absence of immunosuppressive therapy. Although the long-term renal prognosis of class V nephritis is more favorable than that of class III or class IV nephritis, class V patients are more likely to suffer from morbid complications of the nephrotic syndrome, including cardiovascular disease, thromboembolic disease, and hyperlipidemia. Several epidemiologic studies have defined demographic, clinical, and histopathologic factors associated with renal outcome in patients with lupus nephritis. Studies have shown that African-Americans and Hispanics/Latinos generally experience a worse renal prognosis than white and Chinese populations. The reasons for this disparity most likely involve a combination of genetic and socioeconomic factors. A retrospective analysis of 65 patients at the NIH suggested that age greater than 30 years, African-American race, low hematocrit, elevated serum creatinine, and low C3 complement were associated with increased probability of renal failure. The histologic features of cellular crescents and interstitial fibrosis were also associated with worse renal prognosis.50

PLEUROPULMONARY INVOLVEMENT

Pleuropulmonary manifestations of SLE are diverse and can involve any aspect of the lung (Table 80-5).

<table>
<thead>
<tr>
<th>Manifestation</th>
<th>Key Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleuritis</td>
<td>May occur with or without effusion</td>
</tr>
<tr>
<td></td>
<td>May correlate with elevated serum C-reactive protein</td>
</tr>
<tr>
<td>Pleural effusion</td>
<td>May be asymptomatic</td>
</tr>
<tr>
<td></td>
<td>Usually small, bilateral, exudative</td>
</tr>
<tr>
<td></td>
<td>Common feature of drug-induced lupus</td>
</tr>
<tr>
<td></td>
<td>Must exclude infection, malignancy, heart failure</td>
</tr>
<tr>
<td>Acute pneumonitis</td>
<td>Severe respiratory illness with fever, cough, pulmonary infiltrates, hypoxemia</td>
</tr>
<tr>
<td></td>
<td>Pleural effusion may be present</td>
</tr>
<tr>
<td></td>
<td>High mortality rate</td>
</tr>
<tr>
<td>Chronic interstitial lung disease</td>
<td>Bronchopulmonary lavage might be necessary to exclude infection</td>
</tr>
<tr>
<td></td>
<td>May develop after acute pneumonitis or in a more insidious fashion</td>
</tr>
<tr>
<td></td>
<td>Presents as dyspnea on exertion, pleuritic chest pain, nonproductive cough</td>
</tr>
<tr>
<td>Diffuse alveolar hemorrhage</td>
<td>Presents as dyspnea and cough, alveolar infiltrates, fall in blood hemoglobin level</td>
</tr>
<tr>
<td></td>
<td>Hemothysis may not be present</td>
</tr>
<tr>
<td></td>
<td>Diffusion capacity of carbon monoxide typically increased</td>
</tr>
<tr>
<td></td>
<td>Bronchoscopic lavage confirms the diagnosis and excludes infection</td>
</tr>
<tr>
<td>Pulmonary arterial hypertension</td>
<td>Presents as dyspnea on exertion, fatigue, chest pain, nonproductive cough</td>
</tr>
<tr>
<td></td>
<td>Diagnosis should be confirmed with right heart catheterization</td>
</tr>
<tr>
<td>Shrinking lung syndrome</td>
<td>Exclude secondary causes of pulmonary hypertension, including thromboembolic disease</td>
</tr>
<tr>
<td></td>
<td>Dyspnea, low lung volumes, elevation of hemi-diaphragms in absence of lung parenchymal involvement</td>
</tr>
</tbody>
</table>
bronchoalveolar lavage can be useful in detecting pulmonary pathogens. A “tree and bud” pattern on high-resolution computed tomography (HRCT) may suggest the presence of an atypical pneumonia. Acute lupus pneumonitis is associated with significant morbidity and mortality. One series of 12 patients reported a mortality rate of 50% with deaths from respiratory failure, opportunistic infection, and thromboembolic events. Three of the surviving patients progressed to chronic interstitial pneumonitis.70

Chronic Interstitial Lung Disease

Chronic interstitial lung disease is a rare manifestation of SLE. It occurs more commonly in other connective tissue diseases such as systemic sclerosis, rheumatoid arthritis, and polymyositis/dermatomyositis. Interstitial lung disease in the setting of SLE can develop after one or more episodes of acute pneumonitis but can also occur in an insidious fashion.70 Symptoms are similar to those seen in patients with idiopathic interstitial lung disease and include dyspnea on exertion, pleuritic chest pain, and chronic, nonproductive cough. The diagnosis of interstitial lung disease is often made on the basis of clinical–radiologic findings; lung biopsy is not routinely performed. Chest radiography might be normal early in the disease but can show reticular opacities. Pulmonary function studies show a restrictive pattern with reduction in total lung capacity and reduction in the diffusion capacity of carbon monoxide (DLCO). HRCT is more sensitive than chest radiography in detecting interstitial lung disease and in distinguishing reversible lesions (ground glass opacities) from irreversible fibrotic lesions. Nonspecific interstitial pneumonia (NSIP) and usual interstitial pneumonia (UIP) are the most common patterns detected on histopathology and HRCT. Before making the diagnosis of interstitial lung disease, it is important to exclude infection, pulmonary edema, and malignancy.

Diffuse Alveolar Hemorrhage

Diffuse alveolar hemorrhage (DAH) is a life-threatening manifestation of SLE that occurs in less than 2% of patients. It is characterized by acute or subacute onset of dyspnea and cough in the setting of new alveolar infiltrates on chest radiography and a fall in blood hemoglobin level. Similar to other causes of DAH, hemoptysis is not universally present. Although most patients are too ill to receive this test, the DLCO is typically increased in the setting of DAH owing to the presence of extravascular hemoglobin within the alveoli. Bronchoscopy with bronchoalveolar lavage (BAL) is important in ruling out infection and confirming the diagnosis. Characteristic findings include visualization of blood in the airways and serosanguineous BAL fluid that does not clear with continued lavage. Hemosiderin-laden macrophages may be seen in the BAL fluid. Various histopathologic patterns have been described in lupus DAH, including bland pulmonary hemorrhage, capillaritis, diffuse alveolar damage, and vasculitis of small arterioles and small muscular pulmonary arteries. DAH usually occurs in the setting of serologically and clinically active SLE, and lupus nephritis is the most common concurrent SLE manifestation. However, DAH occasionally may be the initial manifestation of SLE. Mechanical ventilation is often required, and infectious complications are common. Despite aggressive therapy, mortality from DAH continues to be 50%.

Pulmonary Arterial Hypertension

Pulmonary arterial hypertension (PAH) is a rare, devastating complication of SLE that is defined as a mean pulmonary artery pressure greater than 25 mm Hg at rest on right heart catheterization. Other key findings on heart catheterization include a normal pulmonary capillary wedge pressure and elevated pulmonary capillary resistance. Symptoms of PAH include dyspnea on exertion, fatigue, chest pain, and nonproductive cough. Physical examination findings may include a pronounced second pulmonary heart sound, a left parasternal lift, and signs of a volume-overloaded state. Chest radiography and HRCT are important in excluding lupus pneumonitis. Chest radiography may show cardiomegaly and a prominent pulmonary artery segment. The electrocardiogram often shows right axis deviation. Pulmonary function studies demonstrate a reduction in DLCO. Although transthoracic Doppler echocardiography is a decent screening test for PAH, the diagnosis should be confirmed by right heart catheterization. Similar to interstitial lung disease, PAH more commonly occurs in association with scleroderma and mixed connective tissue disease.

In an SLE patient who has been diagnosed with PAH, an evaluation must be performed for secondary causes of pulmonary hypertension. Ventilation and perfusion (V/Q) lung scan and/or helical computed tomography are useful in excluding chronic thromboembolic disease. Echocardiography can rule out left heart failure and intracardiac shunting. A sleep study can be useful in ruling out obstructive sleep apnea. An evaluation for interstitial lung disease is necessary. Some studies suggest that PAH occurs more commonly in patients with Raynaud’s phenomenon.

Other

Shrinking lung syndrome occurs in a small subset of SLE patients and should be considered when evaluating an SLE patient with unexplained dyspnea and pleuritic chest pain.72,73 The cause of the disorder remains controversial. Diaphragmatic myopathy, abnormal chest wall expansion, phrenic neuropathy, and pleural inflammation/fibrosis have been reported as possible factors. The prognosis of this syndrome seems to be good, and progressive respiratory failure is uncommon.

Although symptomatic bronchiolar disease is uncommon in SLE, abnormalities in pulmonary function studies have been reported in up to two-thirds of SLE patients.74 One study of nonsmoking SLE patients found that 24% of patients had pulmonary function studies consistent with small airway disease. Rare case reports of bronchiolitis obliterans organizing pneumonia (BOOP) in the setting of SLE have been described.75

CARDIOVASCULAR INVOLVEMENT

Cardiovascular disease is a frequent complication of SLE and may involve the pericardium, myocardium, valves, and coronary arteries.
**Pericarditis**

Pericarditis, with or without an effusion, is the most common cardiac manifestation of SLE, occurring in more than 50% of SLE patients at some point during the course of their disease.\(^5\) Pericardial effusions are usually small and asymptomatic with relatively few reported cases. However, acute pericarditis can be associated with myocardial ischemia, pleuritis, pneumonitis, or pulmonary embolism. Histopathologic evidence of pericarditis is much more common than clinically symptomatic disease during life.\(^77\) Symptomatic pericarditis classically presents as sharp, precordial chest pain that is improved in the upright position. A pericardial rub and tachycardia may be detected on cardiac auscultation. The electrocardiogram demonstrates diffuse ST segment elevation. Similar to pleuritis, pericarditis usually occurs in the setting of active SLE in other organ systems. Although rare, SLE pericarditis complicated by large effusions and tamponade physiology have been reported. Purulent effusions necessitating pericardiocentesis have also been described, but rarely. The differential diagnosis of pericardial chest pain in an SLE patient includes costochondritis, gastroesophageal reflux disease, pulmonary embolism, myocardial ischemia, pleuritis, pneumonitis, and pulmonary hypertension.

**Myocarditis**

Myocarditis, an uncommon manifestation of SLE, should be suspected in a patient presenting with various combinations of the following clinical features: unexplained heart failure or cardiomegaly, unexplained tachycardia, and unexplained electrocardiographic abnormalities. Echocardiography can confirm the presence of systolic or diastolic dysfunction and/or global hypokinesis. If myocarditis is suspected, an endomyocardial biopsy may be helpful in confirming the diagnosis and excluding other causes of cardiomyopathy such as hydroxychloroquine toxicity. The distinguishing pathologic finding of hydroxychloroquine toxicity is myocyte vacuolization in the absence of active myocarditis. Histopathologic findings of SLE myocarditis include perivascular and interstitial mononuclear cell infiltration and sometimes fibrosis and scar.\(^77\)

**Valvular Abnormalities**

Several valvular abnormalities have been described in patients with SLE, including Libman-Sacks endocarditis (also known as atypical verrucous endocarditis), valvular thickening, valvular regurgitation, and valvular stenosis. One transesophageal echocardiographic (TEE) study demonstrated a prevalence of valvular abnormalities of 61% in SLE patients compared with 9% of controls, with vegetations present in 43% of SLE patients compared with none of the controls.\(^9\) Valvular thickening with a predilection for the mitral and aortic valves was the most common abnormality, occurring in 50% of SLE patients. Valvular regurgitation and stenosis were detected in 25% and 4% of patients, respectively.\(^9\) In this study, the presence and progression of valvular disease were not associated with SLE disease activity or treatment. Over a follow-up period of up to 5 years, some valvular abnormalities resolved and some new lesions occurred. The combined incidence of stroke, peripheral embolism, congestive heart failure, infective endocarditis, need for valve replacement, and death was 22% among patients with valvular disease compared with 15% in those without valvular disease.\(^78\)

Libman-Sacks endocarditis has been recognized in multiple pathologic studies as a characteristic valvular abnormality in SLE. Libman-Sacks verrucae typically appear as pea-sized, flat or raised, granular lesions that occur most commonly on the ventricular aspects of the mitral valve posterior leaflet.\(^77\) The verrucae often extend onto the adjacent left ventricular mural endocardium and may lead to adherence of the leaflet and chordae tendineae to the ventricular mural endocardium, resulting in valvular regurgitation. All four valves may be involved, but recent studies suggest a predominance of left-sided lesions. The lesions are frequently clinically silent because they are typically found on the undersurface of valve leaflets, surrounded by fibrous tissue. Histologically, two types of verrucae have been described: (1) active lesions consisting of fibrin clumps with infiltrating lymphocytes and plasma cells, and (2) healed lesions consisting of dense vascularized fibrous tissue with or without calcification.\(^77\) Combinations of active and healed lesions also occur. Verrucae typically do not contain polymorphonuclear cells; thus, the presence of such cells should prompt consideration of infectious endocarditis. Immunopathologic studies have demonstrated immunoglobulin and complement deposition in a granular pattern at the base of the valve, along the valve leaflet, and within the verruca itself.\(^78\) Cardiac murmurs are frequently heard in patients with SLE. They may simply result from high-flow states such as fever and anemia, or they may reflect cardiac pathology such as mitral valve prolapse or infective endocarditis. When a new murmur is evaluated, a transthoracic echocardiogram (TTE) is an appropriate first test. However, TEE should be utilized in the event of a nondiagnostic TTE or in a patient with suspected thromboembolic events. TEE has been shown to be superior to TTE for detection of Libman-Sacks endocarditis.\(^80\) Although thromboembolic events are believed to be rare complications of Libman-Sacks endocarditis, one study showed that valvular heart disease detected on TTE was associated with the presence of cerebral infarcts on MRI.\(^81\) It remains uncertain whether the incidence of valvular disease is increased in SLE patients who also have circulating antiphospholipid antibodies.

**Coronary Artery Disease**

Both intramural and extramural coronary artery disease is increased in patients with SLE. Necropsy studies have demonstrated fibrous intimal proliferation of small intramural coronary arteries and obstruction of these arteries with hyaline material.\(^77\) These lesions are similar to those observed in pathologic studies of renal and central nervous system tissue in SLE patients. The large epicardial coronary arteries may be obstructed owing to arterial emboli, in situ thrombosis, vasculitis, or atherosclerotic disease. True coronary artery vasculitis is exceedingly rare. In contrast, atherosclerotic disease is a well-recognized complication of long-standing SLE.\(^82\) Autopsy studies have demonstrated atherosclerosis in 25% to 40% of SLE patients.\(^77,83\) One epidemiologic study demonstrated that young women with
SLE have a 50-fold higher risk of myocardial infarction compared with age-matched controls. A multicenter inception cohort determined that male sex and older age at SLE diagnosis were significantly associated with the presence of atherosclerotic disease. Although patients with SLE are more likely to have classic atherosclerotic risk factors such as hypertension and exposure to corticosteroids, these risk factors alone do not fully account for the increased risk of atherosclerosis seen in SLE patients. Thus, SLE itself is believed to be an independent risk factor.

The possibility of coronary artery disease must be considered in any SLE patient presenting with chest pain and/or shortness of breath, and one should have a low threshold for a functional evaluation with a cardiac stress test. Cardiac catheterization might also be necessary for diagnosis and therapeutic intervention. It is important to evaluate these patients for the presence of antiphospholipid antibodies because coronary artery thrombosis may be a manifestation of the antiphospholipid antibody syndrome. Evaluation for and treatment of modifiable risk factors such as obesity, smoking, hypertension, and hyperlipidemia are important in mitigating the development and progression of atherosclerotic disease.

**NEUROPSYCHIATRIC INVOLVEMENT**

**General Considerations**

Neuropsychiatric lupus (NPSLE) consists of a broad range of neurologic and psychiatric manifestations that can involve any aspect of the central or peripheral nervous system. With the intention of improving the terminology and classification of NPSLE, an American College of Rheumatology (ACR) subcommittee categorized NPSLE into 19 distinct syndromes encompassing the central (CNS) and peripheral (PNS) nervous system. The extent of this classification system underscores the complexity of NPSLE. CNS disorders range from diffuse processes such as acute confusional state, headache, psychosis, and mood disorders to more focal processes such as seizures, myelopathy, and chorea. It is notable that the ACR classification system has removed the cryptic term “lupus cerebritis” from the vernacular. Although the ACR case definitions are very helpful in providing a framework in which to think about NPSLE, accurate attribution of neuropsychiatric manifestations remains challenging. It is often difficult to distinguish whether neuropsychiatric symptoms are due to active SLE or to other factors such as infection, metabolic abnormalities, severe hypertension, adverse effects of medications, or independent neurologic or psychiatric problems. No laboratory or imaging study is sufficiently sensitive or specific to confirm the diagnosis of neuropsychiatric SLE. Instead, the diagnosis is based on a thorough clinical evaluation that is corroborated by findings (or lack thereof) on brain imaging, serologic testing, lumbar puncture, and neuropsychiatric assessment.

**Pathogenesis**

Multiple pathogenic mechanisms are undoubtedly involved in the various NPSLE syndromes, but in most cases the precise pathogenesis is unknown. Many of the manifestations can be grouped into two broad categories: primary vascular injury and primary inflammatory injury. A combination of the two categories can also occur. Vascular injury includes damage to both large and small vessels via thromboembolic events, often as a consequence of antiphospholipid syndrome. In some patients, vasculopathy of small vessels may be due to vascular hyalinization, perivascular inflammation, and endothelial proliferation. In contrast, inflammatory-mediated injury might result from increased permeability of the blood-brain barrier, intrathecal production of inflammatory cytokines, and damage from antineuronal antibodies. Histopathologic studies have demonstrated multiple CNS abnormalities, including large and small multifocal infarctions, hemorrhage, bland small vessel vasculopathy, cortical atrophy, brain edema, and demyelination. True vasculitis of cerebral vessels is rare. Several autoantibodies have been implicated in the pathogenesis of some neuropsychiatric manifestations, particularly psychosis, but they lack sufficient sensitivity or specificity to guide diagnosis.

**Approach to Diagnosis**

The diagnostic evaluation of a patient with potential NPSLE is tailored to the presenting neuropsychiatric manifestation. Depending on the manifestation, consultation with a neurologist and/or a psychiatrist is advised. The most important first step is exclusion of an alternative explanation for the neuropsychiatric symptoms/signs. Lumbar puncture with cerebrospinal fluid (CSF) examination is useful for exclusion of an infectious origin. Mild lymphocytic pleocytosis and elevated CSF protein are sometimes observed but are not uniformly present. CSF findings are not sensitive or specific enough to confirm a diagnosis of neuropsychiatric SLE. It is important to recognize that infection is a common cause of CNS symptoms in SLE patients who are hospitalized with altered mental status, and must be rigorously ruled out. Progressive multifocal leukoencephalopathy (PML) is a rare infection that also merits consideration in an SLE patient presenting with symptoms of CNS dysfunction. PML occurs more frequently in SLE patients than in patients with other rheumatic diseases, even in the absence of significant immunosuppressive therapy. Thus, in an SLE patient with unexplained new
neurologic symptoms, polymerase chain reaction (PCR) of the CSF for the presence of John Cunningham (JC) virus should be considered. In situations where PCR is negative, brain biopsy might be needed to confirm the diagnosis. Electromyography and nerve conduction studies are important in the setting of suspected peripheral neuropathy. Electroencephalogram (EEG) is necessary in the evaluation of seizure. Neuropsychological testing may be helpful in the setting of suspected cognitive dysfunction.

MRI is the preferred imaging modality in patients with suspected NPSLE. The most common noted abnormalities are small, hyperintense, T2-weighted, focal white matter lesions located in the periventricular and subcortical white matter of the frontoparietal region of the brain. However, these findings are nonspecific and can be observed in other disease processes such as atherosclerotic vascular disease and multiple sclerosis. Other common MRI findings include cortical atrophy, ventricular dilation, cerebral edema, diffuse white matter abnormalities, focal atrophy, infarction, leukoencephalopathy, and hemorrhage. MRI is especially useful in detecting the presence of infarcts, hemorrhage, and myelopathy and sometimes can help to exclude infectious conditions such as brain abscess. MRI is most likely to show abnormalities in the setting of focal neurologic deficits, seizures, chronic cognitive dysfunction, and antiphospholipid antibody–mediated disease and is less likely to show abnormalities in the setting of headache, acute confusional state, and psychiatric syndromes.

Estimates of the prevalence of NPSLE have varied widely in the literature, largely depending on the extent to which headache and/or mild cognitive abnormalities have been included in the analysis. However, the preponderance of these studies suggests that CNS manifestations predominate over PNS manifestations. Several of the more common syndromes and associated differential diagnoses are described in the following paragraphs.

**Selected Neuropsychiatric Lupus Syndromes**

Headaches are reported in more than 50% of SLE patients, but attribution of the headache to SLE is extremely difficult. Both migraine and tension-type headaches have been described. One meta-analysis determined that the prevalence of primary headache syndromes was not different between SLE and control patients, and that headache was not related to SLE disease activity. The evaluation of headache in an SLE patient should be similar to that in a non-SLE patient and should be directed by the presence or absence of worrisome features such as fever, meningismus, altered mental status, and focal neurologic signs.

Cognitive dysfunction, manifested primarily by deficits in thinking, memory, and concentration, is being increasingly recognized in SLE patients. Some have estimated a prevalence of up to 80%, although serious cognitive impairment is much less common. Some studies suggest that cognitive dysfunction may be associated with the presence of antiphospholipid antibodies, but a causal relationship has not been definitively established. Documentation of the presence and extent of cognitive dysfunction via neuropsychiatric testing can be useful in establishing a baseline in a particular patient that can be followed over time, particularly when a therapeutic intervention is being considered.

Psychiatric disorders such as psychosis, depression, and anxiety can occur in SLE, and consultation with a psychiatrist is highly recommended in the evaluation of patients with these symptoms. An SLE patient presenting with psychosis represents a distinct diagnostic and therapeutic challenge. The differential diagnosis includes CNS infection, primary schizophrenia, systemic metabolic abnormalities, and psychosis occurring as a side effect of corticosteroid therapy or illicit drugs. Steroid-induced psychosis is dose dependent and typically occurs within the first 2 weeks of treatment initiation.

Although rare, demyelinating syndromes such as optic neuritis and myelitis can occur as part of the spectrum of NPSLE. Optic neuritis is characterized by pain with eye movement, central visual field loss, and a waxing and waning course. Optic neuritis should be differentiated from ischemic optic neuropathy, which typically presents with acute, painless loss of vision and lack of significant improvement in vision over time. Myelitis is characterized by the onset of bilateral lower extremity paresthesia, numbness, and weakness that can rapidly progress to involve the upper limbs and the muscles of respiration. A sensory level is usually noted, and autonomic involvement of the bowel and bladder is common. Band-like pain or discomfort around the abdomen is a characteristic symptom. It is important to differentiate myelitis from other causes of myelopathy, including infection, structural spinal cord abnormalities, and vascular insult to the spinal cord. Spinal cord MRI is the critical first test. CSF examination is important to rule out infection.

The combination of optic neuritis and myelitis presents a special clinical challenge because this combination of features can occur in the setting of SLE, multiple sclerosis (MS), or neuromyelitis optica (NMO). Complicating matters, the clinical presentation of antiphospholipid antibody syndrome can mimic demyelinating disease. A thorough clinical history, MRI, CSF analysis, and serologic testing are helpful in distinguishing these entities. NPSLE and MS can have identical white matter lesions on brain MRI. Spinal cord lesions may differ in that MS spinal cord lesions usually span an area measuring less than two spinal segments, but SLE lesions are often longitudinally extensive with involvement of more than three spinal cord segments. Although oligoclonal bands may be present in the CSF of both SLE and MS patients, the absence of oligoclonal bands and the presence of a pleocytosis significantly lessen the likelihood of MS. Although a positive ANA has been described in up to 27% of patients with MS, more specific subserologies such as anti-dsDNA and anti-Sm favor the diagnosis of SLE. NMO is defined as manifested by two of the following three features: presence of the NMO IgG antibody (antibody to aquaporin 4), absence of brain lesions diagnostic of MS, and longitudinally extensive myelitis on MRI. The NMO antibody has a specificity of greater than 90% for the diagnosis of NMO. However, SLE and NMO may occur concurrently in a given patient.

Patients with SLE are at increased risk of stroke, with ischemic stroke being more common than intracerebral hemorrhage. Studies have demonstrated an association between the presence of antiphospholipid antibodies (aPL) and/or valvular heart disease and the risk of stroke. Brain MRI is a critical test in the diagnosis of ischemic or
hemorrhagic stroke, and magnetic resonance angiography (MRA) can detect vessel aneurysms. Echocardiography, carotid ultrasound, and electrocardiography are important diagnostic tests in the setting of suspected thromboembolic cerebrovascular disease.

Peripheral neuropathy has been observed in up to 20% of SLE patients and typically is characterized by a symmetric, length-dependent sensory or sensorimotor polyneuropathy. Vasculitis of the vasa nervorum and demyelination are two well-recognized pathogenic mechanisms. When the results of nerve conduction studies (NCS) are normal, small-diameter nerve fibers are likely involved. Small-fiber involvement frequently presents as fluctuating numbness and tingling of the upper extremities and hands. A devastating, large-fiber vasculitic neuropathy can also develop in patients with SLE. Autonomic neuropathies, cranial neuropathies, and abnormalities of the neuromuscular junction resembling myasthenia gravis have also been described.

GASTROINTESTINAL INVOLVEMENT

SLE can involve any part of the gastrointestinal system. Dysphagia is noted in up to 13% of patients, and manometric studies have detected abnormalities of esophageal motility. Decreased peristalsis is most commonly observed in the upper one-third of the esophagus. In contrast to scleroderma, involvement of the lower esophageal sphincter is rare in SLE. A variety of potential pathogenic mechanisms have been described, including muscle atrophy, inflammation of esophageal muscle, and ischemic or vasculitic damage to the Auerbach plexus.

Abdominal pain, sometimes accompanied by nausea and vomiting, has been reported in up to 40% of SLE patients and can be due to SLE-related causes, medication side effects, and non–SLE-related causes such as infection. When evaluating an SLE patient with abdominal pain, it is critical to rule out non-SLE conditions. It is important to note that when patients are treated with corticosteroids and/or other immunosuppressive medications, clinical signs of an acute abdomen such as rebound tenderness can be masked. Thus, delay in diagnosis is common. SLE-related causes of abdominal pain may include peritonitis, pancreatitis, mesenteric vasculitis, and intestinal pseudo-obstruction. Although autopsy studies have revealed evidence of peritoneal inflammation in up to 72% of SLE patients, the presence of ascites is rare. If an SLE patient presents with abdominal pain and ascites, paracentesis is warranted to rule out infection. Peritonitis can also occur in the setting of mesenteric ischemia, bowel infarction, and pancreatitis. Thus, abdominal imaging is an important part of the initial evaluation.

Pancreatitis due to SLE is uncommon and usually is associated with active SLE in other organs. When considering the possible diagnosis of pancreatitis, it is important to note that elevated serum amylase may be misleading in that it has been observed in SLE patients in the absence of pancreatitis. Although corticosteroids and azathioprine have been associated with the development of pancreatitis in non-SLE patients, these medications do not seem to play a major role in the development of pancreatitis in SLE patients. It is important to rule out non-SLE causes of pancreatitis such as biliary disease, alcohol consumption, and hypertriglyceridemia.

Mesenteric vasculitis is a very rare manifestation of SLE that can present with a range of symptoms from cramping, bloating, and anorexia to an acute abdomen with diarrhea and gastrointestinal hemorrhage. Accurate diagnosis and prompt treatment are essential to prevent the potential catastrophic complications of necrotic bowel, perforation, and sepsis. Abdominal radiography may show distention of bowel loops, thickening and thumbprinting of the bowel wall, and/or free air in the abdomen. Ultrasoundography may be useful in demonstrating bowel wall edema and thickening. Abdominal CT is thought to be the most useful imaging modality for the early diagnosis of mesenteric ischemia and can demonstrate prominence of mesenteric vessels with a palisade pattern supplying dilated bowel loops, ascites, and bowel wall thickening with a double halo sign. Gastroscopy and colonoscopy can sometimes reveal findings of ischemia and ulceration. Because lupus mesenteric vasculitis typically involves the small vessels (arterioles and venules) of the bowel submucosa, mesenteric angiography is usually nondiagnostic. However, angiography can be helpful in ruling out larger vessel causes of mesenteric ischemia, such as polyarteritis nodosa, atherosclerotic disease, or thrombosis resulting from antiphospholipid antibody disease.

Liver test abnormalities have been described in up to 60% of SLE patients at some point during the course of their illness, but clinically significant liver disease is rarely a direct manifestation of SLE. For this reason, the presence of liver disease should prompt a search for non-SLE causes, including medications such as nonsteroidal anti-inflammatory drugs, methotrexate, and azathioprine. Abnormal liver enzymes may be caused by hepatic steatosis as a result of obesity, concomitant diabetes mellitus, or treatment with corticosteroids. Infections such as viral hepatitis, cytomegalovirus (CMV), and Epstein-Barr virus (EBV) must also be excluded. Once medications and infections have been ruled out as possible culprits, persistent liver test abnormalities should prompt an investigation with an abdominal ultrasound and possibly a liver biopsy. Lupus hepatitis is believed to be a distinct entity from autoimmune hepatitis. Lupus hepatitis is typically characterized by the presence of lobular inflammation with a paucity of lymphoid infiltrates. These findings contrast with those of autoimmune hepatitis, in which periportal (interface) inflammation and dense lymphoid infiltrates dominate. Although ANA is frequently seen in these disorders, anti–smooth muscle and anti-LKM antibodies are more frequently noted in autoimmune hepatitis than in lupus hepatitis. Rarely, nodular regenerative hyperplasia complicates SLE. This disorder causes diffuse nodularity of the liver with little fibrosis and can result in portal hypertension. Nodular regenerative hyperplasia is also associated with the presence of antiphospholipid antibodies in some patients. Vascular disorders of the liver such as Budd-Chiari syndrome, hepatic veno-occlusive disease, and hepatic infarction have been described, especially in the setting of antiphospholipid antibodies.

Other rare gastrointestinal manifestations of SLE include intestinal pseudo-obstruction and protein-losing enteropathy. Intestinal pseudo-obstruction is characterized by decreased intestinal motility caused by dysfunction of the visceral smooth muscle or enteric nervous system.
SLE can affect the eye in a variety of ways. The most common ocular manifestation is keratoconjunctivitis sicca (KCS), which can occur in the presence or absence of secondary Sjögren’s syndrome. Retinal abnormalities can be detected on ophthalmoscopic examination as retinal hemorrhages, vasculitic-appearing lesions, cotton wool spots, and hard exudates. SLE retinopathy is believed to be an immune complex–mediated vasculopathy and/or the result of microthrombotic events. The presence of retinal abnormalities has been shown to correlate with lupus nephritis, CNS lupus, and the presence of antiphospholipid antibodies. Episcleritis and scleritis can occur in SLE. Uveitis is extremely rare. Discoid lupus can involve the lower eyelid and conjunctiva. Glucocorticoids and antimarial agents, two medications commonly used for the treatment of SLE, can affect the eye. Posterior subcapsular cataracts and elevated intraocular pressure are well-described complications of glucocorticoid therapy, and maculopathy is a rare but serious complication of the use of hydroxychloroquine and chloroquine. The risk of retinal toxicity is low if the daily dose of chloroquine is kept below 3 mg/kg or ideal body weight and the daily dose of hydroxychloroquine is kept at or below 6.5 mg/kg of ideal body weight.

**HEMATOLOGIC**

Hematologic involvement is common in SLE; all three blood cell lines can be affected. When evaluating a patient with the hematologic abnormalities as described later, it is always necessary to consider the potential of myelosuppression from medications such as methotrexate, azathioprine, mycophenolate mofetil, and cyclophosphamide. In addition, corticosteroids are a common cause of lymphopenia and leukocytosis secondary to neutrophilia.

**Anemia**

Anemia of chronic disease (ACD) is the most common anemia in SLE. It is a normochromic, normocytic anemia characterized by the presence of low serum iron, low transferrin, and normal to increased serum ferritin. ACD may coexist with anemias resulting from other processes. Autoimmune hemolytic anemia (AIHA) should be suspected in the setting of the following laboratory abnormalities: increased serum unconjugated bilirubin, increased lactate dehydrogenase (LDH), increased reticulocyte count, and reduced serum haptoglobin. The direct Coombs’ test is typically positive and usually is mediated by warm-reacting IgG anti-erythrocyte antibodies. The peripheral blood smear demonstrates spherocytosis. Some reports have suggested an association between AIHA and the presence of antiphospholipid antibodies. A positive direct Coombs’ test can occur without hemolysis. AIHA may be the presenting manifestation of SLE, or it may predate full-blown SLE by many years.

Microangiopathic hemolytic anemia (MAHA), characterized by the presence of schistocytes on peripheral blood smear, should prompt consideration of thrombotic thrombocytopenic purpura–hemolytic uremic syndrome (TTP-HUS). TTP is a syndrome consisting of MAHA, thrombocytopenia, fever, neurologic symptoms, and renal involvement, and may be associated with SLE. Because MAHA, thrombocytopenia, neurologic symptoms, and renal involvement can also occur in catastrophic antiphospholipid antibody syndrome (CAPS), antiphospholipid antibodies should always be measured as part of the evaluation. Blood loss, renal insufficiency, pure red cell aplasia, and medication-induced myelotoxicity are additional potential causes of anemia in SLE patients.

**Leukopenia**

Leukopenia occurs in approximately 50% of SLE patients and can occur secondary to lymphopenia and/or neutropenia. One study of 158 newly diagnosed, clinically active SLE patients demonstrated that 75% of patients had lymphocyte counts lower than 1500 cells/µL, and that lymphopenia eventually developed in 93% of patients. The presence of lymphocytotoxic antibodies in some SLE patients correlates with lymphopenia and with disease exacerbation. Lymphopenia may be a side effect of treatment with glucocorticoids or other immunosuppressive agents. Neutropenia due to SLE can result from immune-mediated destruction or marrow suppression.

**Thrombocytopenia**

Mild thrombocytopenia is noted in up to 50% of SLE patients, but severe thrombocytopenia can also occur. Thrombocytopenia can be the result of immune-mediated platelet destruction similar to immune thrombocytopenic purpura (ITP). The platelet IIb/IIa antigen is the primary target. Thrombocytopenia can also be caused by a consumptive process such as TTP or splenomegaly. Antithrombopoietin antibodies have been found in the sera of some SLE patients and have been correlated with lower platelet counts. Chronic, low-level thrombocytopenia is a characteristic feature of the antiphospholipid antibody syndrome. Similar to AIHA, isolated ITP may pre-date the development of complete SLE by several years.

**LYMPHADENOPATHY AND SPLENOMEGALY**

Lymphadenopathy commonly occurs in association with active SLE and is characterized by the presence of enlarged, soft, nontender lymph nodes. Lymphadenopathy can be focal or generalized; the cervical, axillary, and inguinal regions are typically involved. Lymph node histopathology demonstrates reactive hyperplasia and varying degrees of coagulative necrosis. The presence of hematoxylin bodies is specific for SLE. Histologic features of Castleman’s disease have been reported. The differential diagnosis of lymphadenopathy in an SLE patient includes infection and/or a
lymphoproliferative process; lymph node biopsy is sometimes required for diagnosis. Splenomegaly can be observed in patients with SLE and may be associated with hepatomegaly. Histopathologic studies demonstrate periarterial fibrosis (onion-skin lesions). Splenic atrophy and functional asplenism have also been reported.117

**DIAGNOSIS**

Establishing the diagnosis of SLE can be challenging because of its heterogeneous disease manifestations and waxing and waning clinical course. No clinical manifestation or laboratory test can serve as a definitive diagnostic test. Instead, SLE is diagnosed on the basis of a constellation of characteristic symptoms, signs, and laboratory findings in the appropriate clinical context. Although the ACR classification criteria (see Table 80-1) cannot always be relied upon for diagnostic purposes in individual patients, they serve as useful reminders of the wide variety of clinical features that can be seen in SLE.

**Serologic Tests**

Serologic tests play an important role in the diagnosis of SLE. SLE is the prototypic systemic humoral autoimmune disease. As such, it is characterized by production of a wide variety of autoantibodies, which often provide important diagnostic information118 (Table 80-7). The hallmark serologic feature is the presence of ANAs, as reflected by a positive ANA test. The gold standard method for detecting ANA is indirect immunofluorescence using a human epithelial cell tumor line (HEp2 cell line). With this method, the ANA test is highly sensitive in that it is positive in more than 95% of people with SLE. Because of a desire for automation and cost savings, some laboratories are utilizing the enzyme-linked immunosorbent assay (ELISA) as the method of testing for ANA. However, the ELISA method is less accurate than the immunofluorescence method, resulting in a higher false-negative rate. Positive ANA tests also occur in many other autoimmune diseases, including rheumatoid arthritis, scleroderma, polymyositis, and autoimmune thyroiditis, among others. ANAs are also detectable in low titers (<1:80) in many people without autoimmune disease, especially in the elderly.119 Therefore, a positive test is not sufficient to establish the diagnosis of SLE. On the other hand, a negative test can be helpful in ruling out SLE. Although “ANA-negative” SLE has been reported, it is very rare with the immunofluorescence method of testing. In those rare instances, other tests (e.g., anti-Ro/SSA) confirm the presence of lupus-associated autoantibodies.120

Once it has been established that ANAs are present, it is important to determine which particular nuclear antigens may be the target of the autoantibodies, because some of these antigen-specific responses provide great diagnostic specificity. The most important of these tests is the test for antibodies to double-stranded DNA (anti-dsDNA). Anti-dsDNA antibodies are present in no more than 50% to 60% of patients with lupus, so their absence does not exclude the possibility of SLE. However, the presence of these antibodies is highly specific for SLE and therefore can be very helpful in establishing a definitive diagnosis. Similarly, antibodies to the Sm antigen have great specificity for SLE, but these antibodies are present in even fewer SLE patients (=30%). The Sm antigen is a component of extractable nuclear antigens (ENAs), a term that refers to a heterogeneous mixture of non-DNA nuclear antigens that can be “extracted” from cells in the laboratory. These antigens are primarily ribonucleoproteins that can be divided into two major subsets based on their susceptibility to digestion by ribonuclease. The ribonuclease-sensitive antigens are designated RNP. Ribonuclease-resistant antigens are designated Sm (because these antibodies were first detected in a patient named Smith). Unlike Sm, anti-RNP is not specific for SLE. However, high titers of anti-RNP antibodies can be helpful in supporting the diagnosis of MCTD.121

Numerous other autoantibodies can be found in patients with SLE. Antibodies to cytoplasmic antigens, such as Ro and La (SSA and SSB), can be found in some patients with SLE. Although these autoantibodies lack both sensitivity and specificity for SLE, they sometimes are associated with distinct clinical syndromes. The best example of such a relationship involves the presence of anti-Ro antibodies in more than 90% of cases of neonatal lupus.122,123 Anti-Ro antibodies are also seen with increased frequency in patients with SLE.124 Other autoantibodies may be directed against cell surface molecules or against circulating proteins. For example, antibodies to blood components can be responsible for hemolytic anemia, neutropenia, or thrombocytopenia, and antibodies to phospholipids can be detected in some SLE patients with or without antiphospholipid syndrome (see Chapter 82). It should be noted that rheumatoid

**Table 80-7** Autoantibodies and Clinical Significance in Systemic Lupus Erythematosus (SLE)

<table>
<thead>
<tr>
<th>Autoantibody</th>
<th>Prevalence in SLE</th>
<th>Clinical Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antinuclear Antibody</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-dsDNA</td>
<td>60%</td>
<td>95% specificity for SLE; fluctuates with disease activity; associated with glomerulonephritis</td>
</tr>
<tr>
<td>Anti-Smith</td>
<td>20%-30%</td>
<td>99% specificity for SLE; associated with anti-U1RNP antibodies</td>
</tr>
<tr>
<td>Anti-U1RNP</td>
<td>30%</td>
<td>Antibody associated with mixed connective tissue disease and lower frequency of glomerulonephritis</td>
</tr>
<tr>
<td>Anti-Ro/SSA</td>
<td>30%</td>
<td>Associated with Sjögren’s syndrome, photosensitivity, SCLE, neonatal lupus, congenital heart block</td>
</tr>
<tr>
<td>Anti-La/SSB</td>
<td>20%</td>
<td>Associated with Sjögren’s syndrome, SCLE, neonatal lupus, congenital heart block, anti-Ro/SSA</td>
</tr>
<tr>
<td>Antihistone</td>
<td>70%</td>
<td>Also associated with drug-induced lupus</td>
</tr>
<tr>
<td>Antiphospholipid</td>
<td>30%</td>
<td>Associated with arterial and venous thrombosis, pregnancy morbidity</td>
</tr>
</tbody>
</table>

SCLE, subacute cutaneous lupus erythematosus.
factor (anti-IgG) can be found in 15% to 20% of people with SLE, whether or not joint disease is present. Anti-CCP antibodies can also be present.

Complement consumption arising from immune complex disease may lead to hypocomplementemia in patients with SLE. Because hypocomplementemia is rare in other diseases, its presence in a patient with SLE can provide valuable supportive evidence for the diagnosis. Moreover, hypocomplementemia most likely reflects complement activation by immune complexes, its presence is often a sign of active disease. However, hereditary complement deficiencies may be found in patients with SLE (C1q, C2, C4), so absence of a particular complement component does not always reflect consumption. For this reason, it is often necessary to measure more than one complement component (e.g., C3 and C4) before concluding that hypocomplementemia is due to active disease.

The utility of serologic tests in assessing disease activity and predicting disease flares remains a topic of controversy. In the absence of a hereditary complement deficiency, hypocomplementemia is a reliable indicator that the disease is active, but normal complement levels do not rule out active disease. Titters of anti-dsDNA antibodies correlate with disease activity in some patients, but not in others. One recent study attempted to resolve the long-standing debate about the prognostic value of changes in lupus serology. In this study, patients with clinically quiescent lupus underwent monthly monitoring for levels of anti-dsDNA, C3a, C3, C4, and CH50 to identify patients with serologically active, clinically quiescent disease. These patients were then randomized to treatment with corticosteroids or placebo to determine whether treatment of serologically active disease could prevent impending clinical flares. The results were equivocal. Some patients with serologically active disease flared, and some flares were apparently prevented. However, in most control subjects, serologic deterioration was not followed by a clinical flare, and most of the flares that occurred in the original patient population that had been monitored were not preceded by serologic deterioration. Thus, there remains no substitute for knowledge of a particular patient's pattern or whether there is an association between clinical and serologic manifestations in that patient.

DIFFERENTIAL DIAGNOSIS

Because of the involvement of multiple organ systems and the lack of specificity of symptoms and/or signs, many systemic diseases can mimic SLE. Thus, before a diagnosis of SLE is established, a comprehensive search for infectious, malignant, and other autoimmune diseases must be undertaken.

Several viral infections can produce symptoms and signs that are present in SLE. In addition, many viral illnesses are associated with the production of autoantibodies. A careful patient history with serologic testing for the potential pathogen should help to secure the correct diagnosis. Parvovirus B19 classically presents with fever, rash, symmetric inflammatory polyarthritis, and cytopenias. Furthermore, the presence of ANA, anti-dsDNA, and hypocomplementemia has been observed in a few cases. Cytomegalovirus and Epstein-Barr virus can mimic SLE in that patients often present with fatigue, cytopenias, abdominal pain, and liver test abnormalities. Acute HIV infection typically presents with fever, diffuse lymphadenopathy, and oral ulcers. Patients with hepatitis B and C can develop inflammatory arthritis and positive autoantibodies.

Malignancy, particularly non-Hodgkin's lymphoma, can manifest with constitutional symptoms, joint pain, cytopenias, lymphadenopathy, rash, and a positive ANA. One must be particularly alert to the possibility of malignancy in an older patient presenting with a new lupus-like syndrome. It is important to ensure that patients undergo appropriate malignancy screening tests.

Other autoimmune diseases such as RA, dermatomyositis, and Still's disease often share similar clinical features with SLE. Differentiating between these disorders might be difficult in early phases of the disease. Patients with RA and SLE may develop a symmetric inflammatory arthritis with a predilection for the wrists and small joints of the hands. ANA and rheumatoid factor (RF) may be elevated in both disorders, although anti-CCP antibody suggests RA, and anti-dsDNA or anti-Sm suggests SLE. The photosensitive, erythematosus rashes of dermatomyositis and SLE can appear clinically and histopathologically identical. A careful patient history and supporting serologies will aid in making the correct diagnosis. Mixed connective tissue disease (MCTD) must also be considered when evaluating a patient for possible SLE. MCTD is a syndrome characterized by a high-titer anti-RNP antibody in conjunction with clinical features that are often present in SLE, scleroderma, and/or polymyositis. Patients frequently present with puffy, swollen hands and Raynaud's phenomenon. In contrast to SLE, patients with MCTD can develop an erosive arthritis that looks very similar to RA.

Careful evaluation for drug-induced lupus should be undertaken in every new patient in whom the diagnosis of SLE is suspected. This is especially important in an older person presenting with a lupus-like syndrome. Arthralgia, myalgia, fever, and serositis are common manifestations. A wide variety of drugs have been implicated in the development of drug-induced lupus; minocycline, procainamide, hydralazine, isoniazid, interferon alpha, and anti-TNF agents are well-known culprits. Hydroxychloroquine is associated with SLE. All of these drugs may cause a positive ANA. Minocycline is occasionally associated with anti-dsDNA antibodies and perinuclear-staining antineutrophil cytoplasmic antibodies (P-ANCA), and anti-TNF agents can cause positive anti-dsDNA antibodies. Antihistone antibodies are present in more than 95% of cases of drug-induced lupus, with the exception of those cases caused by minocycline. However, antihistone antibodies cannot be used to confirm a diagnosis of drug-induced lupus because up to 80% of idiopathic SLE patients will also produce antihistone antibodies.

NEONATAL LUPUS

Neonatal lupus is a passively acquired autoimmune disease of neonates that results from transplacental passage of maternal anti-SSA and/or anti-SSB antibodies. It can occur in mothers with SLE, in those with Sjögren's syndrome, or in patients in whom an autoimmune disease has not been diagnosed. Neonatal lupus can involve multiple
organ systems, including heart, skin, liver, and the hematologic system; the most severe complications are congenital complete heart block and cardiomyopathy. The term neonatal lupus stems from early observations that the skin lesions in affected newborns were similar to the lesions of SCLE.

Congenital complete heart block is associated with a neonatal mortality rate as high as 20%, and most patients will eventually require a permanent pacemaker. This complication occurs in up to 2% of babies born to mothers who are positive for anti-Ro(SSA) and/or anti-La(SSB) antibodies. Once a woman has given birth to a baby with complete heart block, the risk for recurrence in a subsequent pregnancy is approximately 15%. Evidence from in vitro studies suggests that during fetal development, fetal cardiocytes undergo apoptosis that results in expression of Ro(SSA) and La(SSB) on the cell surface. Binding of anti-Ro(SSA) and/or anti-La(SSB) to fetal cardiocytes leads to inflammatory injury and subsequently to fibrosis of the atrioventricular (AV) node and surrounding tissue. The sinoatrial (SA) node may also be involved. Prospective studies have shown that the vulnerable period for the fetal heart is between 16 and 24 weeks of gestation. Thus, it is recommended that mothers with anti-Ro(SSA) and/or anti-La(SSB) antibodies undergo monitoring with fetal echocardiography beginning at 16 weeks' gestation. The hope has been that detection of early stages of heart block (first-degree and second-degree block) might allow for treatment that would prevent progression to third-degree heart block. Currently, the treatment of choice is maternal administration of a fluorinated glucocorticoid such as dexamethasone. Fluorinated glucocorticoids are preferred because of their ability to cross the placenta and enter the fetal circulation. However, treatment of incomplete fetal heart block remains controversial because the benefits have not been clearly delineated, and glucocorticoids have been associated with several fetal side effects such as intrauterine growth retardation, oligohydramnios, and adrenal suppression. Complete heart block is irreversible even with treatment, and first- or second-degree heart block may or may not reverse with treatment. Complicating matters, complete heart block can occur in the absence of preceding first- or second-degree block. In addition to conduction blocks, structural cardiac abnormalities have been observed in the setting of neonatal lupus, including, but not limited to, patent ductus arteriosus, ventricular septal defect, atrial septal defect, and patent foramen ovale. Myocarditis and pericarditis have also been described.

Rash, a common manifestation of neonatal lupus, consists of erythematous, annular lesions that resemble the annular subtype of SCLE. The rash typically occurs on the scalp, face, trunk, and extremities with a predilection for the periorbital area; it often develops after exposure of the newborn to ultraviolet light. Lesions typically occur within the first 4 to 6 weeks of life but may be present at birth. The rash is self-limiting and does not necessitate treatment. Lesions tend to resolve by 6 months of age, at which time maternal anti-Ro(SSA) and/or anti-La(SSB) antibodies are no longer present in the baby's circulation. Less common manifestations of neonatal lupus include hepatic, hematologic, and neurologic involvement. Hepatic manifestations include asymptomatic elevation of liver function tests, hepatitis, hepatomegaly, cholestasis, and cirrhosis. Hematologic manifestations include thrombocytopenia, autoimmune hemolytic anemia, and leukopenia. Neurologic complications, including myelopathy, seizures, and aseptic meningitis, have been reported.

Selected References


References


CHAPTER 80 | CLINICAL FEATURES OF SYSTEMIC LUPUS ERYTHEMATOSUS


**Treatment of Systemic Lupus Erythematosus**

GEORGE BERTSIAS • ANTONIOS FANOURIAKIS • DIMITRIS T. BOUMPAS

The management of systemic lupus erythematosus (SLE) is challenging due to the clinical heterogeneity and the unpredictable course of the disease. SLE activity usually follows the flare pattern, which is characterized by a relapsing-remitting course. However, an equal number of patients have continuously active disease, and only a few have long periods of disease quiescence. Despite improvements in overall survival (85% to 90% during the first 10 years), some SLE patients are still at risk for premature death. Persistent inflammation inevitably results in irreversible major organ damage, which is linked to decreased quality of life and increased mortality. Accordingly, therapeutic strategies should aim at reducing overall burden of systemic inflammation. Achieving these goals requires (1) accurate assessment of disease activity and flares, (2) stratification according to severity of target organ involvement, (3) use of safe and effective drugs to induce remission promptly and prevent flares, and (4) prevention and management of disease and treatment-related comorbidities.

In general, patients with mild lupus manifestations (skin, joint, and mucosal involvement) are treated with antimalarials or disease-modifying antirheumatic drugs (DMARDs), alone or in combination with low-dose oral glucocorticoids (GCs). Severe SLE with major organ involvement requires an initial period of intensive immunosuppressive therapy (induction therapy) to control aberrant immunologic activity and halt tissue injury, followed by a longer period of less intensive and less toxic maintenance therapy, to consolidate remission and prevent flares. Immunosuppressive therapy enables for the use of lower GC doses, thus reducing its deleterious effects.

SLE patients experience poor quality of life, which is only in part associated with disease activity and organ damage. Important contributors include fatigue, fibromyalgia, depression, and cognitive dysfunction. Treating physicians should regularly address these issues and engage symptomatic or remedial therapies as indicated. The realization that a significant proportion of patients features disease- and treatment-related comorbidities has shifted attention to adjunct therapies and primary prevention strategies such as renoprotective and cardiovascular disease risk reduction measures.
Health professionals are increasingly encouraged to involve patients in treatment decisions, recognizing them as experts with a unique knowledge of their own health and their preferences for treatments, health states, and outcomes. This approach may be particularly challenging in patients with severe lupus who may benefit from cytotoxic therapy, which is, however, associated with significant toxicity. This was demonstrated in a study of 93 well-educated women with mild lupus and good health status who were presented with descriptions of cyclophosphamide (CYC) and azathioprine (AZA) and were then asked to indicate their preferred choice of hypothetical treatment. The study patients had a strong preference for full disclosure of medication risks and treatment alternatives, and they preferred a collaborative role in decision making involving their health care. Nearly all (98%) participants chose AZA over CYC when both drugs conferred an equal probability of renal survival. Although most subjects switched preferences to CYC for better renal survival, a significant proportion (15% to 31%) were unwilling to switch to CYC for improved short- or long-term renal survival. Preference for the long-term benefits of CYC was greater among college than high school graduates, which could reflect a better understanding of the probabilities presented in the study among those with a higher education level. This underscores the importance of providing patients with tailored information to ensure that all patients accurately perceive the risks and benefits related to prescribed medications.

On the other hand, the relative paucity of high-quality evidence to guide therapeutic decisions for various disease manifestations has been a cause for practice pattern variation among lupus experts. Many opinions about the treatment of lupus are based on personal perception rather than valid observation. To facilitate physicians who are in the care of these patients, the European League Against Rheumatism (EULAR) Task Force on SLE has developed management recommendations based on evidence and expert opinion (see Evidence and Expert-based Recommendations in SLE later).

**DRUGS USED IN THE TREATMENT OF SYSTEMIC LUPUS ERYTHEMATOSUS**

**KEY POINTS**

The target dose of glucocorticoids should be 0.25 mg/kg every other day for 2 to 3 months, which is acceptable for long-term use. Concomitant use of immunosuppressive agents facilitates tapering and decreases cumulative toxicity.

Antimalarials are effective for skin and mucocutaneous manifestations. Their use has been associated with reduced organ damage accrual.

Azathioprine (AZA) is effective as an induction and a maintenance regimen in mild to moderate SLE including nephritis.

The combination of high-dose glucocorticoids with pulses of intravenous CYC remains the standard of care for severe SLE with major organ involvement.

**MMF is at least equally efficacious and has a better toxicity profile than CYC in the treatment of moderately severe PLN.**

**Calcineurin inhibitors (cyclosporin A, tacrolimus), used alone or in combination with other immunosuppressive agents, have demonstrated efficacy in refractory to cytotoxic therapy lupus nephritis.**

**Based on evidence from uncontrolled studies, the American College of Rheumatology and the European League Against Rheumatism–European Renal Association have included rituximab (anti-CD20 mAb) as a therapeutic option for selected cases of lupus nephritis refractory to conventional immunosuppressive treatment.**

**There is an unprecedented array of promising biologic therapies currently in development in SLE.**

**Glucocorticoids**

GCs exert broad inhibitory effects on immune responses mediated by T and B cells, as well as on the effector functions of monocytes and neutrophils. On the basis of these effects and their rapid onset of action, GCs have been remarkably efficacious in managing acute SLE manifestations. However, only two small randomized controlled trials (RCTs) have been conducted to demonstrate their efficacy in lupus.

Low-dose GC therapy (≤7.5 mg prednisone equivalent per day) is generally used when other initial therapies (antimalarials) are not tolerated or are inadequate to control disease activity. In moderate to severe disease, GCs are used as either single or background therapy in combination with immunosuppressive agents, at doses 0.5 to 1 mg/kg prednisone equivalent in a single dose usually in the morning. When combined with immunosuppressive agents, the GC dose should rarely exceed 0.5 to 0.6 mg/kg prednisone due to concerns for infections and other toxicities. Tapering of GC dose starts after the first 4 to 6 weeks of therapy, targeting a dose of 0.25 mg/kg every other day at 2 to 3 months, which is acceptable for long-term use. Concomitant use of immunosuppressive agents facilitates tapering and decreases cumulative GC toxicity.

Tseng and colleagues examined the effect of a short course of moderate-dose GC in preventing flares in clinically stable but serologically active SLE. Severe flares, requiring increase in prednisone dose or addition of an immunosuppressive agent, or both, occurred in 6 out of 20 patients on placebo, as compared with none of the 21 patients who took prednisone (30 mg for 2 weeks, 20 mg for 1 week, and 10 mg for 1 week). The results agree with those of Bootsma and colleagues and suggest that in clinically stable but serologically active SLE patients, short-term GC therapy may avert a severe flare. This effect must be balanced against risks for overtreating patients with higher cumulative GC doses.

In severe, rapidly progressing disease or when doses greater than 0.6 mg/kg/day prednisone equivalent are required to control disease activity, GC pulse therapy may be introduced. Intravenous (IV) pulses of methylprednisolone (MP) (250 to 1000 mg daily for 1 to 3 consecutive days) are used, although there is no strong evidence to support a survival advantage. In addition to expediting
remission, IV-MP pulses may allow for the use of lower GC doses during the induction period. In a trial from the National Institutes of Health (NIH), 82 patients with moderate to severe proliferative lupus nephritis (PLN) were randomized to receive sequential induction-maintenance therapy with IV-MP alone, IV-MP plus IV-CYC, or IV-CYC alone. Renal remission was significantly more common in both IV-CYC groups regardless of whether or not they received IV-MP. In the extended follow-up of the trial cohort, an analysis of protocol completers found that doubling of serum creatinine (SCr) was significantly lower in the combination group than in the IV-CYC group (relative risk [RR] 0.095). These findings indicate that combining IV-CYC with IV-MP in the treatment of lupus nephritis (LN) may confer an advantage in long-term renal outcomes without added toxicity.

GC toxicity may involve early (mood effects, acne, myalgias, infections); later (metabolic); and late-onset (osteoarthritis, avascular bone necrosis, cataract, cardiovascular disease) adverse events. Although confounded by increased systemic disease activity, most effects are considered to be dependent on the cumulative dose and duration and, to a lesser extent, route of administration. Tailoring GC use on the basis of individual patient profile (disease activity, risk for toxicity) and prompt initiation of primary prevention measures (antiosteoporotic treatment) is of paramount importance.

**Antimalarials and DMARD Therapy**

**Hydroxychloroquine**

Antimalarial drugs, mainly chloroquine and hydroxychloroquine (HCQ), are commonly prescribed to SLE patients with skin and joint manifestations but are increasingly identified as an adjuvant treatment for achieving remission in severe lupus. A systematic review concluded that use of antimalarials resulted in a greater than 50% reduction in general SLE disease activity and to a moderate reduction in severe flares and GC dose. Beneficial effects on the lipid profile and subclinical atherosclerosis markers have also been described. Intriguingly, prospective observational studies have reported an inverse association between use of HCQ and accrual of irreversible organ damage, as measured using the Systemic Lupus International Collaborating Clinics–American College of Rheumatology Damage Index (SDI) (adjusted hazard ratio [HR], 0.73), and overall mortality rates (adjusted HR, 0.14 to 0.32). Although these findings may be confounded by milder forms of disease usually present in SLE patients who are prescribed antimalarials, the presumptive mode of action of antimalarials by inhibition of innate immunity pathways provides a plausible explanation for their multifld beneficial effects. HCQ is well tolerated with low rates of mild gastrointestinal and skin adverse events. Retinal toxicity is uncommon (estimated at 0.1% in patients who received HCQ for more than 10 years), but routine ophthalmologic evaluation is recommended (annual screening during the first 5 years of usage is recommended for individuals who are treated with high HCQ dose [>6.5 mg/kg], for those treated more than 5 years, or for those who have other complicating factors).

**Methotrexate**

Methotrexate (MTX), an antifolate agent commonly prescribed for rheumatoid arthritis (RA), has been used as steroid-sparing treatment for articular and cutaneous manifestations of SLE. MTX is administered weekly either orally or parenterally. Concomitant administration of folic acid (2.5 to 5 mg/week, not until 24 hours after the intake of MTX) is recommended to minimize toxicity (Table 81-1). Fortin and colleagues evaluated the efficacy of MTX in a 12-month placebo-controlled RCT in 86 SLE patients. Patients had mild to moderate disease, with musculoskeletal (93%), cardiovascular (74%), and hematologic (69%) manifestations. Approximately half of the patients in each group were on oral prednisone, whereas 41 patients in the placebo group versus 27 in the MTX group were on antimalarials. Among participants with comparable baseline prednisone dose, those on MTX received on average 1.33 mg/day less prednisone during the trial period compared with those in the placebo group. Fewer patients in the MTX group were also started on GCs (5% vs. 26% in the placebo group). MTX use was associated with a reduction in the mean during-trial Systemic Lupus Activity Measure (SLAM) score of 0.86 units. Together, these data suggest that MTX could be a reasonable alternative steroid-sparing agent in mild to moderate SLE.

**Leflunomide**

Leflunomide, currently used in RA, has been used in LN refractory or intolerant to standard immunosuppressive therapy. Leflunomide requires a loading dose of 100 mg/day for 3 days followed by 20 mg/day thereafter. In a multicenter observational study, 110 patients with biopsy-proven PLN were assigned to either oral leflunomide or IV-CYC (monthly pulses 0.5 g/m²), both in combination with oral prednisone (0.8 mg/kg/day for 4 weeks, then tapered). After 6 months, complete and partial remission rates were 21% and 52% in the leflunomide group and 18% and 55% in the IV-CYC group, respectively. Repeat kidney biopsies showed significant reduction in activity index and pathologic transformation of 10 cases of diffuse to focal PLN. Similar rates of adverse events were observed in the two study groups and included mostly herpes zoster infection, alopecia, and hypertension. Better-designed RCTs are necessary to establish the efficacy, if any, of leflunomide in PLN. The drug is teratogenic and is contraindicated in patients who are trying to become or are pregnant.

**Cytotoxic Therapy**

**Cyclophosphamide**

Pharmacology and Route of Administration. CYC is an alkylating agent that depletes both T and B cells and reduces the production of autoantibodies in lupus. Both oral and IV administration of CYC result in similar plasma concentrations, and the serum half-life is approximately 6 hours. CYC is metabolized to various active metabolites by cytochrome P-450 in the liver or other tissues such as transitional epithelial cells of the bladder or lymphocytes. Drugs that induce hepatic microsomal enzymes (barbiturates,
alcohol, phenytoin, rifampin) may accelerate the metabolism of CYC into its active metabolites and thus increase its pharmacologic and toxic effects. Conversely, drugs that inhibit the hepatic microsomal enzymes (antimalarials, tricyclic antidepressants, and allopurinol) may slow the conversion of CYC to active metabolites. Approximately 20% of the drug is excreted by the kidney, whereas 80% is processed by the liver. Dose modification is necessary for patients with renal impairment but not in liver disease.

The NIH trials demonstrated equivalent efficacy yet lower toxicity with monthly IV (0.5 to 1 g/m²) versus oral regimens, which led to the predominant use of IV-CYC in clinical practice (Table 81-4). Reversible myelotoxicity is common and dose related. After pulse therapy, the nadir of

<table>
<thead>
<tr>
<th>Drug</th>
<th>Dosage</th>
<th>Dose Adjustment</th>
<th>Toxicities Requiring Monitoring</th>
<th>Baseline Evaluation</th>
<th>Laboratory Monitoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azathioprine</td>
<td>50-100 mg/day in 1-3 doses with food</td>
<td>↓ 25% if eGFR 10-30 mL/min; ↓ 50% if eGFR &lt; 10 mL/min</td>
<td>Myelosuppression, hepatotoxicity, lymphoproliferative diseases</td>
<td>CBC, platelets, Scr, AST or ALT</td>
<td>CBC and platelets every 2 wk with changes in dosage; baseline tests every 1-3 mo</td>
</tr>
<tr>
<td>Mycophenolate mofetil</td>
<td>1-3 g/day in 2 divided doses with food</td>
<td>Maximum 1 g/day if eGFR &lt; 25 mL/min</td>
<td>Myelosuppression, hematotoxicity, infection</td>
<td>CBC, platelet, Scr, AST or ALT</td>
<td>CBCs and platelets every 1-2 wk with changes in dosage; baseline tests every 1-3 mo</td>
</tr>
<tr>
<td>Cyclophosphamide</td>
<td>50-150 mg/day in a single dose with breakfast. Lots of fluids (at least 3 L water/day), empty bladder before bedtime</td>
<td>↓ 25% if eGFR 25-50 mL/min; ↓ 30-50% if eGFR &lt; 25 mL/min; ↓ 25% if serum Bil 3.1-5 mg/dL or transaminases &gt;3 times ULN</td>
<td>Myelosuppression, hemorrhagic cystitis, myeloproliferative disease, malignancies</td>
<td>CBC, platelet, Scr, AST or ALT, urinalysis</td>
<td>CBC with differential every 1-2 wk, with changes in dosage and then every 1-3 mo; keep WBC &gt; 4000/mm³ with dose adjustment; urinalysis, AST or ALT every 3 mo; urinalysis every 6-12 mo following cessation</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>7.5-15 mg/wk in 1-3 doses with food or milk/water</td>
<td>↓ 50% if eGFR 10-50 mL/min; avoid use if eGFR &lt; 10 mL/min; avoid use in hepatic dysfunction (serum Bil 3.1-5 mg/dL or transaminases &gt;3 times ULN)</td>
<td>Myelosuppression, hepatic fibrosis, pneumonitis</td>
<td>CxR, hepatitis B/C serology in high-risk patients, AST or ALT, SAlb, ALP, Scr</td>
<td>CBC with platelet, AST, SAlb, Scr every 1-3 mo</td>
</tr>
<tr>
<td>Cyclosporin A</td>
<td>100-400 mg/day in 2 doses at the same time every day with meal or between meals</td>
<td>Avoid in impaired renal function</td>
<td>Renal insufficiency, anemia, hypertension</td>
<td>CBC, SCr, uric acid, AST or ALT, SAlb, ALP, blood pressure</td>
<td>CBC every 2 wk until dose is stable, then monthly; CBC, potassium, AST or ALT, SAlb, and ALP every 1-3 mo; drug levels only with doses &gt; 3 mg/kg/day</td>
</tr>
<tr>
<td>Tacrolimus</td>
<td>1-3 mg/day in 2 doses at the same time every day</td>
<td>Cautious use in liver or renal insufficiency</td>
<td>Renal insufficiency, neurotoxicity, malignancy, infections, hyperkalemia</td>
<td>SCr, potassium, AST or ALT, glucose, blood pressure</td>
<td></td>
</tr>
<tr>
<td>Leflunomide</td>
<td>100 mg/day in a single dose for 3 days, then 10-20 mg/day</td>
<td>Avoid in hepatic dysfunction (serum Bil 3.1-5 mg/dL or transaminases &gt;3 times ULN)</td>
<td>Myelosuppression, hepatotoxicity, fatal toxicity</td>
<td>CBC, SCr, AST or ALT, SAlb, ALP</td>
<td></td>
</tr>
<tr>
<td>Rituximab</td>
<td>1000 mg on day 1 and 15</td>
<td>None</td>
<td>HBV reactivation (rare)</td>
<td>CBC, SCr, AST or ALT, HBV serology (high-risk patients), TST</td>
<td></td>
</tr>
</tbody>
</table>

ALP, alkaline phosphatase; ALT, alanine transaminase; AST, aspartate transaminase; Bil, bilirubin; CBC, complete blood cell count; CxR, chest x-ray; eGFR, estimated glomerular filtration rate; HBV, hepatitis B; LFTs, liver function tests; MTX, methotrexate; SAlb, serum albumin; SCr, serum creatinine; TST, tuberculin skin testing; ULN, upper limit of normal; WBC, white blood cell count.
lymphocyte count occurs on days 7 to 10 and that of granulocyte count on days 10 to 14. The risk of infection increases with a white blood cell count less than 3000 cells/mm³, so the dose should be adjusted to keep it above this level (see Table 81-1). A prompt recovery from granulocytopenia usually occurs after 21 to 28 days. Thrombocytopenia is rare in CYC monotherapy. Reversible alopecia and nausea are common, whereas infections (especially herpes zoster), gonadal toxicity, and malignancy (including bladder toxicity and carcinoma) are less frequent though much more serious adverse events (see Comorbidities in Systemic Lupus Erythematosus and Women’s Health Issues later).

**Use in Lupus Nephritis.** RCTs with long-term follow-up have shown that intermittent pulse IV-CYC therapy is effective for moderate to severe PLN.³⁵³ CYC may retard progressive renal scarring, preserve renal function, and reduce the risk for the development of end-stage renal disease (ESRD) requiring dialysis or renal transplantation. Following induction therapy, a maintenance regimen is necessary to decrease the risk of flares.³⁴ NIH studies have demonstrated that combination pulse therapy with IV-CYC and IV-MP improves renal outcomes without increasing toxicity.¹⁸,¹⁹ On the basis of these studies, the authors propose 7 monthly pulses of IV-CYC (0.5 to 1 g/m²) followed by quarterly pulses for at least 1 year beyond remission. For patients with moderate to severe disease, monthly pulses of IV-MP are given during the induction period.

Because of toxicity concerns and the appreciation that the disease may be less severe in whites, European investigators sought alternative IV-CYC protocols. In the EuroLupus Nephritis Trial involving mostly patients with milder forms of disease (mean SCr, 1.2 mg/dL; mean proteinuria, 3 g/day), less intensive regimens of IV-CYC (6 fortnightly pulses at a fixed dose of 500 mg each in combination with three daily doses of 750 mg of IV-MP) followed with AZA as maintenance had comparable efficacy and less toxicity than high-dose IV-CYC (8 pulses).³⁵ Mean survival rate at 10 years was 92% for both groups; ESRD and doubling of SCr rates did not differ between the two groups.³⁶ Therefore, low-dose IV-CYC may be an alternative option for white patients with moderately severe LN.

Austin and colleagues³⁷ compared cyclosporin A (5 mg/kg/day, then adjusted according to changes in SCr), IV-CYC (0.5 to 1 g/m² × 6 monthly doses), and GC alone in an RCT in 42 patients with lupus membranous nephropathy (LMN) (median GFR, 83 ml/min/1.73 m²; median proteinuria, 5.4 g/day). All patients received alternate-day oral prednisone (1 mg/kg every other day for 5 weeks, then tapered to 0.25 mg/kg every other day). At 1 year, the cumulative probability of remission was 27% with prednisone, 60% with IV-CYC, and 83% with CsA. Rates of nephrotic syndrome relapse per 100 patient-months were 2.0 with CsA versus 0.2 with IV-CYC. Thus although IV-CYC and CsA are equally effective as induction therapy in LMN, CsA may require maintenance therapy (with lower doses of CsA, or AZA, or MMF) to prevent relapses.

**Use in Extrarenal Disease.** CYC has demonstrated efficacy in life-threatening extrarenal lupus manifestations such as severe thrombocytopenia (platelet count <20,000/mm³), neurologic disease, abdominal vasculitis, acute pneumonitis/alveolar hemorrhage, and extensive skin disease.³⁸,³⁹ Barile-Fabris and colleagues⁴⁰ have reported superiority of IV-CYC against IV-MP for severe nonthrombotic neurologic SLE. In this trial, 32 patients were randomized to receive 3 pulses of 1 g of IV-MP followed by one of the following two treatments: pulses of 1 g IV-MP (monthly for 4 months, then every 2 to 3 months for 1 year) or IV-CYC (0.75 g/m² monthly for 1 year and then every 3 months for another year). Seizures were the most common syndrome; other manifestations included peripheral neuropathy, optic neuritis, transverse myelitis, brain stem disease, coma, and internuclear ophthalmoplegia. Clinical response was observed in 18 of 19 patients who received IV-CYC as compared with 7 of 13 who received IV-MP. Thus the combination of pulses of IV-MP with IV-CYC is considered as treatment of choice for severe inflammatory neurologic SLE.

**Other Agents**

**Chlorambucil.** Chlorambucil (CAB) is an aromatic alkylating agent with substitution of the N-methyl group of mechloretamine with phenylbutyric acid. The drug is given orally (0.1 to 0.2 mg/kg/day) with good absorption. The effects on immune functions are comparable with those described for CYC. Adverse events are also similar to those of CYC, except for bladder toxicity. More prolonged and less predictable bone marrow suppression can be observed. CAB has been associated with increased risk for leukemia. There is limited experience with the use of CAB in SLE, yet favorable outcomes in renal and extrarenal manifestations such as neuropsychiatric, vasculitis, and multiorgan involvement have been reported. In a retrospective study of 19 patients with predominantly LMN, Moroni and colleagues⁴⁰ showed that CAB combined with alternate-month cycles of IV-MP was more effective than IV-MP alone in inducing remission of nephrotic syndrome (64% vs. 38%) and preserving renal function over a period of 83 months. However, the use of CAB for LMN or other lupus manifestations is limited.

**Fludarabine.** Fludarabine induces profound immunosuppression by depleting T and B lymphocytes and is used in the treatment of hematologic malignancies. A single pilot study in patients with active PLN was terminated prematurely due to severe hematologic toxicity.³⁷ It is unlikely that fludarabine will be used in SLE.

**Antimetabolites Calcineurin Inhibitors**

**Azathioprine**

AZA interferes with the de novo synthesis of inosinic acid and inhibits the conversion of purine bases such as inosine to adenine and guanine ribonucleotides. AZA in doses of 2 to 2.5 mg/kg/day has been remarkably safe in the long term without significantly increasing the risk for infection, whereas it is associated with a marginally increased risk for malignancy. Gastrointestinal complaints are frequent, leading 15% to 30% of patients to discontinue the drug within 6 months. Mild liver enzyme elevation may occur, but severe liver injury is rare. Reversible, dose-related bone marrow toxicity is also common; leukopenia is encountered in approximately 4.5% and thrombocytopenia in 2% of patients receiving low-dose AZA (see Table 81-1). Notably,
AZA toxicity is idiosyncratic and has been associated with genetic polymorphisms resulting in decreased thiopurine methyltransferase (TPMT) activity and impaired ability to detoxify intermediate metabolites. Concomitant use of allopurinol substantially increases AZA toxicity and should be avoided.

In lupus, manifestations such as mild PLN, thrombocytopenia with platelet count in the range of 20 to $50 \times 10^9$ mm$^{-3}$, and serositis may respond to AZA, usually in combination with moderate to high GC doses (Tables 81-2 and 81-3). Its efficacy as an induction-maintenance regimen has been tested in low-risk European patients with PLN who were randomized to receive pulse IV-CYC plus prednisone versus IV-MP plus AZA plus a tapering dose of prednisone. After 2 years, both groups received maintenance therapy with AZA plus prednisone. The two groups did not differ in terms of induction of remission, mean SCr and proteinuria levels during the first 2 years. However, after a median follow-up duration of 5.7 years, rates of doubling of baseline SCr and of renal relapses were higher in the AZA group. Thus the authors believe that AZA may be used in mild LN and in patients strongly opposed to CYC.

AZA has also been considered a safe and efficacious option for maintenance of remission in SLE including cases of moderately severe PLN. Two RCTs have compared AZA versus mycophenolate mofetil (MMF) as maintenance regimens in PLN. In the MAINTAIN trial, which included only European patients, both agents were equally efficacious in terms of time-to-renal flare, number of severe flares, renal remission, and doubling of SCr. In contrast, the ALMS trial, which included a larger number of patients with multietnic backgrounds, reported increased renal flares in the AZA versus the MMF group (see Mycophenolate Mofetil later). To this end, both agents can be used as maintenance therapy on the basis of availability, clinical experience, and potential for pregnancy because MMF is associated with an increased risk of spontaneous abortion and fetal malformation. On the basis of their significant difference in cost, patients with mild to moderate LN, especially white individuals, could first be treated with AZA.

### Mycophenolate Mofetil

**Pharmacology.** MMF is a prodrug of mycophenolic acid, a potent inhibitor of inosine monophosphate dehydrogenase that is indispensable for the de novo synthesis of guanosine nucleotides. The lack of any salvage nucleotide synthesis pathway in lymphocytes renders them a selective target for MMF. Conversely, other tissues with high proliferative activity (skin, intestine, neutrophils) that have an intrinsic salvage guanosine synthesis pathway can escape the anti-proliferative effects, which explains its more favorable toxicity profile compared with CYC. MMF has excellent oral bioavailability with peak levels occurring within 1 to 2 hours after administration and half-life of 17 hours. Although therapeutic drug monitoring to guide MMF dosing has been proposed in renal transplantation, validation of therapeutic MPA monitoring in LN is still required. Antacids and cholestyramine decrease the bioavailability

### Table 81-2 Indications for Immunosuppressive Therapy in Systemic Lupus Erythematosus

<table>
<thead>
<tr>
<th>General Indications</th>
<th>Specific Organ Involvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Involvement of major organs or extensive involvement of nonmajor organs (skin) refractory to other agents, or both</td>
<td>Renal</td>
</tr>
<tr>
<td>Failure to respond to or inability to taper corticosteroids to acceptable doses for long-term use</td>
<td>Proliferative or membranous nephritis (nephritic or nephritic syndrome), or both</td>
</tr>
<tr>
<td></td>
<td>Hematologic</td>
</tr>
<tr>
<td></td>
<td>Severe thrombocytopenia (platelets &lt;20,000/mm$^3$)</td>
</tr>
<tr>
<td></td>
<td>Thrombotic thrombocytopenic purpura–like syndrome</td>
</tr>
<tr>
<td></td>
<td>Severe hemolytic or aplastic anemia, or immune neutropenia not responding to glucocorticoids</td>
</tr>
<tr>
<td></td>
<td>Pulmonary</td>
</tr>
<tr>
<td></td>
<td>Lupus pneumonitis or alveolar hemorrhage, or both</td>
</tr>
<tr>
<td></td>
<td>Cardiac</td>
</tr>
<tr>
<td></td>
<td>Myocarditis with depressed left ventricular function, pericarditis with impending tamponade</td>
</tr>
<tr>
<td></td>
<td>Gastrointestinal</td>
</tr>
<tr>
<td></td>
<td>Abdominal vasculitis</td>
</tr>
<tr>
<td></td>
<td>Nervous System</td>
</tr>
<tr>
<td></td>
<td>Transverse myelitis, cerebritis, optic neuritis, psychosis refractory to corticosteroids, mononeuritis multiplex, severe peripheral neuropathy</td>
</tr>
</tbody>
</table>

### Table 81-3 Recommended Immunosuppressive Therapy for Major Organ Involvement in Systemic Lupus Erythematosus

<table>
<thead>
<tr>
<th>Disease Severity</th>
<th>Induction Therapy</th>
<th>Maintenance Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild</td>
<td>High-dose GC (0.5-1 mg/kg/day prednisone ×4-6 wk, tapered to 0.125 mg/kg every other day within 3 mo) alone or in combination with AZA (1-2 mg/kg/day)</td>
<td>Low-dose GC (prednisone ≤0.125 mg/kg on alternative days) alone or with AZA (1-2 mg/kg/day)</td>
</tr>
<tr>
<td></td>
<td>If no remission within 3 mo, treat as moderately severe</td>
<td>Consider further gradual tapering at the end of each year of remission</td>
</tr>
<tr>
<td>Moderate</td>
<td>MMF (2 g/day) (or AZA) with GC as above; if no remission after the first 6-12 mo, treat as severe</td>
<td>MMF tapered to 1.5 g/day for 6-12 mo and then to 1 g/day; consider further tapering at the end of each year in remission</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alternative: AZA (1-2 mg/kg/day)</td>
</tr>
<tr>
<td>Severe</td>
<td>Pulse IV-CYC alone or in combination with pulse IV-MP for the first 6 mo (background GC 0.5 mg/kg/day for 4 wk, then taper)</td>
<td>Quarterly pulses of IV-CYC for at least 1 year beyond remission</td>
</tr>
<tr>
<td></td>
<td>If no response, consider adding RTX or switch to MMF</td>
<td>Alternative: AZA (1-2 mg/kg/day), MMF (1-2 g/day)</td>
</tr>
</tbody>
</table>

AZA, azathioprine; CYC, cyclophosphamide; GC, glucocorticoid; IV, intravenous; MMF, mycophenolate mofetil; MP, methylprednisolone; RTX, rituximab.
of MMF, and the coadministration with AZA should be avoided. MMF may be teratogenic and should not be administered during pregnancy.

**Use in Lupus Nephritis**

**Induction Therapy.** Initial RCTs indicated equal or even superior efficacy of MMF over CYC in inducing remission in PLN, but their findings were limited by flaws in the design such as the low number of patients, the underrepresentation of severe forms of LN, and the short follow-up. The Aspreva Lupus Management Study (ALMS), one of the largest and most racially diverse RCT in LN that included a total of 370 patients, 27% of whom had estimated GFR less than 60 mL/min/1.73 m², failed to demonstrate superiority of MMF (2 to 3 g/day) over monthly pulses of IV-CYC (0.5 to 1 g/m²). Both groups received oral prednisone, with a defined taper from a maximum starting dose of 60 mg/day. At 6 months, response rates were similar for both groups (56% for MMF, 53% for IV-CYC) and there were no differences in adverse events. Subsequently, three meta-analyses of RCTs have concluded that MMF is as effective as CYC (pooled relative risk [RR] for complete remission ranging 1.49 to 1.61) and has a better safety profile (pooled RR, 0.15 to 0.17 for amenorrhea; 0.41 to 0.78 for leukopenia; 0.77 to 0.83 for infections) as induction therapy for PLN. MMF may therefore be considered as induction therapy in moderately severe PLN, especially when gonadal toxicity is an issue.

MMF has demonstrated antiproteinuric effects in LMN, but there is lack of large RCTs to formally test its efficacy. A pooled analysis of 84 patients with pure LMN who participated in the ALMS trial and in the study by Ginzler and colleagues demonstrated comparable remission rates (percentage change of proteinuria and SCr were the primary end points) in MMF- and IV-CYC-treated patients. More data are necessary to delineate the role of MMF in the management of LN.

**Maintenance Therapy.** Contreras and colleagues compared MMF versus AZA or quarterly pulses of IV-CYC as maintenance therapy in PLN following remission with 7 IV-CYC pulses. In this trial of 59 patients, 95% were black or Hispanic, 78% had diffuse PLN, with average SCr 1.6 mg/dL and serum albumin 2.7 mg/dL. After a follow-up of 72 months, MMF and AZA were superior to IV-CYC in terms of relapse-free survival (78% for MMF, 58% for AZA, 4% for IV-CYC), mortality, infections, and amenorrhea. This study was criticized for the insufficient number of patients to demonstrate superiority, the use of lower doses of IV-CYC, and the use of higher doses of GC, which may have decreased efficacy and increased the risk for infections. Two large multicenter RCTs have compared MMF versus AZA as maintenance regimen in PLN. The MAINTAIN Nephritis Trial included 105 European patients with moderately severe class III to IV PLN (10% had baseline SCr >1.4 mg/dL, 39% had proteinuria ≥3 g/day) who received induction therapy with 3 daily 750-mg IV-MP pulses followed by oral GC and 6 fortnightly IV-CYC pulses (500 mg per pulse).
On the basis of randomization performed at baseline, AZA (target dose: 2 mg/kg/day) or MMF (target dose: 2 g/day) was given at week 12 to all patients. Over a 3-year period, the two groups did not differ in terms of time-to-renal flare, number of severe flares, renal remission, or doubling of SCr. Adverse events did not differ between the groups except for blood cytopenias, which were more frequent in the AZA group. These results are different from those reported for the ALMS maintenance part. This study included 227 patients (44% nonwhites) and showed a failure rate of 32% in the AZA group versus 16% in the MMF group at 3 years after successful induction therapy during the first part of the study with either MMF or IV-CYC. Differences in the study design and the induction protocol, the number and ethnicity of the included patients, and the outcome measures may account for the discrepant results. Of note, an RCT in ANCA-positive vasculitis found that MMF was less effective than AZA for maintaining disease remission following induction therapy with IV-CYC and GC. To this end, both MMF and AZA may be used for maintenance of remission in PLN on the basis of availability and potential for pregnancy.

Use in Extrarenal Lupus. A systematic review of open-label trials concluded that MMF may be effective for refractory skin and blood manifestations in SLE. A posthoc analysis of the ALMS trial data in LN patients showed that MMF was equally efficacious with IV-CYC pulse therapy (both in combination with tapered prednisone) on general disease activity. At week 24, BILAG-defined remission was achieved in the general (100% in MMF vs. 94% in IV-CYC), mucocutaneous (84% vs. 93%), musculoskeletal (91% vs. 96%), and hematologic (60% vs. 67%) domains. Normalization of C3/C4 and anti-dsDNA titers also occurred at a similar rate. Conversely, MMF showed no efficacy in preventing extrarenal flares in 75 patients who were followed at a single center for 5 years for renal (71%) or nonrenal disease. A substantial number of patients experienced flares during the second and third years of treatment, particularly in hematologic, mucocutaneous, and musculoskeletal domains. While awaiting additional data to define the efficacy of MMF in extrarenal lupus, the drug may be used in patients with moderately severe lupus who are intolerant or have not adequately responded to AZA.

Cyclosporin A

Pharmacology. Cyclosporin A (CsA), a fungus-derived calcineurin inhibitor that deactivates T cells, also reduces antigen presentation and autoantibody production by lupus B cells. Following oral administration, peak serum concentration occurs within 1 to 8 hours. Drug concentration is measured in whole blood, but this is rarely necessary in autoimmune diseases, unless CsA is used in doses greater than or equal to 3 mg/kg/day. Clinical response occurs 1 to 2 months after treatment initiation. Several drugs interact with CsA, leading to reduced (rifampin, phenytoin, phenobarbital, nafcillin) or increased (erythromycin, clarithromycin, azoles, calcium channel blockers, amiodarone, allopurinol, colchicine) drug concentrations. The drugs may also augment CsA nephrotoxic effects (NSAIDs, aminoglycosides, quinolones, angiotensin-converting enzyme [ACE] inhibitors, amphotericin B). Common adverse events include mild gastrointestinal complaints, hirsutism, gingival hyperplasia, and mild elevation in serum alkaline phosphatase levels. Tremor, paresthesias, electrolyte disturbances (hyperkalemia and hypomagnesemia), and hypertension may also occur. Hypertension occurs in nearly 20% of patients receiving CsA and is controlled by either reduction of the dose or antihypertensive treatment. A major adverse effect is nephrotoxicity, which is reversible after adjustment of the dose or drug discontinuation, and CsA should be avoided in patients with impaired renal function (see Table 81-1).

Use in Proliferative Lupus Nephritis. Uncontrolled studies have demonstrated efficacy of CsA when used in conjunction with GC or in between quarterly doses of IV-CYC in refractory-to-conventional treatment PLN. Beneficial effects include reduction in proteinuria, stabilization of renal function, improvement in overall disease activity, and a modest steroid-sparing effect. Moroni and colleagues prospectively studied 31 LN patients (n = 24 with class III/IV nephritis) who were treated with CsA (5 mg/kg/day in two equal doses and then adjusted to trough level 80 to 120 mg/mL) and low-dose prednisone. After a mean of 7 months, all but two patients achieved complete response, defined as proteinuria less than 1 g/day and improved or stabilized renal function. About half of these patients, however, experienced a renal flare after CsA withdrawal.

Moroni and colleagues compared AZA with CsA as maintenance therapies in 69 patients with diffuse PLN and preserved renal function. All patients received induction therapy with 3 daily pulses of 1 g IV-MP, followed by prednisone and oral CYC for 3 months. They were then assigned to receive either CsA (4 mg/kg/day for 1 month and then tapered to 2.5 to 3 mg/kg/day) or AZA (1.5 to 2 mg/kg/day) for 2 to 4 years. The two groups did not differ in flare-ups, proteinuria, and blood pressure levels. Both agents were well tolerated. In another study, 40 patients with newly diagnosed PLN and mild renal insufficiency were randomly assigned to sequential induction and maintenance therapy with either CYC or CsA. The CYC regimen included 8 pulses IV-CYC (10 mg/kg) administered within 9 months, followed by 4 to 5 oral CYC boluses; CsA was given orally 4 to 5 mg/kg/day for 9 months and then tapered to 3.75 to 1.25 mg/kg/day within the next 9 months. Both groups received oral MP (0.8 mg/kg/day, then tapered). In the intention-to-treat analysis, 16 patients in the CYC group (76%) and 13 patients in the CsA group (68%) achieved complete or partial response at 9 months (induction phase). At the end of maintenance phase (18 months), the respective percentages were 52% for CYC and 95% for CsA. The trend for more favorable response in the CsA group was due to a higher proportion of patients achieving a 50% or greater decrease in proteinuria (38% in CYC group vs. 74% in CsA group). Despite its methodology flaws, this trial provides evidence for efficacy of CsA in mild to moderate proteinuric PLN with preserved renal function.

Use in Membranous Lupus Nephropathy. Balow and Austin performed an RCT in 42 patients with MLN to compare prednisone alone or in combination with CsA or monthly pulses of IV-CYC. Remission rates at 1 year were 27% for prednisone, 60% for IV-CYC, and 83% for CsA. During follow-up, however, rates of nephrotic syndrome...
relapse were 10-fold higher in the CsA than the IV-CYC group, suggesting that despite its effectiveness as induction therapy, CsA may require maintenance therapy to prevent relapses in MLN.

**Use in Extrarenal Lupus.** Uncontrolled studies of short duration have shown improvement in disease activity, anti-dsDNA titers, and cytopenias with modest GC reduction in SLE patients who received low-dose CsA.\(^6\,\,6\) The BILAG group performed a multicenter, nonblinded RCT to compare the steroid-sparing effect of CsA (titrated to 2.5 to 3.5 mg/kg/day in two divided doses and adjusted to changes in SCr) versus AZA (2 to 2.5 mg/kg/day) in severe lupus requiring 15 or greater mg/day prednisone.\(^6\) Eighty-nine patients (66% whites) were randomized; 66% had active disease (defined as a BILAG A or B in any systems), and 34% entered the study because a different steroid-sparing agent was required. At 12 months, the unadjusted mean reduction in prednisolone dose was 9.5 ± 8.1 mg in the CsA group and 10.2 ± 6.2 mg in the AZA group. These improvements, however, were deemed as suboptimal for both drugs because almost 50% of patients with active disease at baseline failed to respond. The two groups did not differ in any of the secondary end points, namely disease activity, response to treatment, flares, damage accrual, and quality-of-life measures. In terms of safety, 23 patients (49%) who received CsA developed hypertension and another 6 patients (13%) showed a rise in SCr, both successfully managed by CsA dose reduction or addition of antihypertensive treatment. Thus CsA may be used as an alternative steroid-sparing agent to AZA in SLE patients, but with monitoring of the renal function and blood pressure.

**Tacrolimus**

Tacrolimus is a 10 to 100 times more potent calcineurin inhibitor than CsA. Systemic administration has been associated with dose-dependent reversible nephrotoxicity and blood pressure elevation, albeit less often than CsA. Other reported adverse effects include cardiomyopathy in children, anxiety, seizures, delirium and tremor, diabetes, and hyperlipidemia. In a pilot study of 10 SLE patients with skin and musculoskeletal disease, tacrolimus administered at doses of 1 to 3 mg/day for 1 year resulted in a significant reduction in SLE Disease Activity Index (SLEDAI) score and the dose of GC.\(^6\) Tacrolimus has also demonstrated beneficial effects in refractory-to-conventional therapy PLN and LMN.\(^4\)\,\,6\,\,6\) Miyasaka and colleagues\(^6\) conducted a placebo-controlled, double-blind trial in 63 patients with active mild-to-moderate nephritis requiring 10 or greater mg/day prednisone. They were randomized to receive a 28-week course of either tacrolimus (3 mg/day) or placebo in combination with GC (≤10 mg/day). In intention-to-treat analysis, only tacrolimus-treated patients had significant improvement in the author-defined nephritis disease activity index; 4 out of 27 patients in the tacrolimus group as compared with 1 out of 33 in the placebo group achieved proteinuria less than 0.3 g/day. Higher rates of GI toxicity and hyperglycemia were observed in the tacrolimus group.

Tacrolimus has been used in combination with MMF in severe or resistant LN. Cortés-Hernandez and colleagues\(^6\) prospectively studied 70 patients with moderately severe PLN who received induction therapy with 3 daily pulses of IV-MP (1 g/dose) and oral MMF (2 g/day), in conjunction with oral prednisone (1 mg/kg/day for 4 weeks, then tapered). In case of renal flare or treatment failure despite an increase in MMF dose, oral tacrolimus (0.075 mg/kg/day, adjusted to trough level 5 to 10 ng/mL) was added. At the end of the 65-month follow-up, tacrolimus had been started in 17 patients (24%), with 12 of them achieving complete or partial response after an average of 24 months. Tacrolimus/MMF combination therapy (tacrolimus 4 mg/day, MMF 2 g/day) has also been compared against IV-CYC (6 to 9 pulses of 1 g/m²) in patients with mixed class IV + V LN.\(^6\) In this study, patients had mean estimated GFR of 98 mL/min, proteinuria 4.4 g/day, and most had previously been treated with MMF or CYC. Both groups received 3 daily pulses of IV-MP (0.5 g/day) and then switched to oral prednisone. After 6 months, 10 patients in the “multitarget” group versus 1 patient in the IV-CYC group achieved complete remission. Combination therapy was well tolerated, and no major effects were observed. This is in contrast, however, with the significant toxicity reported in other patients with LN who received the same combination therapy.\(^7\,\,6\) Therefore additional studies with a larger number of patients and longer follow-up are necessary to establish the efficacy, safety, and specific indications of such a multidrug approach in LN.

**Biologic Therapies**

### B Cell–Depleting Therapies

**Rituximab.** Rituximab (RTX) is a chimeric mouse-human monoclonal antibody targeting the cell membrane protein CD20, which is expressed in all developmental stages of B cells, except for the hematopoietic stem cell and the plasma cell. Its mechanism of action involves cytotoxicity through complement activation, antibody-dependent cell-mediated cytotoxicity, and induction of apoptosis. RTX has been used in more than 450 SLE patients with refractory-to-cytotoxic therapy SLE, mainly as add-on therapy to GCs or other immunosuppressive agents, or both.\(^1\,\,7\) Clinical response was noted in more than 80%, with manifestations such as neuropsychiatric disease,\(^1\,\,5\) PLN, and autoimmune cytopenias, showing higher response rates. Data from a French lupus registry including a total of 136 patients also indicated response rates of at least 70% for various manifestations, with autoimmune cytopenias exhibiting the most favorable outcomes (85% for hemolytic anemia, 92% for thrombocytopenia). A clinically significant decrease in SLEDAI by 3 or more units was observed in 71%.\(^7\) With regard to LN, a meta-analysis of observational studies found an overall complete and partial response rate with RTX therapy of 69%, with lower rates observed in LMN.\(^7\) Two European cohorts, however, have reported comparable response rates in patients with PLN and LMN who received RTX in combination with steroids and IV-CYC.\(^7\)
Unexpectedly, two RCTs failed to demonstrate efficacy of RTX in SLE. The EXPLORER trial assessed the effects of RTX over 52 weeks in patients with active moderate to severe extrarenal SLE. A total of 257 patients were randomized to receive RTX or placebo (two biweekly infusions of 1000 mg at baseline and at 6 months), in combination with background immunosuppression (AZA, MMF, MTX) and prednisone at a dose of 0.5 to 1 mg/kg/day, not tapered until day 16 after the first RTX infusion. At week 52, no difference was observed between RTX and placebo in the primary (BILAG-defined major or partial clinical response) or any of the secondary end points. The LUNAR trial evaluated the efficacy of RTX versus placebo, both in combination with MMF (3 g/day), in 144 patients with active PLN. By week 52, there was no significant difference between the two groups in terms of complete or partial renal response. Both studies were criticized for suboptimal design, mainly due to background therapies with high doses of oral GC and immunosuppressive drugs, which could have masked RTX effects. Alternatively, RTX may be more efficacious in patients with aggressive disease refractory to conventional treatment, as were most patients included in open-label trials.

RTX is generally well tolerated with mild infusion reactions being the most common adverse event, usually prevented by premedication with antihistamines and GCs. Mild infections are common (up to 20%), but the overall risk for serious or opportunistic infections is not significantly increased. Following the report of a few cases of progressive multifocal leukoencephalopathy (PML) in RTX-treated patients with RA and SLE, the U.S. Food and Drug Administration (FDA) has issued an alert about possible RTX and PML links. More data are necessary to evaluate whether RTX indeed increases the inherent risk for PML in SLE. Together, although RTX cannot be considered first-line therapy for mild to moderate SLE, its use may be justified in severe refractory cases on the basis of mounting evidence from open-label trials. Accordingly, the American College of Rheumatology (ACR) and the EULAR–European Renal Association have included rituximab (anti-CD20 mAb) as a therapeutic option for selected cases of LN refractory to conventional immunosuppressive treatment.

Ocrelizumab. The BELONG trial evaluated the efficacy and safety of ocrelizumab (humanized anti-CD20 monoclonal antibody) versus placebo, both in combination with either MMF (up to 3 g/day) or IV-CYC (Euro-Lupus protocol) in 378 patients with PLN. The study was prematurely terminated due to increased rates of serious and opportunistic infections in ocrelizumab-treated patients, both when combined with MMF and IV-CYC.

B Cell Inhibitors

Epratuzumab. Epratuzumab is a recombinant anti-CD22 monoclonal antibody that modulates lupus B cell function. CD22 is a membrane co-receptor for the B cell receptor mediating inhibitory signaling through attenuation of calcium efflux. In 14 SLE patients with moderately active disease, epratuzumab therapy resulted in reduction in BILAG scores by greater than or equal to 50% in more than 70% of patients at 6 weeks and in nearly 40% of patients at 18 weeks. Drug tolerability was acceptable, and mild infections occurred in a minority of participants. Positive results have been announced for a phase IIb placebo-controlled, dose-defining trial in moderate to severe SLE. Using a composite BILAG-based response index, patients who received epratuzumab 600 mg weekly or 1200 mg biweekly achieved almost twofold higher response rates than the placebo group at the end of the 12-week treatment cycle. A similar rate of adverse events was recorded in the two groups. Phase III trials in moderate and severe SLE are under way.

Belimumab. The B lymphocyte stimulator protein (BlyS, also known as B cell activation factor [BAFF]) and the proliferation-inducing ligand (APRIL) are growth factors important for B cell survival and maturation. BlyS binds to three different B cell receptors, namely the transmembrane activator and calcium-modulating cyclophillin ligand interactor (TACI), the B cell maturation antigen (BCMA), and the BAFF receptor (BAFF-R), whereas APRIL signals through TACI and BCMA. Belimumab is a fully human anti-BlyS monoclonal antibody. An initial phase II trial of belimumab in combination with GCs and/or immunosuppressive agents in 449 active SLE patients did not meet its primary end points. In a posthoc analysis, however, serologically active patients (ANA titer ≥:80 or anti-dsDNA positive) had a significantly higher response by week 52 in terms of SLEDAI and physician’s global assessment. These two parameters were combined with BILAG to define a novel activity index, the SLE responder index (SRI), which revealed favorable response rates for belimumab. Two subsequent phase III placebo-controlled RCTs (BLISS-52, BLISS-76) evaluated two different doses of belimumab (1 mg/kg, 10 mg/kg; dosed IV on days 0, 14, 28, and then every 28 days) on top of standard therapy in more than 1500 serologically positive SLE patients, using the SRI as the primary end point. Belimumab treatment resulted in a modest albeit significant reduction in disease activity and needs for additional GC, coupled with an increase in time-to-first flare. In BLISS-52, 58% achieved the primary end point in the high-dose belimumab arm versus 44% in the placebo group; the respective figures in BLISS-76 were 43% for belimumab and 34% for placebo. There were no significant differences in infectious adverse events between the two drugs. These data opened the way for approval of this agent for the treatment of SLE.

Atacicept. Atacicept (TACI-Ig) is a recombinant fusion protein of the extracellular domain of TACI and the human IgG1.Fc domain. TACI mediates signals from both BlyS and APRIL, thus affecting memory B cells, plasma cells, and immunoglobulin production. This was illustrated in two phase Ib placebo-controlled trials in patients with mild to moderate SLE. Both studies found no difference in adverse events between the two study groups, with the exception of mild injection-site reactions in patients receiving atacicept. However, a phase II/III trial of atacicept in combination with MMF in LN was prematurely terminated due to increased infection rates.

Co-stimulation Blockade

CD40-Ligand Blockade. CD40-ligand (CD40L) is expressed on activated T cells and stimulates antigen-
presenting cells including B cells through engagement with CD40. Anti-CD40L treatment resulted in improvement of renal disease and increased survival in NZB/W F1 lupus mice. Two fully humanized monoclonal anti-CD40L antibodies have been developed for therapeutic trials in humans: BG9588 and IDEC-1310. The latter showed no efficacy in mild to moderate SLE. The encouraging results by the use of BG9588 (improvement in serologic activity and decrease in hematuria in 28 patients with PLN) were overpowered by the occurrence of serious thromboembolic events.

**Abatacept.** Abatacept (CTLA4-Ig) is a recombinant protein that comprises the extracellular domain of human CTLA-4 fused to the Fc portion of human IgG1 and antagonizes CD28-mediated T cells. It is approved for the treatment of moderate to severe RA and juvenile idiopathic arthritis. CTLA4-Ig delayed disease progression in lupus-prone mice, especially when combined with CYC. A placebo-controlled, phase IIb trial of abatacept in combination with GC and background immunosuppressive therapy in 175 SLE patients with skin, joint, cardiovascular, and respiratory manifestations failed to meet its primary end point. After 1 year, the proportion of patients with new BILAG A/B flare after steroid taper did not differ between the two groups. However, the trial design may have under mined potential efficacy of abatacept because patients were started on high doses of GC (30 mg/day) and tapering began on day 29 or 57. Posthoc analysis showed that patients with polyarthritis benefited more from abatacept treatment. Serious adverse events were significantly more frequent in the abatacept group (20% vs. 7% in placebo) including bronchitis, diverticulitis, and gastroenteritis. Two RCTs are currently under way to evaluate abatacept in combination with MMF or IV-CYC in LN.

**Anticytokine Therapy**

**Tumor Necrosis Factor Inhibitors.** Tumor necrosis factor (TNF) has divergent effects on the immune systemic in lupus. In an open-label study, seven SLE patients with moderately active disease received four doses of 300 mg infliximab on day 0 and at weeks 2, 6, and 10, in combination with AZA or MTX. Anti-dsDNA and other autoantibodies increased in most patients, peaking 4 to 10 weeks after the last infliximab infusion but returning to baseline levels thereafter. In the prospective follow-up of 13 patients, 6 out of 9 patients with LN had a long-term (up to 5 years) response after four infusions of infliximab in combination with AZA. All five patients with severe arthritis responded, but only for 2 months after the last infusion. Long-term therapy was associated with high rates of serious adverse events. In another small open-label study in resistant proliferative or membranous LN, treatment with infliximab in combination with oral prednisone and CsA resulted in transient reduction of proteinuria and stabilization of renal function. Together, it is unlikely that TNF inhibition will be used routinely in SLE treatment.

**Interferon Inhibition.** Type 1 interferon (IFN-α) has been implicated in lupus pathogenesis through breakdown of immune tolerance. Serum IFN levels are increased and correlate with disease activity, and gene expression studies have identified an “interferon signature” in a subset of SLE patients. A phase I trial evaluated the effects of a single dose of anti-IFN-α monoclonal antibody (MEDI-545) in SLE. It was noted to downregulate IFN-inducible genes and other signaling pathways such as GM-CSF, TNF, IL-10, IL-1β, and BAFF. Two phase II trials are ongoing to assess safety and efficacy of anti-IFN therapy in lupus.

**Anti-IL-6 Therapy.** In lupus-prone mice, blockade of IL-6 or its receptor reduced anti-dsDNA levels, ameliorated proteinuria, and increased survival. Tocilizumab, a humanized monoclonal antibody against the α-chain of the IL-6 receptor, has shown efficacy in RA. Illei and colleagues tested tocilizumab in 16 patients with moderately active SLE, in combination with low-dose GC. Patients received tocilizumab (2, 4, 8 mg/kg) biweekly for 12 weeks and were monitored for an additional 8 weeks. Results revealed a correlation between clinical and serologic efficacy, evidenced by reduction in SLAM and SLEDAI indices, acute-phase reactants, and anti-dsDNA levels. Dose-dependent neutropenia (median decrease in neutrophil count by 56% in the 8 mg/kg group) and frequent infections (upper respiratory and urinary tract) were observed. To resolve issues on efficacy and safety, a larger trial is necessary.

**Anti-IL-10 Therapy.** IL-10 is upregulated in SLE and correlates with disease activity, although its exact pathogenic role has not been elucidated. An open-label trial in six steroid-dependent SLE patients showed improvement in disease activity reported up to 6 months after the administration of an anti-IL-10 murine monoclonal antibody (BN10) for 21 days. However, all patients developed antibodies against BN10, and new trials are awaited with a human anti-IL-10 monoclonal antibody.

**Other Therapies**

**Intravenous Immunoglobulin**

Intravenous immunoglobulin (IVIG) exerts immunosuppressive effects by interaction with anti-idiotypic antibodies, interference with the complement and cytokines, cytolysis of target cells, induction of apoptosis through Fc receptors, and modulation of co-stimulatory molecules. In an RCT of 14 patients, IVIG (400 mg/kg for 5 consecutive days monthly for 18 months) was as effective as pulse IV-CYC (1 g/m² every 2 months for 6 months, then every 3 months for 1 year) as maintenance therapy of PLN. In another study, high-dose IVIG (2 g/kg divided over 5 days) in 1 to 8 monthly courses was administered in 20 SLE patients with cytopenias, massive proteinuria, arthritis, fever, arthralgia, mood changes, and psychosis. IVIG therapy resulted in improvement in SLEDAI and dose of GC. Common adverse events include fever, myalgia, headache, and arthralgia; less common are aseptic meningitis, nephropathy in patients who receive sucrose-containing preparations, and thromboembolic complications in older patients with atherosclerotic risk factors. The drug is contraindicated in IgA deficiency.

**Synthetic Tolerogens**

Tolerogenic peptides aim at restoring immune tolerance in lupus. Abetimus sodium (LJP-394) contains four identical dsDNA strands covalently linked to a small molecule
platform. It is thought to reduce anti-dsDNA antibodies by binding and clearance of soluble anti-dsDNA antibodies and B cell receptor cross-linking on dsDNA-specific B cells. Initial studies showed decrease of anti-dsDNA levels, prolongation of the time-to-renal flare, and reduced renal flares, especially in patients with high-affinity anti-dsDNA antibodies. However, a phase III placebo-controlled trial in 317 LN patients failed to meet its primary end point (prolongation of time to renal flare), although abetimus-treated patients had 21% fewer flares, reduced proteinuria, and improved SLEDAI scores. The drug was well tolerated, and a 900-mg dosage was introduced in another trial, with no additional benefit. The spliceosomal peptide P140 is another tolerogen that has demonstrated benefit in early SLE trials. In a dose-escalation study in 20 patients with moderate SLE, the drug significantly decreased anti-dsDNA levels and slightly reduced SLEDAI. The results of an ongoing phase IIIb trial using SRI as efficacy index are awaited.

**Dehydroepiandrosterone**

Dehydroepiandrosterone (DHEA) is a naturally occurring inactive steroid in adrenal glands, testes, and ovaries. Seven RCTs including a total of 842 patients have been performed to assess efficacy and safety of DHEA in SLE. Their meta-analysis showed that the drug had modest clinical effect only in mild to moderate disease. DHEA treatment also resulted in improvement in health-related quality of life measurements.

**Clofazimine**

Clofazimine (CFZ) (100 mg/day), an antimicrobial used in the treatment of leprosy, has shown efficacy in the treatment of skin involvement in lupus. However, it should be reserved for patients with exclusively cutaneous disease because it may provoke severe lupus flares.

**Management of Specific Systemic Lupus Erythematosus Manifestations and Treatment Algorithms**

**Key Points**

- Skin manifestations in SLE usually respond to sun exposure prophylaxis, topical glucocorticoids, and systemic antimalarials.
- Risk stratification based on renal pathology, demographic, and clinical and laboratory characteristics enables the identification of LN patients at high risk for renal dysfunction who may benefit from aggressive cytotoxic therapy.
- In moderately severe PLN, mycophenolate mofetil may be preferred as induction regimen, especially when gonadal toxicity is a concern. Failure to achieve response after the initial 6 months of therapy should evoke decisions about intensifying or altering immunosuppressive therapy.

**Combination of monthly pulses of intravenous CYC and pulse intravenous methylprednisolone is the treatment of choice for severe LN. If substantial improvement occurs after the first 6 months, maintenance therapy with azathioprine or mycophenolate mofetil may be started.**

**Glucocorticoids alone or in combination with immunosuppressive agents are recommended for neuropsychiatric events felt to reflect an immune/inflammatory process; antiplatelet and/or anticoagulation therapy is recommended for events related to antiphospholipid antibodies.**

**Antiplatelet or anticoagulation therapies, or both, are necessary in patients with antiphospholipid syndrome to prevent recurrent events, but the intensity of such therapies remains controversial.**

**Mucocutaneous and Joint Disease**

No single therapeutic agent has been officially approved for cutaneous lupus erythematosus (CLE). Mild malar rash and other photosensitive rashes usually respond to prophylaxis from sun exposure, but the use of sunscreens with high sun protection factor cannot be overemphasized. Topical GCs reduce redness and scaling. In an RCT of 78 patients with discoid lupus erythematosus (DLE), high-potency topical fluocinonide 0.05% was more effective than low-potency hydrocortisone 1% (response rates 27% vs. 10%, respectively). Calcineurin inhibitors (tacrolimus, pimecrolimus) are a useful alternative, especially for the face, because they cause less atrophic and rosacea-like effects than topical steroids.

In refractory to topical therapy skin disease, systemic antimalarials may be used alone or in combination with oral GCs (up to 20 mg/day). HCQ was equally effective and had more favorable safety profile than the oral retinoid acitretin in an RCT of 58 DLE patients. Other systemic treatment remains largely empiric. MTX has shown efficacy in refractory subacute CLE and DLE. Alternative choices include retinoids, dapsone, MMF, and IV-CYC. Thalidomide and IVIG should be reserved for patients with severe recalcitrant CLE due to potential serious neurotoxicity and high cost, respectively (Figure 81-1).

Lupus arthritis follows the pattern of nonerosive symmetric polyarthritis primarily affecting the small joints of the hands and feet. In mild arthritis, initial therapy should be based on antimalarials. In persistent or aggressive disease, utilization of DMARDs is advocated. In a prospective controlled study, SLEDAI scores, need for steroids, and articular involvement were significantly improved in SLE patients receiving MTX for 6 months. Leflunomide is used less often. Anti-TNF agents have been implicated for anti-dsDNA antibody development or even drug-induced lupus. RTX may be used in severe lupus arthritis refractory to DMARDs.

**Lupus Nephritis**

**Induction Therapy**

Current therapeutic strategies in LN include an initial induction phase aimed at substantially improving disease
activity (or even attaining remission), followed by a maintenance phase, in which the goal is to maximize the therapeutic effect and consolidate the response. Risk stratification, according to renal pathology and patient demographic, clinical and laboratory characteristics, enables the identification of patients at risk for renal dysfunction or ESRD, or both, who may benefit from aggressive cytotoxic therapy (Table 81-5).


For mild class I/II disease, a limited trial of prednisone (0.5 to 1 mg/kg/day for 4 to 6 weeks and then tapered to alternate day 0.125 to 0.25 mg/kg if remission occurs) is indicated. The use of three consecutive daily pulses of IV-MP may expedite remission and allow for the use of lower GC doses (0.5 mg/kg/day). Addition of immunosuppressive therapy (AZA 1 to 2 mg/kg/day) either from the beginning or during GC tapering may decrease cumulative steroid dose. If the patient does not achieve complete remission (clearing of cellular casts and proteinuria, normalization of complement, minimal lupus activity) within 3 months or if nephritis worsens, therapy with MMF or monthly pulse IV-CYC should be initiated. Delay in immunosuppressive therapy because of a partial response beyond the 3 to 4 months may have an adverse impact on response to therapy, thus increasing the risk for flare.

### Table 81-5 Severity of Lupus Nephritis

<table>
<thead>
<tr>
<th>Classification</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proliferative Nephritis</strong></td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>Class III nephritis without severe histologic features (crescents, fibrinoid necrosis); low chronicity index (≤3); normal renal function; non-nephrotic range proteinuria</td>
</tr>
<tr>
<td>Moderately severe</td>
<td>Mild disease as defined above with partial or no response after the initial induction therapy, or delayed remission (&gt;12 mo), or focal proliferative nephritis with adverse histologic features or reproducible SCr increase ≥30%, or Class IV nephritis without adverse histologic features</td>
</tr>
<tr>
<td>Severe</td>
<td>Moderately severe as defined above but not remitting after 6-12 mo of therapy, or proliferative disease with impaired renal function and fibrinoid necrosis or crescents in &gt;25% of glomeruli, or mixed membranous and proliferative nephritis, or proliferative nephritis with high chronicity alone (chronicity index &gt;4) or in combination with high activity (chronicity index &gt;3 and activity index &gt;10), or rapidly progressive glomerulonephritis (doubling of SCr within 2-3 mo)</td>
</tr>
</tbody>
</table>

| Membranous Nephropathy              |                                                                             |
| Mild                                | Non-nephrotic range proteinuria with normal renal function                  |
| Moderate                            | Nephrotic range proteinuria with normal renal function at presentation      |
| Severe                              | Nephrotic range proteinuria with impaired renal function at presentation (≥30% increase in SCr) |

*Concomitant therapy with corticosteroids or immunosuppressive drugs, or both, may modify urinary sediment and histologic findings and should be taken into consideration. SCr, serum creatinine.

In moderately severe PLN, meta-analyses of RCTs comparing MMF with CYC confirmed the efficacy of the former for induction therapy; claims of superiority toward CYC were not substantiated. MMF is increasingly used as a first-line treatment for most PLN cases due to its favorable toxicity profile, whereas CYC is reserved for the most severe ones. For whites, the low-dose (Euro-Lupus) IV-CYC regimen may be equally efficacious and less toxic than the high-dose (NIH) regimen. Both earlier and recent studies underscore the significance of early response to therapy at 6 months (defined as a decrease in SCr level and proteinuria <1 g/day) as a strong predictor of long-term renal outcome. Thus failure to achieve response after the initial 6 months of therapy should precipitate decisions about switching from MMF to pulse IV-CYC.

In severe LN, the authors recommend induction therapy with 7 monthly pulses of IV-CYC in combination with 3 initial pulses of IV-MP at the start, followed by pulses of IV-MP in combination with IV-CYC at monthly intervals (1 pulse/month) for the first 6 to 12 months (see Table 81-3). High doses of prednisone (0.5 to 1 mg/kg/day tapered after 4 weeks) should be continued during the induction period. Addition of plasma exchange is of no benefit in terms of survival. For patients strongly opposing pulses of IV-CYC, alternative induction regimens include (1) daily oral MMF (1 g twice a day for 12 months with increase to 3 g/day if no response in 6 to 8 weeks) or (2) monthly pulses of IV-MP (1 g/m² daily) for three doses and then at monthly
Figure 81-2  Recommended treatment of proliferative lupus nephritis. Renal response is assessed at 6 months. Complete response (CR) is defined as decrease in proteinuria to less than 1 g/day (or <0.3 g/day if nephritis was diagnosed in the past 6 months) with normal serum albumin concentrations; inactive urine sediment; and improved or stable renal function. Partial response (PR) is defined as significant change in proteinuria (if nephrotic at baseline ≥50% decrease in proteinuria to < 3 g/day; if non-nephrotic at baseline but not meeting the CR criteria) and improved or stable renal function. All induction regimens include pulse intravenous methylprednisolone (IV-MP) (1 g/pulse × 3) followed by oral prednisone (0.5 to 0.6 mg/kg for the first 4 weeks of induction, then tapered). AZA, azathioprine; IV-CYC, intravenous cyclophosphamide; MMF, mycophenolate mofetil; NR, no response.

**Maintenance Therapy**

Following induction of remission, patients with mild to moderate disease may be treated with low-dose GC (7.5 to 15 mg prednisone on alternate day) or AZA, with CsA being an alternative agent. For maintenance therapy in moderately severe disease, MMF or AZA could be used on the basis of availability and potential for pregnancy because MMF is associated with an increased risk of spontaneous abortion and fetal malformation. Because of the significant difference in the cost between the two drugs, patients with less severe LN could first be treated with AZA, especially white individuals.

In severe disease, our preferred approach is quarterly pulse IV-CYC until 1 year beyond remission. If substantial improvement occurs, this could be followed by AZA or MMF after the first 6 months. Microscopic hematuria or non-nephrotic proteinuria may not clear for several months, even when most other clinical parameters have remitted. Remission may occur at an average of 1.5 to 2 years after therapy. Thus discontinuing pulse IV-CYC therapy because of lack of achieving remission before completing at least 2 years of treatment is not justified unless there is definite disease worsening (reproducible increase in SCr and/or >50% increase in proteinuria at the nephrotic range). In selected patients with severe disease who have achieved remission, MMF may be used as a maintenance regimen (see Figure 81-2).

**Lupus Membranous Nephropathy**

LN with mesangial expansion (pure membranous) and low-grade proteinuria (<2 g/day) carries a low risk for progression to ESRD (20% at 10 years) and may require renin-angiotensin axis blockade and low-dose GC. Persistent nephrotic-range proteinuria, abnormal renal function, and black race have been implicated as high-risk features and are indications for immunosuppressive therapy. Uncontrolled studies indicate beneficial effects of the combination of GC with AZA or calcineurin inhibitors in mild to moderate LN. In the RCT of Austin and colleagues, oral CsA and pulse IV-CYC were equally successful in inducing remission. Nephrotic syndrome relapses were more common in the CsA group, suggesting that these patients should receive maintenance therapy with lower CsA doses or other immunosuppressive agent. MMF has emerged as an efficacious regimen with more favorable safety profile for both induction and maintenance therapy of class V LN on the basis of the pooled data from two RCTs comparing MMF versus IV-CYC in moderately severe LN. However, this will have to be proven in larger trials with longer follow-up. Mixed membranous and proliferative histology (especially when diffuse proliferation is present) has a worse prognosis, even from pure proliferative disease. These patients should be aggressively treated as those with PLN. Bao and colleagues have reported interesting results using the combination of prednisone, MMF, and tacrolimus as induction therapy for mixed class IV+V LN. In this trial, multitarget therapy was associated with higher rates of complete response compared with pulse IV-CYC. These results need to be confirmed before adopting this regimen.

**Treatment of Renal Flares**

Approximately 30% to 50% of patients with moderate to severe LN will relapse after achieving partial or complete remission. Nephritic flares are characterized by active urine sediment and reproducible increase of SCr (≥30% increase) and may adversely affect renal prognosis; proteinuric flares without significant changes in renal function have a more benign prognosis. Risk factors for progression to ESRD after a nephritic flare are patients with marked loss of renal function (SCr >2 mg/dL) at the time of response, partial response to therapy, and high chronicity and activity.
indices at renal biopsy. Flares are more common in African-American patients and patients with undetectable serum C3/C4 levels. However, pre-emptive treatment in the face of abnormal serology (rising anti-dsDNA or reduced C3/C4 titters, or both) may result in overtreating a large number of patients. Mild to moderate flares (stable SCr, subnephrotic proteinuria) may be treated with GC in combination with AZA or MMF. Calcineurin inhibitors (alone or added to existing immunosuppressive therapy) have also demonstrated efficacy. For severe nephritic flares, reinstatement of cytotoxic therapy with monthly pulses of IV-CYC and IV-MP is the authors’ preferred approach, with MMF considered as an alternative agent. RTX has also been successfully used in a small case series.

Central Nervous System Disease

Less than 40% of neuropsychiatric events in SLE patients can be attributed to lupus, whereas the remaining cases represent complications of the disease or its therapy, or they may be caused by infections, metabolic abnormalities, and drug adverse effects. Neuropsychiatric SLE (NPSLE) is a clinical challenge, and difficulties include the correct attribution of neuropsychiatric syndromes to SLE, the selection of proper diagnostic examinations, and optimal treatment. To facilitate the management of these patients, EULAR has published recommendations using an evidence-based approach followed by expert consensus (Table 81-6). All nonlupus contributing factors should be identified and treated, and symptomatic therapy should be considered if appropriate. Moderate to high doses of GCs alone or in combination with immunosuppressive therapy (AZA for mild to moderate cases, IV-CYC for severe ones) may be used for neuropsychiatric events felt to reflect an immune/inflammatory process (particularly acute confusional state, aseptic meningitis, myelitis, optic neuritis, refractory seizure disorder, peripheral neuropathies, psychosis) or when they occur in the context of active generalized lupus, following exclusion of non–SLE-related causes. Although a placebo-controlled trial reported improved cognition in five out of eight SLE patients with inactive disease and mild cognitive dysfunction who were treated with low-dose prednisone (0.5 mg/kg/day for 21 days, then tapered), GCs should not be routinely administered in these patients unless to control concurrent SLE or other overt NPSLE activity. In severe NPSLE refractory to cytotoxic therapy, the use of plasma exchange, IVIG, and RTX has been reported in uncontrolled studies, with varying rates of success. Antiplaatelet or anticoagulation therapy are recommended for NPSLE related to aPL antibodies, especially for thrombotic cerebrovascular disease (see Antiphospholipid Syndrome later).

Hematologic Disease

Peripheral cytopenias are common but usually mild in SLE. A thorough clinical and laboratory evaluation is necessary to exclude offending drugs or other secondary causes. Mild

Table 81-6  Approach to Systemic Lupus Erythematosus (SLE) Patients with Neuropsychiatric Manifestations

<table>
<thead>
<tr>
<th>What Are the Risk Factors for NPSLE?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized (non-CNS) lupus activity or damage, previous or other concurrent major NPSLE manifestation(s), persistently positive moderate-to-high titers of aPL antibodies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>When to Suspect NPSLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any SLE patient at risk who presents with new-onset neurologic or psychiatric manifestations without an apparent cause</td>
</tr>
<tr>
<td>In patients with subtle or mild signs or symptoms, a high index of suspicion is required to exclude underlying overt NPSLE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Is It NPSLE?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild manifestations (headache, mood disorders, anxiety, mild cognitive dysfunction, polyneuropathy without electrophysiologic confirmation) are common (up to 40%) but are not usually related to lupus</td>
</tr>
<tr>
<td>Non–SLE-related causes (infections, metabolic disturbances, drug adverse effects) must be excluded</td>
</tr>
<tr>
<td>Most (40%-50%) lupus-related events occur at onset or during the first 2-4 yr after SLE diagnosis, common (50%-60%) in the presence of generalized lupus activity</td>
</tr>
<tr>
<td>Attribution to lupus more likely when NPSLE risk factors are present</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What Diagnostic Workup Is Indicated?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnetic resonance imaging is the preferred neuroimaging test and may help to identify:</td>
</tr>
<tr>
<td>Ischemic/thrombotic, demyelinating, or infectious processes</td>
</tr>
<tr>
<td>T2-weighted white matter lesions: in the absence of other confounding factors (increased age, long-standing SLE, atherosclerotic risk factors, heart valve disease) may reflect underlying CNS lupus activity (especially when ≥5 in number, ≥6-8 mm in size, bihemispheric)</td>
</tr>
<tr>
<td>Cerebrospinal fluid analysis should be performed when CNS infection is suspected; mild abnormalities common in active NPSLE</td>
</tr>
<tr>
<td>Other tests as indicated: electroencephalogram to diagnose seizure disorder, neuropsychologic tests to assess cognitive dysfunction, nerve conduction studies for peripheral neuropathy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What Is the Treatment of NPSLE?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control aggravating factors (infection, dehydration, metabolic abnormalities, hypertension)</td>
</tr>
<tr>
<td>Control symptoms (anticonvulsants, antidepressants, antipsychotics)</td>
</tr>
<tr>
<td>Glucocorticoids and/or immunosuppressive therapy in cases of:</td>
</tr>
<tr>
<td>Acute confusional state, aseptic meningitis, myelitis, optic neuritis, refractory seizure disorder, peripheral neuropathies, severe psychosis</td>
</tr>
<tr>
<td>Control generalized (non-CNS) lupus activity</td>
</tr>
<tr>
<td>Antithrombotic or antiplatelet therapy</td>
</tr>
<tr>
<td>aPL-associated NPSLE (particularly cerebrovascular disease, ischemic optic neuropathy, chorea) or when antiphospholipid syndrome–associated thrombotic events are present</td>
</tr>
</tbody>
</table>

CNS, central nervous system; NPSLE, neuropsychiatric systemic lupus erythematosus.
cytopenias require no specific therapy other than regular monitoring. In more severe cases (platelet count < 50 × 10^3/mm^3 or active bleeding, neutrophil count < 1000/mm^3), GCs (1 mg/kg/day with gradual tapering) are the mainstay of treatment. Pulsed IV-MP followed by lower doses of prednisone (0.6 mg/kg/day) may be used alternatively. Steroid-sparing agents (AZA, CsA) can be added during steroid tapering. Patients with steroid-resistant thrombocytopenia may be candidates for splenectomy, especially when there is no significant lupus activity in other organs. Early reports from the 1980s have shown a satisfactory response to splenectomy with progressive rise in platelet counts. Prophylactic measures for avoidance of infectious complications in patients undergoing splenectomy are mandatory. These include immunization against encapsulated bacteria (Streptococcus pneumoniae, Neisseria meningitidis, Haemophilus influenzae) and influenza virus, as well as prophylactic antibiotic coverage when necessary.

Resistant life-threatening cytopenias may require potent immunosuppressive therapy. Monthly pulses of IV-CYC have been shown to reverse severe refractory autoimmune thrombocytopenia. In severe neutropenia, however, the risk for potential leukocyte toxicity of CYC should be considered. The use of lower-dose IV-CYC (Euro-Lupus protocol) is associated with a better safety profile and is advocated by some centers. RTX may be considered in patients with refractory neutropenias, especially when the use of IV-CYC is hampered by its potential myelotoxicity. Of note, RTX may also be associated with both early- and late-onset neutropenia, which is usually self-limited.

Treatment of immune thrombocytopenia may sail into a new era with the development of novel thrombopoietin mimetic agents that enhance platelet production. Two such agents, romiplostim and eltrombopag, were shown to be superior than standard of care in large RCTs and have been approved for the treatment of immune thrombocytopenia. They resulted in a greater incidence of sustained platelet response, less bleeding, fewer transfusions, reduced requirement for other treatments (including splenectomy), and greater improvement in quality of life. Adverse events were minimal. The ultimate place these remedies will take in the armamentarium against immune thrombocytopenia (before or after splenectomy) will depend on their long-term efficacy/safety profile and their cost compared with the potentially curative and relatively safe choice of splenectomy.

Supportive treatment may be necessary for severe cytopenias or associated complications. Febrile neutropenia should be treated with broad-spectrum antibiotics and, if neutrophil counts are less than 500/mm^3, with human G-CSF. Serious hemolytic anemia (Hb < 7 g/dL) may require red blood cell transfusions, whereas platelet transfusions are best avoided unless invasive procedures are planned.

**Antiphospholipid Syndrome**

aPL antibodies (anticardiolipin [aCL], anti-β2-glycoprotein I [anti-β2-GPI], and lupus anticoagulant [LAC]) are encountered in 30% to 40% of SLE and are associated with increased risk for thrombo-occlusive incidents. The combination of vascular thrombosis or obstetric morbidity, or both, and persistently positive aPL antibodies measured at least 12 weeks apart, defines antiphospholipid syndrome (APS).

**Thrombotic Antiphospholipid Syndrome**

APS-associated thrombosis requires antiplatelet and/or anticoagulation therapy to prevent recurrent events, but the intensity of such therapies remains controversial. On the basis of the results of two systematic reviews, patients with definite APS and first venous or arterial noncerebral/noncoronary thrombosis should receive oral anticoagulation (warfarin) at a target international normalized ratio (INR) 2 to 3, although some experts recommend higher-intensity anticoagulation following arterial thrombotic events. Data from the Antiphospholipid Antibodies and Stroke Study (APASS) and large RCTs in the general population suggest that antithrombotic therapy is not superior to antiplatelet therapy for secondary thromboprophylaxis after noncardioembolic stroke or transient ischemic attack. Because many patients studied were elderly and had low titers of aPL antibodies, determined at a single time point, the conclusions may be limited to these populations. Acute coronary artery syndromes should be treated according to the evidence base for the general population. For patients with recurrent thrombotic events, anticoagulation should target INR 3 to 4, especially if they have high-risk aPL profile (LAC or aCL IgG at higher titers, or anti-β2-GPI plus LAC or aCL). Additional atherothrombotic risk factors should be aggressively controlled. The role of newer classes of anticoagulants (direct thrombin inhibitors, oral direct factor Xa inhibitors)—currently licensed for the management of venous thromboembolism in the general population—for secondary APS thromboprophylaxis remains to be determined.

HCQ, in addition to its anti-inflammatory effects, has been proposed to exert antithrombotic properties through inhibiting platelet aggregation and arachidonic acid release from stimulated platelets. Use of HCQ has been associated with reduced rates of thrombosis in both aPL-positive and aPL-negative patients. A systematic review of epidemiologic studies found evidence for antithrombotic effect of HCQ in SLE patients, especially in studies accounting for exposure previous to the event. These results have not been confirmed in large prospective cohort studies and controlled studies are necessary to determine the effectiveness of HCQ for primary thromboprophylaxis. RCTs have also shown a protective effect of rosuvastatin (20 mg/day) against thrombosis (including cardiovascular events and venous thromboembolism) in healthy adults with normal low-density lipoprotein levels and elevated C-reactive protein (>2 mg/dL). These findings justify the conduct of clinical trials of statins in aPL-positive patients. In small case series, RTX has demonstrated beneficial effects in severe APS cases including resolution of symptoms (especially thrombocytopenia) and reduction in aPL antibody titers.

**Pregnancy in Antiphospholipid Syndrome**

Pregnant SLE-APS patients are at increased risk for complications including maternal thrombosis, recurrent spontaneous abortions before 10 weeks’ gestation, and late...
adverse pregnancy outcomes such as fetal death, pre-eclampsia, fetal growth restriction, and preterm birth.\textsuperscript{15,151} For women with APS and a history of pregnancy complications or thrombosis, or both, a meta-analysis of three RCTs concluded that the combination of unfractionated heparin and aspirin confers a significant benefit in live births (odds ratio [OR] for first trimester loss, 0.26; number needed to treat [NNT], 4).\textsuperscript{152} The pooled effect of low-molecular-weight heparin was also favorable (OR 0.70) but not statistically significant. Conversely, combination therapy of either unfractionated or low-molecular-weight heparin with aspirin had no effect in prevention of late-pregnancy losses. Results from RCTs do not specifically define optimum treatment for women with fetal death (>10 weeks’ gestation) or previous early delivery (<34 weeks’ gestation) because of severe pre-eclampsia or placental insufficiency. Nonetheless, most experts recommend low-dose aspirin and either prophylactic or intermediate-dose heparin. A beneficial effect of low-dose aspirin in primary prevention of thrombotic events and miscarriage in SLE with persistently positive moderate to high titers of aPL antibodies has been suggested by some,\textsuperscript{148,151} but not all,\textsuperscript{154} studies. Neither aspirin combined with low-molecular-weight heparin nor aspirin alone improved the live birth rate, as compared with placebo, among women with unexplained recurrent miscarriage.\textsuperscript{155}

Vitamin K antagonists are teratogenic and should be avoided between 6 and 12 weeks of gestation; even after 12 weeks of gestation they should be used cautiously due to increased risk for fatal bleeding. Antithrombotic coverage of the postpartum period is recommended for all APS patients irrespective of their thrombotic history. Women with previous thrombosis will need long-term anticoagulation, and treatment is generally switched to warfarin as soon as the patient is clinically stable after delivery.\textsuperscript{139,156} In patients with no previous thrombosis, the recommendation is prophylactic-dose heparin or low-molecular-weight heparin therapy for 4 to 6 weeks after delivery, although warfarin is an option. Both heparin and warfarin are safe during breastfeeding.

Other Antiphospholipid Syndrome Manifestations

A distinct type of small-vessel vaso-occlusive nephropathy with histologic features of thrombotic microangiopathy and chronic vascular lesions, termed APS nephropathy, should be suspected in aPL-positive patients with hypertension, proteinuria (usually mild to moderate), hematuria, and renal impairment.\textsuperscript{151,152} These patients may benefit from aggressive immunosuppressive therapy along with anticoagulation, but development of ESRD is common.\textsuperscript{159} A minority of patients (<1%) can suffer from a potentially life-threatening variant of APS, catastrophic APS (CAPS), which is characterized by multiple small-vessel thromboses resulting in multiorgan failure.\textsuperscript{160} Approximately 70% of patients manifest renal involvement, usually resulting in severe hypertension, proteinuria, hematuria, and renal impairment. The condition is associated with 50% mortality rate and may be treated with combination of anticoagulation and immunosuppressive therapy (high-dose GC alone or in combination with CYC) plus attempts to reduce aPL titers (plasma exchange, IVIG).

### ADDITIONAL ISSUES

#### Key Points

- Pregnancy outcome in SLE is optimal when disease is clinically quiescent for at least 6 months.
- The risks of treatment during pregnancy must be weighed against the risks of an untreated SLE flare, with potential deleterious effects on the mother and the fetus.
- The treatment of pediatric-onset SLE is similar to that in adult SLE patients.

#### Treatment of Refractory Systemic Lupus Erythematosus

IV-CYC in combination with IV-MP has long been considered the treatment of choice for most patients with severe, life-threatening lupus. Mounting evidence from uncontrolled studies supports the beneficial effects of RTX (alone or in combination with IV-CYC) in refractory SLE.\textsuperscript{161-163} Higher response rates have been reported for skin, blood, central nervous system (CNS), and renal manifestations.\textsuperscript{74} Advantages from the use of RTX include its favorable toxicity profile and the rapid onset of action.\textsuperscript{73} MMF may rescue a few refractory SLE patients including LN cases with inadequate response to CYC. However, its efficacy in critically ill patients requires further documentation. Calcineurin inhibitors have also shown efficacy in resistant-to-standard immunosuppressive therapy lupus,\textsuperscript{164} particularly LN,\textsuperscript{165} and benefits of combination therapy with MMF were reported in a small RCT.\textsuperscript{69} For selected patients with CNS involvement, autoimmune thrombocytopenia, or APS, IVIG may be considered as an adjunct therapy. Synchronized therapy with plasmapheresis and pulse IV-CYC has been used in severe PLN cases.\textsuperscript{166}

High-dose chemotherapy with autologous hematopoietic stem cell transplantation (HSCT) has been used in SLE.\textsuperscript{167} The rationale is to maximally suppress the immune system with an immunoablative regimen (usually high-dose CYC combined with equine antithymocyte globulin and pulse IV-MP) and then rescue the patient from prolonged cytopenias by the infusion of mobilized CD34-enriched stem cells. A few patients who received HSCT experienced improvement in disease activity, but relapses were common.\textsuperscript{168} In a pilot study, 15 patients refractory to conventional treatment (n = 14 with nephritis) received a small dose of allogeneic bone marrow–derived mesenchymal stem cells (1 × 10^6/kg by IV injection).\textsuperscript{169} Prednisolone was tapered from week 2 onwards, and IV-CYC was continued with larger intervals. Rapid and sustained improvements in autoantibody levels, proteinuria, and nonrenal manifestations were reported, with no significant acute toxicity. Until more data are available, stem cell transplantation remains an experimental procedure that needs to be considered in critically ill lupus patients, ideally in experienced centers.

#### Treatment of Lupus in Pregnancy

Pregnancy may increase disease activity and precipitate the appearance of flares (13% to 74%), which are usually mild.
Lupus pregnancies are considered high risk for maternal and fetal complications. A recent meta-analysis reviewing 2571 pregnancies in SLE patients reported unsuccessful pregnancy and preterm birth rates of 23% and 39%, respectively, while maternal complications included nephritis (16%), hypertension (16%), and pre-eclampsia (7%). A history of LN and the presence of aPL antibodies were associated with development of hypertension and premature birth.

The management of a pregnant woman with an SLE flare is challenging and should be dealt with on a multidisciplinary basis (Table 81-7). The risks of treatment with potential deleterious effects on the mother and the fetus must be balanced against the risks of untreated disease. Mild flares involving the skin, joints, and blood are usually treated with GC. The latter are relatively safe, although they carry a small risk of harm (increased risk of cleft palate in children, intrauterine growth retardation, maternal hypertension, diabetes) and thus should be used in the lowest enough doses to control disease activity. HCQ has a good safety record despite crossing the placenta. AZA, although listed as FDA category D, can be used with caution as a steroid-sparing agent. The same holds true for CsA (category C). In contrast, CYC is a major teratogen and should not be used unless there is no available alternative for organ-threatening disease in the mother, preferentially in the late second or third trimester. Fetal loss due to CYC toxicity may occur. MMF and MTX are listed as category D and X, respectively, and should be avoided. For biologic agents, data regarding safety during pregnancy are lacking and their use is generally discouraged. Pregnant SLE patients with a severe renal flare (active urinary sediment, increase in SCr) can be treated with high-dose GC and antihypertensives, excluding ACE inhibitors, whereas AZA can be used cautiously to allow steroid tapering. Prompt delivery of the fetus at the earliest safe time point possible is warranted.

Neonatal lupus occurs in some babies born by mothers with anti-SSA/Ro or anti-SSB/La antibodies, or both. The most serious complication is neonatal complete heart block (CHB), which occurs in up to 2% of such pregnancies and carries a 20% to 30% mortality risk. The risk of recurrence in a mother having already borne a child with CHB ranges from 14% to 19%. Therefore all women with SLE should be evaluated for the presence of anti-Ro/La antibodies. The most vulnerable period for the development of conduction disturbances is between week 18 and 26 of gestation, during which anti-Ro/La-positive patients should be evaluated with weekly fetal Doppler echocardiography for the prompt identification of conduction anomalies, mainly PR interval prolongation. If CHB develops, no modality has been shown to reverse it. In contrast, a large observational study in 118 pregnant women with anti-Ro antibodies showed that first-degree block may be reversed by fluorinated GCs such as dexamethasone (4 mg/day) initiated at the time of diagnosis and continued throughout pregnancy. Two trials found no beneficial effect of IVIG on prevention of recurrence of congenital heart block in Ro/SSA-positive mothers.

Regarding breastfeeding, the American Academy of Pediatrics (AAP) states that nursing is permissible for women receiving GC but the interval between dose and nursing should be at least 4 hours if the prednisone dosage is greater than 20 mg/day. Antimalarials may be continued during lactation, whereas AZA is not recommended due to possible risk for immunosuppression, carcinogenicity, and growth restriction of the child. MTX, MMF, CsA, leflunomide, and CYC are also contraindicated.

The high-risk features of SLE pregnancy and the possible risk for adverse outcomes highlight the importance of family counseling and planning before pregnancy is sought. It is generally accepted that pregnancy outcome is optimal when disease is clinically quiescent for at least 6 months and, in cases of kidney involvement, if renal function is normal or near normal. Until this desired goal is reached, common effective means of contraception should be taken to avoid unplanned pregnancies (see Women’s Health Issues).

### Table 81-7 Approach to the Management of Pregnancy in Systemic Lupus Erythematosus

<table>
<thead>
<tr>
<th>Planning of pregnancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensure that lupus is inactive for at least 6 mo</td>
</tr>
<tr>
<td>Reassure patient: small risk for major flare</td>
</tr>
<tr>
<td>Discourage pregnancy if Scr &gt;2 mg/dL</td>
</tr>
<tr>
<td>Determine aPL antibodies and other antibodies that may be of relevance (anti-SSA, anti-SSB)</td>
</tr>
<tr>
<td>Obtain baseline serology and chemistry labs (Scr, SAlb, uric acid, anti-dsDNA, C3/C4)</td>
</tr>
<tr>
<td>Be aware of the small risk for CHB, especially in women with both anti-SSA and anti-SSB antibodies or with a prior episode of CHB.</td>
</tr>
<tr>
<td>In such cases, may monitor for CHB between 16 and 24 wk of gestation</td>
</tr>
<tr>
<td>Monitor closely blood pressure and proteinuria. Should this develop, differentiate between active nephritis and pre-eclampsia</td>
</tr>
<tr>
<td>Presence of generalized lupus activity, active urine sediment, and low serum complement are in favor of lupus nephritis</td>
</tr>
<tr>
<td>For patients with antiphospholipid syndrome, consider combined heparin and aspirin to reduce risk for pregnancy loss and thrombosis. Patients with aPL antibodies may be treated with aspirin, although there are no adequate data to support its use</td>
</tr>
</tbody>
</table>

CHB, congenital heart block.

### Treatment of Pediatric Systemic Lupus Erythematosus

Pediatric-onset SLE (pSLE) represents 15% to 20% of all SLE cases and is associated with higher disease severity and more rapid damage accrual than adult-onset SLE. The use of antimalarials (HCQ 4 to 6 mg/kg/day) is helpful in children with skin disease and arthritis. In mild nephritis (class II or mild class III LN), normal renal function, normal blood pressure, and non-nephrotic proteinuria, oral GCs (51 mg/kg/day) may suffice to control disease. Initial pulses of IV-MP (30 mg/kg/day up to a maximum dose of 1 g/day for 3 consecutive days) may allow for lower oral prednisone dose (0.5 mg/kg/day). In patients with class III nephritis and high activity index or class IV disease, pulse therapy with IV-MP and IV-CYC (0.75 to 1 g/m²) should be considered. Lehman and colleagues studied clinical and histologic progression of class III/IV nephritis in 16 children who completed at least 3 years of continuous treatment with IV-CYC. Significant improvements in disease activity, renal function, and renal activity index were observed. In a retrospective study of 28 patients with class IV LN who received IV-CYC and oral GC, repeat biopsy showed...
histologic improvement in 20 out of 25 children. At last follow-up, 3.5 ± 2.3 years after second biopsy, 26 patients retained normal renal function. In PLN with preserved renal function and non-nephrotic proteinuria, MMF could be used. A single study has reported good outcomes with the combination of MMF and CsA in severe class III/IV LN patients who responded inadequately to initial induction therapy with MMF and GC.

Both AZA and MMF are efficacious maintenance regimens in pediatric PLN. Baskin and colleagues studied 20 pSLE patients with severe PLN who received 6-month induction therapy with pulse IV-MP (30 mg/kg, followed by oral prednisone 0.5 to 1.0 mg/kg/day) and IV-CYC (500 mg/m²). AZA (2 mg/kg/day) was commenced as a remission-maintaining regimen in all patients; in 10 patients, treatment subsequently switched to oral MMF (1200 mg/m²/day) because of intolerance or lack of efficacy. Following this approach, 14 patients (70%) achieved complete remission and another 3 (15%) had partial remission. An alternate-day regimen of 10 to 15 mg oral prednisone should be aimed at during the maintenance phase.

Renal flares remain an important issue that has not been adequately addressed. Alternative to a second course of IV-CYC and GC, which may be associated with significant long-term toxicity, is the combination of IV MTX and CYC administered for 9 months. Although data for RTX are limited, it may be used (600 mg/m², maximum dose of 1 g/m² on days 1 and 15) in refractory renal and extrarenal disease. RTX can be used as monotherapy or in combination with CYC, under the following dosage scheme: 500 to 750 mg/m² on days 2 and 16.

Significant CNS involvement usually warrants combination of high-dose steroids with an immunosuppressive agent such as CYC or AZA. To date, there is no study documenting an increased rate of serious adverse effects such as life-threatening infections or secondary neoplasia in children with SLE receiving cytotoxic therapy compared with adults. All children with LN should be regularly reviewed to treat accompanying hypertension, proteinuria, and renal dysfunction. Monitoring for normal growth, physical and pubertal development, and bone density is important, especially for patients on long-term treatment with GC.

**KEY POINTS**

The diagnostic approach and management of an SLE patient with possible infection should consider: (1) the dominant clinical syndrome, (2) the history of epidemiologic exposures, and (3) the “net-state” of immunosuppression.

Immunizations are safe and effective in SLE patients but must be avoided in active disease. Inactivated live vaccines are contraindicated in patients taking immunosuppressive drugs or high-dose glucocorticoids.

In SLE patients with end-stage renal disease, hemodialysis may be preferred over chronic ambulatory peritoneal dialysis. Renal transplantation is a potential alternative.

### COMORBIDITIES IN SYSTEMIC LUPUS ERYTHEMATOSUS

#### Infections and Immunizations

#### Risk Factors and General Management

Infections account for 20% to 55% of all deaths in SLE patients. Susceptibility to infections may be due to underlying immune dysregulation and therapeutic factors, particularly high-dose GCs and immunosuppressive drugs. A broad spectrum of infections have been reported in SLE including bacterial, mycobacterial, viral, fungal and parasitic infections, with the respiratory, urinary tract, and CNS as the most commonly involved sites. Risk factors for infections include increased clinical or serologic lupus activity, or both, at baseline; major organ involvement; especially renal and lung involvement; lymphopenia; persistent neutropenia (<1000/mm³); hypoaalbuminemia; high dose of GC (each increase of 10 mg/day prednisone is associated with 11-fold increased risk for serious infection); and prior (within the last 6 months) use of immunosuppressive drugs (especially AZA and CYC).

The evaluation of a lupus patient who receives immunosuppressive therapy and presents with symptoms or signs suggestive of infection possesses diagnostic and therapeutic challenges (Figure 81-3). This is complicated by the fact that active SLE per se is associated with increased risk for infection, and on the other hand, viral infections can mimic a lupus flare. Findings that favor the diagnosis of infection include the presence of shaking chills, leukocytosis and/or neutrophilia (especially in the absence of steroid therapy), increased numbers of band forms or metamyelocytes on peripheral blood smear, and concomitant immunosuppressive therapy. The diagnosis of SLE fever is favored by the presence of leukopenia (not explained by cytotoxic therapy), normal or only slightly increased C-reactive protein level, low C3/C4, and elevated anti-DNA antibodies. If fever fails to resolve in a patient receiving prednisone 20 to 40 mg daily, it is likely that that fever is due to infection, not SLE. Elevated serum procalcitonin (PCT) levels have been reported to be predictive of bacterial or mycotic infections, although their diagnostic utility in patients with systemic autoimmune diseases has been questioned. Pending microbiology results, adequate antimicrobial therapy (including broad-spectrum antibiotics in suspected nosocomial infection) is recommended to reduce adverse outcomes.

#### Specific and Opportunistic Infections

**Tuberculosis Infection.** Tuberculosis (TB) infection rates are 6- to 60-fold higher in SLE patients than the general population and may be caused by both Mycobacterium tuberculosis (MTB) and nontuberculous mycobacterium (NTM).

Extrapulmonary involvement is
The advent of the T cell IFN-γ release assays, a more specific test than the tuberculin skin test, may improve the detection of latent TB infection.

*Pneumocystis jiroveci* Infection. The incidence of *Pneumocystis* pneumonia in SLE patients on CYC approximates 0.15%. Risk factors include lymphopenia (≤750/mm³), during lupus treatment, high disease activity, renal involvement, interstitial pulmonary fibrosis, and high prednisone dosage. Some authors recommend *Pneumocystis* pneumonia prophylaxis (one double-strength tablet of trimethoprim-sulfamethoxazole three times a week or 202

**Figure 81-3** Initial assessment and management of systemic lupus erythematous patients who receive immunosuppressive therapy and present with fever or other symptoms and signs suggestive of infection. *a,* Systemic inflammatory response syndrome (SIRS): T ≥38°C or <36°C, tachycardia (heart rate >90/min), tachypnea (respiratory rate >20/min), white blood cells (WBC) >12,000/mm³; comorbidities: age older than 65 years, diabetes, chronic cardiopulmonary disease. †, Consider empirical therapy for *Pneumocystis* pneumonia in severe hypoxemia or diffuse pulmonary infiltrates. ‡, Consider tuberculosis and other opportunistic central nervous system (CNS) infections. AZA, azathioprine; BAL, bronchoalveolar lavage; CBC, complete blood count; CSf, cerebrospinal fluid; G-CSF, granulocyte colony-stimulating factor; GM, galactomannane; H1N1, influenza H1N1; HRCT, high-resolution chest tomography; IV-CYC, intravenous cyclophosphamide; MMF, mycophenolate mofetil; MRSA, methicillin-resistant *Staphylococcus aureus*; MTX, methotrexate; PCR, polymerase chain reaction; T, temperature; VATS, video-assisted thoracoscopy. (Modified from Papadimitraki ED, Bertsias K, Chamilos MD, Boumpas DT: Systemic lupus erythematosus: cytotoxic drugs. In Tsokos G, Buyon JP, Koike T, Lahita RG, editors: Systemic lupus erythematosus, ed 5, St Louis, 2010, Elsevier, pp 1083–1108.)
dapsone 100 mg/day if allergic to sulfamethoxazole) for patients on high-dose GC (≥20 mg/day prednisone or equivalent) alone or in combination with cytotoxic drugs, especially if CD4 count is less than 300 cells/mm³. Of note, risk of allergic reactions to sulfamethoxazole may be increased in SLE patients.

Viral Infections. The most commonly reported viral infections in SLE patients are parvovirus B19 and cytomegalovirus (CMV) (predominantly in severely immunosuppressed patients). The latter may mimic a lupus flare or present with specific organ involvement such as hepatitis, gastrointestinal bleeding, or pulmonary infiltrates. Renal insufficiency, lymphopenia, APS manifestations, treatment with CYC, multiorgan involvement at presentation, and a lower frequency of antiviral treatment have been related to fatal viral infection in SLE cohorts. The introduction of direct antigenic (CMV pp65 antigenemia) and molecular (quantitative polymerase chain reaction [PCR]) testing has revolutionized the diagnosis of CMV infection in the setting of immune suppression. Viral load has proven useful for diagnosis of infection, prediction of disease, and monitoring response to therapy. However, detection of small amounts of viral DNA in peripheral blood by PCR is not necessarily indicative of clinically significant CMV infection, and correct interpretation of test results may require consultation with an infectious diseases specialist. Further confirmatory diagnosis of organ-specific involvement can be made by histologic detection of characteristic inclusion bodies. Increased age (≥60 years), symptomatic infection, and lymphopenia (all three correlating with CMV antigenemia above 5.6 × 10⁹ PMNs) have been associated with fatal outcome, and their presence may therefore prompt initiation of antiviral therapy.

Varicella-zoster virus (VZV) reactivation is an important issue in SLE patients. Risk factors include history of CYC or AZA exposure, the presence of concurrent or previous malignancy, and LN. Shingles dissemination and bacterial superinfection are mostly linked to high-dose GC treatment. SLE patients do not display increased risk for HIV infection, and screening should be based on individualized risk factors. Due to risks of occurrence and reactivation of the infection following immunosuppressive therapy, particularly when steroids are administered, screening for hepatitis C and B virus is prudent before starting such therapies.

Immunizations

Although vaccination may hypothetically induce polyclonal cell activation in lupus precipitating a flare, it is felt to be safe. Vaccines, however, should not be given to patients with active disease. Inactivated live vaccines (measles, mumps, rubella, polio, VZV, and vaccinia [smallpox]) are contraindicated in patients taking immunosuppressive drugs or prednisone, or both, at a dose greater than 20 mg/day. Influenza vaccine is safe and effective, although seroprotection and seroconversion rates are lower than those in healthy controls; booster vaccination does not increase antibody titers in annually vaccinated patients. Pneumococcal vaccine is also safe, but the resultant antibody titers may be decreased in SLE patients; use of GC may contribute to blunted antibody responses. Protective immune response can be achieved safely in SLE patients with both tetanus toxoid and H. influenzae type B in addition to pneumococcus.

According to the Advisory Committee on Immunization Practices (ACIP) guidelines and EULAR guidelines, individuals older than 60 years of age should be vaccinated with a single dose of VZV vaccine. This is a live-attenuated vaccine and should be administered at least 14 days before initiation, or it should be deferred for at least 1 month after discontinuation of high-dose GC or immunosuppressive therapy. Therapy with low-dose MTX (<0.4 mg/kg/week) or AZA (<3 mg/kg/day) is not considered sufficiently immunosuppressive to create vaccine safety concerns. Kuruma and colleagues reported on hepatitis B vaccination in inactive SLE patients treated with low-dose GC and not receiving immunosuppressive drugs, negative for anti-dsDNA and aCL antibodies. Patients were administered a recombinant vaccine. All patients developed protective antibodies, and there were no SLE flares.

Chronic Kidney Disease and End-Stage Renal Disease

Risk Factors and Dialysis

Approximately 10% to 20% of SLE patients will develop ESRD. Clinical predictors are abnormal SCR values at presentation, delay in treatment initiation, failure to achieve remission, and systolic hypertension. Progression of LN to the point of dialysis does not necessarily indicate ESRD because 5% to 10% of patients will recover sufficient renal function to interrupt dialysis at least temporarily. Patients with rapid deterioration of renal function are more likely to have a reversible physiologic (dehydration, infection, acute tubular necrosis) or pathologic (crescentic glomerulonephritis) component accounting for their renal insufficiency. In these patients, immunosuppressive therapy (pulse of IV-MP and IV-CYC 0.4 to 0.5 g/m²) may improve renal function to interrupt dialysis at least temporarily. Patients experience a decline in lupus activity. Discontinuation of GCs and immunosuppressive drugs is essential to minimize the risk for septic complications. Cardiovascular morbidity and mortality are also increased in patients with ESRD, underscoring the need for tight control of atherosclerotic risk factors.

Renal Transplantation

Renal transplantation is a viable alternative for lupus patients with ESRD. Graft and patient survival rates are
comparable with those in other patient groups, although this is not confirmed in all studies. Results are superior with living than cadaveric donor transplantation (5-year graft survival rates 77% to 90% vs. 41% to 58%). If a living donor is available, the possibility for pre-emptive transplantation can be considered because this approach is associated with superior graft and patient outcomes in other kidney diseases. Nevertheless, a period of at least 3 months on dialysis may allow some patients to recover adequate renal function for significant time periods.

There are no prospective studies comparing various immunosuppressive regimens after renal transplantation in SLE. Calcineurin inhibitors should generally be included in the induction phase (6 to 12 months) on the basis of observations of improved graft survival rates since their introduction in lupus transplantation. Due to potential nephrotoxicity and adverse effects on cardiovascular disease (CVD) risk factors, their use should be minimized during the maintenance phase and other immunosuppressive agents (MMF, AZA) should be preferentially employed. Recurrence of LN (usually mild mesangial proliferative disease) in the renal allograft is a rare event (2% to 3%) and not an important cause of graft loss. Risk factors for recurrent LN include non-Hispanic black race, female gender, age younger than 33 years, and living donor transplantation. APS has been associated with post-transplant renal thrombosis and poor graft outcome and warrants anticoagulation therapy.

**Cardiovascular Morbidity**

SLE patients have a 2.3 to 7.5 increased risk for developing coronary heart disease compared with age-matched controls, after adjusting for traditional CVD risk factors. In lupus, traditional CVD risk factors are reinforced by disease-related risk factors such as circulating prothrombotic aPL antibodies, antibodies against high-density lipoprotein cholesterol (HDL-C) interfering with its function, a typical lupus-pattern of dyslipidemia (low HDL-C, high triglycerides, normal or slightly elevated low-density lipoprotein cholesterol (LDL-C), and abnormal serum homocysteine levels), and the proatherogenic effect of systemic inflammation.

Tight control of CVD risk factors is recommended for SLE patients, according to risk stratification and the presence of comorbidities. Hypertension requires diligent monitoring and treatment because it has been associated with worse outcomes in LN. For patients with nephritis and proteinuria greater than 1 g/day, target blood pressure should be less than 110/70 mm Hg. For most cases, ACE inhibitors or angiotensin receptor blockers (ARBs), or both, are the initial treatments of choice. For dyslipidemia, Wajed and colleagues advocated for the institution of dietary or lipid-lowering therapy toward a target LDL-C less than 101 mg/dL in all SLE patients without stratifying for presence of other risk factors. Similarly, many clinicians follow the National Cholesterol Education Program Adult Treatment Panel III guidelines for the use of statins in primary CVD prevention, considering SLE as a coronary heart disease risk factor equivalent to diabetes. However, this approach is not universally accepted because it may lead to overtreatment of LDL-C levels. Nonetheless, aggressive treatment of dyslipidemia (target LDL-C <100 mg/dL and triglycerides <150 mg/dL) is recommended for patients with multiple risk factors, especially those with moderate or severe lupus. Although CVD is the leading cause of death among women and it carries a higher mortality than men (44% vs. 27%), a large survey revealed that only 13% of the women recognized the magnitude of the problem. This, coupled with the greater challenge the diagnosis of coronary disease poses in women compared with men owing to its often atypical presentation, emphasizes the importance of sensitization of lupus women regarding their increased probability to develop CVD.

**Osteoporosis**

Ongoing disease activity, premature menopause caused by use of immunosuppressive drugs, relative vitamin D deficiency due to avoidance of sun exposure, and the use of systematic GCs all contribute to reduced bone mineral density (BMD) in SLE patients. Quality indicators regarding osteoporosis prevention and therapy suggest that (1) all SLE patients receiving prednisone greater than or equal to 7.5 mg/day for 3 or more months should undergo BMD testing unless already on antiresorptive or anabolic therapy and receive osteoporosis prophylaxis with calcium and vitamin D supplements and (2) patients receiving prednisone greater than or equal to 7.5 mg/day for 1 or more months and having a central T score of −2.5 or less or a history of fragility fracture should be treated with an antiresorptive or anabolic agent unless contraindications exist. This daily prednisone dosage is higher than the guideline proposed by the ACR for prevention of steroid-induced osteoporosis (5 mg/day) in the general patient population. The authors recommend osteoporosis prophylaxis with daily prednisone doses of greater than or equal to 5 mg for postmenopausal women and greater than or equal to 7.5 mg for premenopausal women.

**Malignancy in Lupus**

Hematologic malignancies (particularly non-Hodgkin's lymphoma [NHL]) and cervical and lung cancer occur more commonly in SLE compared with the general population, followed by rare forms of malignancy affecting the hepatobiliary tract and the vulva/vagina. Immunosuppressive therapy (with the potential for mutagenesis and the impairment of antitumor immune surveillance) and intrinsic SLE-related mechanisms (chronic antigenic stimulation, impaired surveillance) may account for this risk. NHL is associated with SLE (standardized incidence ratio [SIR], 3.6) with the most commonly identified histologic subtype being diffuse large B cell lymphoma, which usually runs an aggressive course. Hodgkin's lymphoma (HL) is also more frequent in SLE (SIR, 3.2). The risk for hematologic malignancies may increase after exposure to immunosuppressive medications, particularly after a period of 5 years following cessation of therapy. Because SLE and lymphomas share clinical manifestations (fever, lymphadenopathy, splenomegaly, cytophenias, and monoclonal expansion of B cells), a high index of suspicion is necessary for early detection of the latter. In such cases, an aggressive
investigation is warranted with appropriate imaging studies and potentially lymph node biopsy.

Cervical dysplasia is increased in women with lupus as a result of impaired clearance of human papillomavirus (HPV) due to exposure to immunosuppressive agents, particularly CYC (increase by 1 g of IV-CYC exposure corresponding to 13% increased risk of cervical intraepithelial neoplasia [CIN]). Therefore SLE should be regarded as a risk factor for cervical malignancy and high-risk HPV infection. The authors would recommend cervical cytology for cancer screening once (EULAR) or twice (U.S. Preventive Services Task Force) in the first year and then annually, adding HPV testing to the first-year-obtained cervical smears and then modifying subsequent screening on the basis of these results (cervical cytology screening every 6 months for women with detectable HPV DNA and annually for others). Although the efficacy of HPV vaccine has not been investigated in patients with autoimmune diseases, EULAR guidelines concluded that HPV vaccination should be considered for women with SLE until the age of 25 years, similar to the general population.

SLE patients have a moderately increased risk for lung cancer (SIR, 1.4), mostly adenocarcinomas, yet this risk is higher in SLE patients who smoke. While waiting for the final verdict on the use of chest computed tomography as a screening test in the smoking population, smoking and nonsmoking patients should be evaluated similarly to the general population. With regard to breast cancer, there is no evidence for increased risk in SLE and thus no particular recommendation should be applied beyond screening guidelines used in the general population.

**Emergencies in Patients with Lupus**

SLE patients may visit the emergency department (ED) for complications related to lupus itself, lupus treatment, or unrelated reasons. Critical questions confronting the clinician are (1) whether the event is related to lupus and (2) whether in the presence of lupus the management should differ. In general, lupus-related emergencies frequently occur when disease is active. For example, approximately 60% of primary NPSLE events occur in the presence of generalized lupus activity. Common symptoms bringing lupus patients to the ED are fever, shortness of breath, and chest pain. Poor compliance, low education level, severity of the underlying disease, and higher damage scores are risk factors for hospitalization.

SLE patients might also develop life-threatening organ dysfunction, related to disease activity or immunosuppressive therapy, severe enough to require admission to the intensive care unit (ICU). Infections are the leading cause of admission (45% to 61%). Hsu and colleagues studied 51 SLE patients admitted to the ICU and identified intracranial hemorrhage, gastrointestinal hemorrhage, and septic shock as poor outcome predictors. High APACHE II scores and use of vasopressors have been associated with patient mortality. Overall mortality in SLE patients admitted to the ICU is high, although it has declined during the past decade (from 47% to 24%). An aggressive search for infection together with recommended empiric antimicrobial therapy is crucial. When it is not clear whether the underlying process is related to infection or active disease, empiric antimicrobial therapy together with high-dose GC (0.5 to 1 mg/kg/day) may suffice. Once infection has been ruled out, pulse steroids or CYC, or both, may be used. Agents such as IVIG, plasmapheresis, and RTX may be considered as adjuvant therapy in life-threatening disease or in cases of deterioration or inadequate response following 1 to 2 weeks of therapy.

**WOMEN’S HEALTH ISSUES**

**KEY POINTS**

- Synthetic gonadotropin releasing hormone analogues significantly decreased rates of gonadal failure in young women with severe SLE treated with CYC.
- Contraceptive methods including hormone agents are generally safe for most SLE patients.

Premature ovarian failure is an age- and dose-dependent adverse effect of CYC therapy, whereas male gonadal toxicity may be observed with as little as 7 g cumulative CYC dose. Strategies to preserve fertility in postpubertal SLE women include hormonal contraceptives, gonadotropin releasing hormone (GnRH) antagonists, and embryo and oocyte cryopreservation. Ovarian tissue banking for future tissue transplantation has been suggested for prepubertal girls. Individual patient preferences should be considered when deciding about fertility preservation. Some authors have suggested that coadministration of GnRH antagonists confers protection against premature ovarian failure and therefore recommend a GnRH antagonist–based protocol in CYC-treated female patients. In men receiving CYC for malignancies, frequency of azoospermia ranges from 30% to 90%. Administration of testosterone and sperm banking represent valid strategies for preservation of testicular function and fertility (see Table 81-4).

Contraceptive methods are generally safe for most SLE patients. Previous research had suggested that hormonal agents might increase the risk of disease flares. However, two RCTs found no increase in flares in those without severe disease activity at study entry, and a systematic review concluded that benefits of use outweigh potential risks for most contraceptive methods. However, women with positive aPL antibodies should avoid combined hormonal methods because of an increased risk for thrombosis. Coexisting conditions such as severe thrombocytopenia, atherosclerosis, hypertension, and venous thrombosis could make certain contraceptives less advisable.

Women with SLE may experience lower urinary tract symptoms, often accompanied by voiding dysfunction (small bladder capacity, low urinary flow rate). Urinary bladder involvement is thought to contribute to recurrent urinary tract infections in some SLE patients. The approach to investigation and management of recurrent urinary tract infections is similar to that in the general population; control of disease activity may prevent bladder dysfunction. Lupus mastitis represents a rare form of lupus panniculitis occurring in 2% of women with SLE and may clinically mimic breast carcinoma. GC treatment is...
generally effective, but refractory forms may necessitate surgical excision.

**EVIDENCE AND EXPERT-BASED RECOMMENDATIONS IN SYSTEMIC LUPUS ERYTHEMATOSUS**

Due to the systemic nature of SLE, multiple medical sub-specialties are involved in the care of these patients, dictating an integrated approach to their care. To this end, EULAR has developed recommendations covering the most important aspects in the diagnosis, management, and monitoring. These recommendations—developed not only for the specialists but for all internists and primary care physicians—are based on a combined research-based evidence approach and expert opinion consensus.

**Future Directions**

This past decade has witnessed major advances in defining risk factors and phenotypes, elucidating pathogenesis, and optimizing treatment. Recognition of adjuvant-like factors that promote the production of type I interferon via Toll-like receptors offers new targets for therapy in addition to those targeting B and T cells. Exploration of the genetic and environmental factors that determine susceptibility to disease may eventually lead to the identification of individuals at risk and elucidate the primary events that cause autoimmunity. In contrast to monogenic diseases, the expansion of personalized medicine in lupus awaits a more complete description of predisposition. Genome-wide, next-generation sequencing efforts now under way will provide within the next few years a more comprehensive description of the relations between genome sequence variation and clinical phenotypes.

Meanwhile, new drugs have been added to the armamentarium against the disease and new therapeutic strategies are aimed at inducing prompt remission with more intense therapy and prevention of flares with less toxic therapies. The introduction of MMF has added a useful new drug while providing additional valuable insights in trial design in lupus. Despite past efforts devoted to arguing on the choice of specific agents, there is finally a consensus that what is more important is a strategy aiming at remission and its maintenance with the treatment that best fits the patient. With the approval of the first biologic agent in March 2011, the disease is finally showing signs of yielding to more targeted therapy and unraveling its heterogeneity and complexity. Most importantly, it has become clearer than ever that optimal long-term outcome requires not only treatment of the disease flares but also management of its comorbidities. To this end, lupus highlights the need for a multidisciplinary approach and superb internal medicine skills.

**Selected References**


Full references for this chapter can be found on [www.expertconsult.com](http://www.expertconsult.com).
References


KEY POINTS

Antiphospholipid antibodies (aPLs) are a family of autoantibodies directed against phospholipid-binding plasma proteins, most commonly β2-glycoprotein I.

The clinical manifestations range from asymptomatic to catastrophic antiphospholipid syndrome (APS).

Stroke is the most common presentation of arterial thrombosis; deep vein thrombosis is the most common venous manifestation.

Pregnancy losses typically occur after 10 weeks’ gestation (fetal loss), but earlier losses also occur.

Catastrophic APS is a rare form of APS that consists of multiple thromboses of medium and small arteries occurring over days.

Diagnosis should be made in the presence of characteristic clinical manifestations and persistently positive aPLs (measured at least 12 weeks apart).

Prevention of secondary thrombosis lacks a risk-stratified approach; the effectiveness of high-intensity anticoagulation in APS patients with vascular events is not supported by prospective controlled studies.

A common strategy to prevent fetal loss in aPL-positive patients with a history of pregnancy morbidities is low-dose aspirin and heparin.

Primary thrombosis prevention in persistently aPL-positive individuals requires a risk-stratified approach; elimination of reversible thrombosis risk factors and prophylaxis during high-risk periods are crucial. The effectiveness of aspirin in persistently aPL-positive patients without vascular events is not supported by prospective controlled studies.

Currently, no evidence indicates that anticoagulation is effective for nonthrombotic manifestations of aPLs, such as thrombocytopenia or heart valve disease.

Catastrophic APS patients usually receive a combination of anticoagulation, corticosteroids, intravenous immunoglobulin (IVIG), and plasma exchange.

Diagnosis of the antiphospholipid syndrome (APS) requires that a patient have both a clinical event (thrombosis or pregnancy morbidity) and the persistent presence of antiphospholipid antibody (aPL), documented by a solid phase serum assay (anticardiolipin or anti-β2-glycoprotein I [anti-β2GPI] immunoglobulin [IgG or IgM]), a coagulation assay (inhibitor of phospholipid-dependent clotting—the lupus anticoagulant test), or both. Preliminary (Sapporo) classification criteria for APS,1 revised in 2004,2 are listed in Table 82-1.

Certain factors are not included as criteria but may be helpful in the diagnosis of individual patients. These include IgA anticardiolipin or anti-β2GPI, valvular heart disease, thrombocytopenia, early preeclampsia, and livedo reticularis (Table 82-2). These factors are rare, nonstandardized, or nonspecific phenomena that are too unreliable for use in clinical studies, but they occur in a sufficient number of patients to support a suspected diagnosis.

APS can occur as an isolated diagnosis, or it can be associated with systemic lupus erythematosus (SLE) or another rheumatic disease. Transient aPL, but probably not the syndrome, can be induced by drugs and infection.7

EPIDEMIOLOGY

Low-titer, usually transient, anticardiolipin occurs in up to 10% of normal blood donors,4,5 and moderate- to high-titer anticardiolipin or a positive lupus anticoagulant test occurs in less than 1%. The prevalence of positive aPL tests increases with age. Ten percent to 40% of SLE patients5 and approximately 20% of rheumatoid arthritis patients8 have positive aPL tests.

Based on a limited number of uncontrolled and non–risk-stratified studies, asymptomatic (no history of vascular or pregnancy events) aPL-positive patients have a 0% to 4% annual risk of thrombosis; patients with other autoimmune diseases such as SLE are at the higher end of the range.7,8 The aPL profile (low vs. high risk for thrombosis) and patients’ clinical characteristics (presence or absence of other acquired or genetic thrombosis risk factors) influence the individual risk of thrombosis.9 Ten percent of first-stroke victims have aPLs,10 especially those who are young (up to 29%),5,11 as do up to 20% of women who have suffered three or more consecutive fetal losses.12 Fourteen percent of patients with recurrent venous thromboembolic disease have aPLs.13

CAUSE

The main antigen to which aPLs bind is not a phospholipid but rather a phospholipid-binding plasma protein, namely, β2GPI (apolipoprotein H). β2GPI is normally present in serum at a concentration of 200 mg/mL, is a member of the complement control protein family, and has five repeating domains and several alleles. An octapeptide in the fifth domain and critical cysteine bonds are necessary for both phospholipid binding and antigenicity14; a first-domain site activates platelets15,16; in vivo, β2GPI binds to phosphatidylinerine on activated or apoptotic cell membranes, including those of trophoblasts, platelets, and endothelial cells. Under physiologic conditions, β2GPI may function...
**Table 82-1** Revised Sapporo Classification Criteria for Antiphospholipid Syndrome

<table>
<thead>
<tr>
<th>Clinical Criteria</th>
<th>Laboratory Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vascular thrombosis*</td>
<td>Lupus anticoagulant present in plasma on two or more occasions at least 12 wk apart, detected according to the guidelines of the International Society on Thrombosis and Hemostasis</td>
</tr>
<tr>
<td>One or more clinical episodes¹ of arterial, venous, or small vessel thrombosis² in any tissue or organ</td>
<td>1. Lupus anticoagulant present in plasma on two or more occasions at least 12 wk apart, detected according to the guidelines of the International Society on Thrombosis and Hemostasis</td>
</tr>
<tr>
<td>Pregnancy morbidity</td>
<td>2. Anti–β₂-glycoprotein I antibody of immunoglobulin (Ig)G or IgM isotype in serum or plasma, present in medium or high titer (&gt;40 GPL or MPL, or &gt;99th percentile), on two or more occasions at least 12 wk apart, measured by a standardized ELISA</td>
</tr>
<tr>
<td>(a) One or more unexplained deaths of a morphologically normal fetus at or beyond the 10th wk of gestation, or</td>
<td>3. Anti-β₂-glycoprotein I antibody of immunoglobulin (Ig)G or IgM isotype in serum or plasma (in titer &gt;99th percentile) present on two or more occasions at least 12 wk apart, measured by a standardized ELISA</td>
</tr>
<tr>
<td>(b) One or more premature births of a morphologically normal neonate before the 34th wk of gestation because of eclampsia, severe preeclampsia, or recognized features of placental insufficiency⁶, or</td>
<td></td>
</tr>
<tr>
<td>(c) Three or more unexplained consecutive spontaneous abortions before the 10th wk of gestation, with maternal anatomic or hormonal abnormalities and paternal and maternal chromosomal causes excluded</td>
<td></td>
</tr>
</tbody>
</table>

Definite APS is present if at least one of the clinical criteria and one of the laboratory criteria are met. Classification of APS should be avoided if less than 12 wk or more than 5 yr separate the positive antiphospholipid antibody test and the clinical manifestation. In studies of populations of patients who have more than one type of pregnancy morbidity, investigators are strongly encouraged to stratify groups of subjects according to a, b, or c above.

*Coexisting inherited or acquired factors for thrombosis are not reasons to exclude patients from APS trials. However, two subgroups of APS patients should be recognized by the presence and absence of additional risk factors for thrombosis. Indicative (but not exhaustive) cases include age (>55 yr in men and >65 yr in women), the presence of any of the established risk factors for cardiovascular disease (hypertension, diabetes mellitus, elevated low-density lipoprotein or low high-density lipoprotein cholesterol, cigarette smoking, family history of premature cardiovascular disease, body mass index >30 kg/m², microalbuminuria, estimated glomerular filtration rate <60 mL/min), inherited thrombophilia, oral contraceptive use, nephritic syndrome, malignancy, immobilization, and surgery. Thus, patients who fulfill these criteria should be stratified according to contributing causes of thrombosis.

1 A thrombotic episode in the past can be considered a clinical criterion, provided that thrombosis is proved by appropriate diagnostic means, and that no alternative diagnosis or cause of thrombosis is found.

2 Superficial venous thrombosis is not included in the clinical criteria.

3 Generally accepted features of placental insufficiency include an abnormal or nonreassuring fetal surveillance test (e.g., nonreactive nonstress test) suggestive of fetal hypoxemia, an abnormal Doppler flow velocimetry waveform analysis suggestive of fetal hypoxemia (e.g., absent end-diastolic flow in the umbilical artery), oligohydramnios (e.g., amniotic fluid index ≤5 cm), or a postnatal birth weight less than the 10th percentile for gestational age.

4 Investigators are strongly advised to classify APS patients participating in studies into one of the following categories: I, more than one laboratory criteria present (any combination); IIa, only lupus anticoagulant present; IIb, only anticardiolipin antibody present; III, only anti–β₂-glycoprotein I antibody present. APS, antiphospholipid syndrome; ELISA, enzyme-linked immunosorbent assay.


in the elimination of apoptotic cells³ and as a natural anticoagulant.⁸

Other, less relevant antigens targeted by aPLs are prothrombin, annexin V, protein C, protein S, high- and low-molecular-weight kininogens, tissue plasminogen activator, factor VII, factor XI, factor XII, complement component C4, and complement factor H.¹⁹

In experimental animal models, passive or active immunization with viral peptides,²⁰ bacterial peptides,²¹ and heterologous β₂-GPI²² induces polyclonal aPLs and clinical events associated with APS. These data suggest that pathologic autoimmune aPL is induced in susceptible humans by infection via molecular mimicry.

However, infection-induced aPLs (syphilitic and non-syphilitic Treponema, Borrelia burgdorferi, human immunodeficiency virus, Leptospira, or parasites) are usually β₂-GPI independent and bind phospholipids directly.²³ Drugs (chlorpromazine, procainamide, quinidine, and phenytoin) and malignancies (lymphoproliferative disorders) can also induce β₂-GPI-independent aPLs. Conversely, autoimmune aPLs bind β₂-GPI or other phospholipid-binding plasma proteins, which in turn bind negatively charged phospholipids such as cardiolipin (β₂-GPI-dependent aPLs).

Low levels of aPLs may be present normally; one of the functions of normal aPLs may be to participate in the physiologic removal of oxidized lipids.

**Table 82-2** Other Features Suggesting the Presence of Antiphospholipid Antibodies

<table>
<thead>
<tr>
<th>Clinical</th>
<th>Laboratory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Livedo reticularis</td>
<td>Lupus anticoagulant present in plasma on two or more occasions at least 12 wk apart, detected according to the guidelines of the International Society on Thrombosis and Hemostasis</td>
</tr>
<tr>
<td>Thrombocytopenia (usually 50,000-100,000 platelets/mm³)</td>
<td>1. Lupus anticoagulant present in plasma on two or more occasions at least 12 wk apart, detected according to the guidelines of the International Society on Thrombosis and Hemostasis</td>
</tr>
<tr>
<td>Autoimmune hemolytic anemia</td>
<td>2. Anti–β₂-glycoprotein I antibody of immunoglobulin (Ig)G or IgM isotype in serum or plasma, present in medium or high titer (&gt;40 GPL or MPL, or &gt;99th percentile), on two or more occasions at least 12 wk apart, measured by a standardized ELISA</td>
</tr>
<tr>
<td>Cardiac valve disease (vegetations or thickening)</td>
<td>3. Anti-β₂-glycoprotein I antibody of immunoglobulin (Ig)G or IgM isotype in serum or plasma (in titer &gt;99th percentile) present on two or more occasions at least 12 wk apart, measured by a standardized ELISA</td>
</tr>
<tr>
<td>Multiple sclerosis–like syndrome, chorea, or other myelopathy</td>
<td></td>
</tr>
</tbody>
</table>

**PATHOGENESIS**

Antiphospholipid antibody is most likely related to thrombosis through multiple mechanisms; a proposed pathogenesis is illustrated in Figure 82-1. The process begins with activation or apoptosis of platelets, endothelial cells, or trophoblasts, during which phosphatidylserine (a negatively charged phospholipid) migrates from the inner to the normally electrically neutral outer cell membrane. Circulating β₂-GPI binds to phosphatidylserine, and then aPL binds to a β₂-GPI dimer.²⁴
Antiphospholipid antibody–β2GPI dimer binding activates the complement cascade extracellularly; initiates an intracellular signaling cascade, probably through the C5a and β2GPI surface receptors; and recruits and activates inflammatory effector cells, including monocytes, neutrophils, and platelets, leading to the release of proinflammatory products (e.g., tumor necrosis factor [TNF], oxidants, proteases) and the induction of a prothrombotic phenotype. The putative receptor of β2GPI binding protein that transduces signals from the cell membrane to the nucleus is not yet identified and may vary among cells. The following candidates have been suggested: apoER2 (a member of the low-density lipoprotein receptor superfamily), annexin A2, and a Toll-like receptor.

In addition, through downregulation of the signal transducer and activator of transcription 5 (STAT5), aPLs inhibit the production of placental prolactin and insulin growth factor–binding protein-1, and they adversely affect the formation of a trophoblast syncytium, placental apoptosis, and trophoblast invasion—all processes that are required for the normal establishment of placental function.

Other possible contributory mechanisms of aPL-mediated thrombosis include inhibition of coagulation cascade reactions catalyzed by phospholipids (e.g., activation of circulating procoagulant proteins, inhibition of protein C and S activation), induction of tissue factor (a physiologic initiator of coagulation) expression on monocytes, polymorphonuclear (PMN) cells, and platelets, resulting in the release of proinflammatory mediators (e.g., tumor necrosis factor [TNF], vascular endothelial growth factor receptor-1) and initiation of the prothrombotic stage. Both nuclear factor κB (NFκB) and p38 mitogen-activated protein kinase (p38 MAPK) may play a role in the intracellular signaling cascade. Antiphospholipid antibodies also downregulate the expression of trophoblast signal transducer and activator of transcription 5 (STAT5), reducing the endometrial stromal cell production of prolactin (PRL) and insulin growth factor binding protein-1 (IGFBP-1).
epidermal growth factor–like growth factor leads to defective placentation.39

Because high-level aPLs may persist for years in asymptomatic persons, it is likely that vascular injury, endothelial cell activation, or both immediately precede the occurrence of thrombosis in those bearing the antibody (second-hit hypothesis). Of note, at least 50% of APS patients with vascular factors possess other acquired thrombosis risk factors at the time of their events.50,51

Persons congenitally lacking β2GPI42 and β2GPI knock-out mice appear normal.43 β2GPI polymorphisms influence the generation of aPLs in individuals, but they have only a weak relationship to the occurrence of APS.44 A cluster of 50 upregulated genes may have an effect on the occurrence of thrombosis in aPL-positive individuals.45

**CLINICAL FEATURES**

Clinical manifestations range from asymptomatic aPL positivity (no history of vascular or pregnancy events) to catastrophic APS (multiple thromboses occurring over days). Thus patients should not be evaluated and managed as if they have a single disease.

**Vascular Occlusion**

APS affects all organ systems. Its principal manifestations are venous or arterial thromboses and pregnancy loss (see Table 82-1). Except for their severity, the youth of affected patients, and the unusual anatomic locations (Budd-Chiari syndrome; sagittal sinus and upper extremity thromboses), venous thromboses in APS do not differ clinically from thromboses attributable to other causes. Similarly, arterial thromboses differ from non–aPL-associated thromboses only by their recurrent nature, unusual locations, and occurrence in young patients. Deep vein thrombosis and stroke are the most common clinical manifestations of APS. Renal thrombotic microangiopathy, glomerular capillary endothelial cell injury, and thrombosis of renal vessels cause proteinuria without cells in the urine or hypocomplementemia and may lead to severe hypertension, renal failure, or both.46

**Pregnancy Morbidity**

Pregnancy losses in patients with aPLs typically occur after 10 weeks' gestation (fetal loss), although earlier losses also occur. These pre-embryonic or embryonic pregnancy losses (<10 weeks' gestation) are commonly due to chromosomal and other genetic defects. Pregnancy in those with APS is often normal until the second trimester, when fetal growth slows and amniotic fluid volume decreases. APS patients may develop severe, early preeclampsia or HELLP (hemolysis, elevated liver enzymes, low platelets) syndrome. Placental infarction is a cause of fetal growth restriction or death; nonthrombotic mechanisms of placental dysfunction also may occur.47 Prior late pregnancy losses predict future losses, independent of the aPL profile.

**Miscellaneous and Noncriteria Manifestations**

Many patients have livedo reticularis or thrombocytopenia (Figure 82-2), although these conditions are not specific for APS. Cardiac valve disease (vegetations, thickening, or both), a late manifestation, may necessitate valve replacement. Its pathogenesis in APS is unknown. Recent studies suggest that APS does not add to the risk of atherosclerosis imparted by SLE.48 Pulmonary hypertension may develop owing to recurrent pulmonary embolism or small vessel thrombosis; rarely, aPL-positive patients may present with diffuse pulmonary hemorrhage. Some patients develop nonfocal neurologic symptoms such as lack of concentration, forgetfulness, and dizzy spells. Multiple small, hypertensive lesions seen on magnetic resonance imaging (MRI), primarily in the periventricular white matter, do not correlate well with clinical symptoms. Rarely, high-affinity antiprothrombin antibodies may cause hemorrhage by depleting prothrombin (lupus anticoagulant hypoprothrombinemia syndrome).49

**Catastrophic Antiphospholipid Syndrome**

Catastrophic APS is a rare, abrupt, life-threatening complication. It consists of multiple thromboses of medium and small arteries occurring (despite apparently adequate anticoagulation) over a period of days and causing stroke; cardiac, hepatic, adrenal, renal, and intestinal infarction; and peripheral gangrene.45 In a review of 220 patients with catastrophic APS, the main clinical manifestations included renal involvement in 154 patients (70%), pulmonary in 146 (66%), cerebral in 133 (60%), cardiac in 115 (52%), and cutaneous in 104 (47%).50 Acute adrenal failure may be the initial clinical event. Proposed formal criteria for this syndrome are shown in Table 82-3.51 Patients often have moderate thrombocytopenia; erythrocytes are less fragmented than in the hemolytic uremic syndrome or thrombotic thrombocytopenic purpura, and fibrin split products are not strikingly elevated. Renal failure and pulmonary hemorrhage may occur. Tissue biopsies show noninflammatory vascular occlusion.

**DIAGNOSIS AND DIAGNOSTIC TESTS**

**Laboratory Studies**

The diagnosis of APS requires a positive lupus anticoagulant test or a moderate- to high-titer anticardiolipin or anti-β2GPI IgG or IgM test in patients with characteristic clinical manifestations. Patients with negative lupus anticoagulant
Hypocomplementemia, erythrocyte casts, and IgM anticardiolipin is standardized, considerable variability exists among commercial laboratories that perform the test, especially for the IgA isotype.51 Low-titer anticardiolipin or anti-β2GPI, transient aPLs, and antibody to noncardiolipin phospholipids (phosphatidylserine, phosphatidylethanolamine) have no or less proven relationship to APS. ELISA tests other than anticardiolipin and anti-β2GPI are neither standardized nor widely accepted as predictors of clinical illness.

A false-positive test for syphilis does not fulfill the laboratory criteria, but it should alert physicians to order the aPL tests previously described, especially in patients with a history of aPL-related clinical manifestations.

Whether to test persons with venous occlusive disease or recurrent fetal loss simultaneously for protein C, protein S, and antithrombin III deficiency or for the factor V Leiden and prothrombin mutations is a matter of economics and clinical likelihood; such testing is advisable when feasible. It is useful to test persons with arterial occlusive disease for hyperhomocysteinemia.

Antinuclear and anti-DNA antibodies occur in approximately 45% of patients clinically diagnosed as having primary APS without an accompanying illness; these antibodies do not mandate the additional diagnosis of SLE if the patient has no clinical indicators of SLE. Thrombocytopenia in APS is usually modest (>50,000/mm3); proteinuria and renal insufficiency occur in patients with thrombotic microangiopathy. Pathologic examination demonstrates small artery and glomerular thrombi and recanalization (Figure 82-3). Hypocomplementemia, erythrocyte casts, and pyuria are not characteristic of thrombotic microangiopathy and, when present, imply lupus glomerulonephritis. Erythrocyte sedimentation rate, hemoglobin, and leukocyte count are usually normal in patients with uncomplicated primary APS, except during acute thrombosis. Prothrombin fragment 1 + 2 and other markers of coagulation activation do not predict impending thrombosis.

**Imaging Studies**

MRI shows vascular occlusion and infarction consistent with clinical symptoms, with no special characteristics (other than multiple, otherwise unexplained cerebral infarctions in a young person). Multiple small, hyperintense white matter lesions are common and do not unequivocally imply brain infarction. Occlusions usually occur in vessels below the resolution limits of angiography; hence, angiography or magnetic resonance angiography is not indicated unless clinical findings suggest medium- or large-vessel disease. Echocardiography or cardiac MRI may show severe Libman-Sacks endocarditis and intracardiac thrombi.59

**Pathology**

Skin, renal, and other tissues show thrombotic occlusion of all caliber arteries and veins, acute and chronic endothelial injury and its sequelae, and recanalization in late lesions. Uteroplacental insufficiency was once thought to be due to thrombosis or spiral artery vasculopathy (atherosclerosis, intimal thickening, fibrinoid necrosis, and absence of physiologic changes in the spiral arteries).60 Consistent with the importance of inflammation in murine models of APS, recent
findings demonstrate inflammatory infiltrates, particularly macrophages, and suggest that inflammation contributes to placental injury in patients.\textsuperscript{61} The finding of necrotizing vasculitis suggests concomitant lupus or other connective tissue disease. No other diagnostic immunofluorescence or electron microscopic findings have been reported.

**Differential Diagnosis**

Infection-induced anticardiolipin is usually transient and is more commonly IgM than IgG.\textsuperscript{62} Transient aPLs or low-titer anticardiolipin is inconclusive for diagnosis. Research laboratories can distinguish autoimmune from infection-induced aPLs by determining the antibody's β2GPI dependence. In a patient who has lupus or lupus-like disease, livedo reticularis, or long-standing thrombotic thrombocytopenia and who has a persistently positive aPL test in addition to APS-related symptoms, it is usually unnecessary to exclude other diagnoses.

Because the prevalence of aPL-positive ELISA tests increases with age, and because the differential diagnosis of vascular occlusion is broader than it is in young adults, particular care is necessary in diagnosing APS in patients older than 60 years. Sustained high-titer anticardiolipin IgG, livedo reticularis, thrombotic thrombocytopenia, coexisting rheumatic disease, and absence of other causes support a diagnosis of APS.

Five percent to 21% of women with recurrent pregnancy losses, and 0.5% to 2% of normal pregnant women, have aPLs. Heritable deficiency of protein C, protein S, and antithrombin III and the presence of the factor V Leiden (A506G), prothrombin (G20210A), and methylene tetrahydrofolate reductase (MTHFR, C677T) mutations are less common causes of fetal loss.\textsuperscript{63} Attribution of pregnancy loss to APS is most certain when no coexisting plausible explanation is known, when the loss occurs after demonstration of a fetal heartbeat (10 weeks), when a significant aPL profile is repeatedly positive before and after pregnancy, and when the placenta shows vasculopathy and infarction. A single pregnancy loss before 10 weeks' gestation in a patient with a low-positive anticardiolipin test is more likely to be attributable to fetal chromosomal abnormalities, infection, or maternal hormonal or anatomic abnormalities.

Independent coagulopathies may further increase thrombotic risk in patients with aPLs. These and other acquired thrombotic risk factors (hypertension, diabetes, nephrotic syndrome, venous insufficiency, immobility) are alternative causes of thromboembolic disease. Arterial occlusion occurs in patients with thrombotic thrombocytopenic purpura, infected or sterile emboli of cardiac or vascular origin, septicemia, hyperhomocysteinemia, myxoma, Takayasu's arteritis, polyarteritis nodosa, and severe Raynaud's disease. The relationship of Sneddon's syndrome (stroke and livedo reticularis, with or without aPLs) to APS is uncertain.

Catastrophic APS has few mimics. Among them are sepsis, disseminated intravascular coagulation, thrombotic thrombocytopenic purpura, hemolytic uremic syndrome,
Antiphospholipid antibodies are rarely present in patients with the alternative diagnoses. Acute adrenal insufficiency is characteristic of APS and Wahrman-Friderichsen syndrome.

**TREATMENT**

**Thrombosis**

Thrombosis recommendations are summarized in Table 82-4. Anticoagulation with heparin is the treatment for acute thrombosis in APS patients. Warfarin, occasionally in association with low-dose aspirin, is used for secondary thrombosis prophylaxis. Two randomized, controlled trials demonstrated that moderate warfarin (international normalized ratio [INR], 2 to 3) and high-intensity warfarin (INR, 3 to 4) are equally protective against recurrence in APS patients after the first thrombosis. The intensity of anticoagulation for aPL-related arterial thrombosis is still a matter for debate because in both studies, patients with arterial events continued less than half of the study population. Although APS patients with arterial thrombosis who are at high risk for recurrence may require high-intensity anticoagulation, in the absence of risk-stratified studies, high risk has no consensus definition and currently is based solely on clinical judgment.

Aspirin is the standard of care after an ischemic stroke or a transient ischemic attack to prevent a recurrence in aPL-negative patients. Although most aPL-positive patients with ischemic strokes receive warfarin, the Antiphospholipid Antibody in Stroke Study (APASS) concluded that for selected aPL-positive patients who do not have atrial fibrillation or high-grade stenosis, aspirin and warfarin (target INR, 2.2) are equivalent in terms of both efficacy and major bleeding complications. The APASS results probably do not apply to conventionally defined APS because the average age of study participants was much higher than that of the average APS population; in addition, the aPL determination was performed only once at study entry, and the inhibitor cutoff for assigning a patient to the positive anticardiolipin group was very low. However, aspirin is an option for older aPL-positive patients who have a single low-titer anticardiolipin test, and whose presentation is one stroke.

Some patients require larger than expected doses of both heparin and warfarin to achieve therapeutic anticoagulation. Uncommonly, positive lupus anticoagulant tests cause the INR to be unreliable. Patients with constantly fluctuating INR levels and/or recurrent events despite therapeutic INR may be monitored by the measurement of anti–factor Xa activity or another appropriate assay.

For well-anticoagulated patients who continue to have thromboses, the addition of aspirin (81 to 325 mg/day) can be considered. Experimental (in vitro and/or animal models) and clinical evidence (in lupus patients) suggests that hydroxychloroquine may decrease the incidence of thrombosis; similarly, experimental evidence indicates that statins can interfere with aPL-mediated thrombosis. However, controlled studies are needed to determine the effectiveness of hydroxychloroquine and statins in aPL-positive patients. Corticosteroids have no established role in the treatment of APS but are used for rheumatic symptoms in patients with accompanying systemic autoimmune illnesses. However, high doses of corticosteroids usually are given empirically to patients with severe thrombocytopenia, hemolytic anemia, and catastrophic APS.

No controlled studies in APS patients have been published for clopidogrel, pentoxifylline, aspirin-dipyridamole, argatroban, hirudin, and other new anticoagulant agents. Neither hirudin nor fondaparinux inactivates complement, and neither drug protects mice with experimental APS. However, high doses of corticosteroids usually are given empirically to patients with severe thrombocytopenia, hemolytic anemia, and catastrophic APS.

Currently available data, although retrospective and not risk-stratified, indicate that lifelong anticoagulation of APS patients who have had vascular events is appropriate. However, recent recognition that some patients have full remission of antibody, and that most thrombotic events have recognizable triggers, raises the possibility of discontinuing anticoagulation in highly selected patients when the triggers are eliminated.

---

**Table 82-4 Treatment Recommendations for Persistently Antiphospholipid Antibody–Positive Individuals**

<table>
<thead>
<tr>
<th>Clinical Circumstance</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptomatic</td>
<td>No treatment*</td>
</tr>
<tr>
<td>Venous thrombosis</td>
<td>Warfarin INR 2.5 indefinitely</td>
</tr>
<tr>
<td>Arterial thrombosis</td>
<td>Warfarin INR 2.5 indefinitely</td>
</tr>
<tr>
<td>Recurrent thrombosis</td>
<td>Warfarin INR 3-4 ± low-dose aspirin</td>
</tr>
<tr>
<td>Pregnancy:</td>
<td></td>
</tr>
<tr>
<td>First pregnancy</td>
<td>No treatment†</td>
</tr>
<tr>
<td>Single pregnancy loss at &lt;10 wk</td>
<td>No treatment†</td>
</tr>
<tr>
<td>≥1 Fetal or ≥3 (pre)-embryonic losses, no thrombosis</td>
<td>Prophylactic heparin† + low-dose aspirin throughout pregnancy, discontinue 6-12 wk postpartum</td>
</tr>
<tr>
<td>Thrombosis regardless of pregnancy history</td>
<td>Therapeutic heparin† or low-dose aspirin throughout pregnancy, warfarin postpartum</td>
</tr>
<tr>
<td>Valve nodules or deformity</td>
<td>No known effective treatment; full anticoagulation if emboli or intracardiac thrombi demonstrated</td>
</tr>
<tr>
<td>Thrombocytopenia &gt;50,000/mm³</td>
<td>No treatment</td>
</tr>
<tr>
<td>Thrombocytopenia &lt;50,000/mm³</td>
<td>Prednisone, IVIG</td>
</tr>
<tr>
<td>Catastrophic antiphospholipid syndrome</td>
<td>Anticoagulation + corticosteroids + IVIG or plasmapheresis</td>
</tr>
</tbody>
</table>

*Aspirin (81 mg/day) may be considered in high-risk patients with multiple non-aPL cardiovascular risk factors.
†Aspirin (81 mg/day) may be considered.
‡Enoxaparin 0.5 mg/kg subcutaneously once daily.
§Enoxaparin 1 mg/kg subcutaneously twice daily or 1.5 mg/kg subcutaneously once daily.
INR, international normalized ratio; IVIG, intravenous immunoglobulin.
**Pregnancy Morbidity**

Heparin is indicated at the diagnosis of pregnancy in an aPL-positive woman who has had prior pregnancy losses attributable to APS. Because warfarin is teratogenic, only unfractionated or low-molecular-weight heparin is used for the treatment of affected pregnancies in the United States; in other countries, converting to warfarin after the first trimester may be considered acceptable. Most physicians with a special interest in this field now use low-molecular-weight heparin owing to decreased risk of thrombocytopenia and osteoporosis.

Patients with prior fetal losses later than 10 gestational weeks should be treated with prophylactic heparin (enoxaparin 30 to 40 mg subcutaneously once daily), together with low-dose aspirin; this regimen increases the fetal survival rate from 50% (untreated) to 80%. Women who have had prior thromboses must be fully anticoagulated (enoxaparin 1 mg/kg subcutaneously twice daily or 1.5 mg/kg subcutaneously once daily) throughout pregnancy because the risk of new thrombosis markedly increases both during pregnancy and during the postpartum period. Even with treatment, prematurity and fetal growth restriction may occur.

In aPL-positive women with prior thrombosis, warfarin is changed to heparin or low-molecular-weight heparin before conception, if possible, or at the first missed menstrual period. In aPL-positive women without prior thrombosis, if heparin is indicated during pregnancy, treatment begins after confirmation of pregnancy, continues until 48 hours before anticipated delivery (to allow epidural anesthesia), and resumes for 8 to 12 weeks during the postpartum period. Some physicians recommend the initiation of heparin before conception; no clinical trial supports this recommendation, however, and the risk associated with longer-duration heparin therapy is considerable. Patients in most published series received low-dose aspirin as well as heparin, but the benefit of adding aspirin is unknown.

Because of the risk of postpartum thrombosis, it is prudent to continue anticoagulation for 8 to 12 weeks during the postpartum period and then to discontinue it by tapering the doses. If desired, conversion from heparin to warfarin may be accomplished after the first or second postpartum week. Breastfeeding is permissible with both heparin and warfarin.

No studies unequivocally justify the treatment of women with aPLs during a first pregnancy, women with only very early losses, or women whose aPL titers are low or transient. Nonetheless, it is common to offer such patients low-dose aspirin.

**Asymptomatic Antiphospholipid Antibody–Positive Individuals**

Persistence of aPLs for decades without clinical events is well documented. The probability that an asymptomatic person incidentally found to have aPLs will eventually develop the syndrome is likely low. Anticoagulation is not indicated for the prophylactic treatment of asymptomatic aPL-positive individuals. For those with moderate- to high-titer antiphospholipid and persistent aPLs, education about the meaning of abnormal tests is appropriate, as is a discussion of warning signs to be reported. Elimination of reversible thrombosis risk factors and prophylaxis during high-risk periods, such as surgical procedures, are crucial. The necessity for and the effectiveness of aspirin are not supported by the current literature; in a recent randomized, double-blind, placebo-controlled trial, low-dose aspirin (81 mg) appeared to be no better than placebo in preventing first thrombotic episodes in persistently asymptomatic aPL-positive patients; the incidence rate of first thrombosis was relatively low.

Although drugs that induce lupus (hydralazine, phenytoin) may also induce aPLs, they may be prescribed for patients with aPLs if no alternatives are available. Drugs that promote thrombosis (estrogen and estrogen-containing oral contraceptives) are not currently deemed safe, even for asymptomatic women serendipitously known to have high-titer antibodies. This advice does not translate to a recommendation to test all normal women before prescribing such medications, but it does suggest that special attention and further evaluation should be provided to those with family histories or clinical suggestions of rheumatic disease, livedo reticularis, biologic false-positive tests for syphilis, or borderline thrombocytopenia. No reliable information is available regarding the safety of progestin-only contraception, “morning after” contraception, or the use of raloxifene, bromocriptine, or leuprolide in APS patients. A small retrospective review of women undergoing artificial reproductive technology (in vitro fertilization) procedures demonstrated no thrombotic events.

**Antiphospholipid Antibody–Positive Individuals with Ambiguous Events**

Some patients with positive aPL tests have clinical events of ambiguous meaning (dizzy or confusional episodes, nonspecific visual disturbances, very early pregnancy loss). No consensus has been reached regarding the treatment of such persons. Because full anticoagulation carries high risk, many physicians prescribe low-dose (81 mg) aspirin, hydroxychloroquine, or both daily. No published data support or repudiate this recommendation. Based on the presumed pathogenesis, some physicians prescribe anticoagulation for patients with livedo reticularis, thrombocytopenia, leg ulcers, thrombotic microangiopathy, or valvulopathy. The efficacy of anticoagulation is unknown in these conditions. One small, descriptive, cross-sectional study provides evidence that B cell depletion with rituximab is well tolerated and can be effective for refractory thrombocytopenia and skin ulcers in aPL-positive patients. An open-label phase IIa descriptive pilot study assessing the effectiveness and safety of rituximab in patients with noncriteria and/or anticoagulation-resistant manifestations of aPL is in progress (ClinicalTrials.gov Identifier: NCT00537290).

**Catastrophic Antiphospholipid Syndrome**

The onset of catastrophic APS is usually sudden and immediately life threatening. Early diagnosis can be a challenge, especially in patients with no history of APS. However,
Antiphospholipid Antibody–Negative Individuals with a Clinical Event

In patients clinically suspected of having APS but with normal antiphospholipin, lupus anticoagulant, and anti-β2-GPI tests, alternative causes of clotting must be sought. Even among patients with concomitant rheumatic disease, APS may not be the cause of recurrent thromboembolism or pregnancy loss. Patients with SLE develop emboli from SLE-related cardiac valvular disease, vasculitis, or atheroma. Other patients have factor V Leiden or some other procoagulant mutation. Recurrent pregnancy losses may be caused by chromosomal abnormalities, uterine infection, diabetes, hypertension, or non-aPL coagulopathy. The concept of "seronegative" APS is not recognized.

OUTCOME

Pulmonary hypertension, neurologic involvement, myocardial ischemia, nephropathy, gangrene of extremities, and catastrophic APS are associated with a worse prognosis. During long-term follow-up, serious morbidity and disability occur in an unpredictable proportion of primary APS patients who experience major vascular events and in those who have delays in diagnosis and treatment. Thus, the long-term functional outcome of primary APS patients is poor; at 10 years, one-third of patients develop permanent organ damage, and one-fifth are unable to perform everyday activities.

In a retrospective study of obstetric APS patients without thrombosis, 35% developed aPL-related clinical events during 8 years of follow-up. The studied populations were highly selected referral populations that may have been biased toward severe disease, but follow-up studies of obstetric patients with autoantibodies show similar results.

Long-term outcomes of children born of APS pregnancies are not known. In many patients with long-standing APS, the development of severe cardiac valvular disease necessitates valve replacement, and rare patients develop renal failure due to thrombotic microangiopathy. Immediate thrombosis may cause loss of a transplanted kidney or other organ; aPL positivity correlates with poor graft survival after renal transplantation in SLE patients.

Serious perioperative complications may occur despite prophylaxis in aPL-positive patients because they are at additional risk for thrombosis when undergoing surgical procedures. Thus perioperative strategies should be clearly identified before any surgical procedure is performed; in addition, pharmacologic and physical antithrombosis interventions should be vigorously employed, periods without anticoagulation should be kept to an absolute minimum, intravascular manipulation for access and monitoring should be minimized, and any deviation from a normal course should be considered a potential disease-related event.

References

1340

PART 11

|

SYSTEMIC LUPUS ERYTHEMATOSUS AND RELATED SYNDROMES

23. Arvieux J, Renaudineau Y, Mane I, et al: Distinguishing features of
anti-beta2 glycoprotein I antibodies between patients with leprosy and
the antiphospholipid syndrome, Thromb Haemost 87:599–605,
2002.
2-glycoprotein 1 increase platelet deposition to collagen via interaction with phospholipids and the apolipoprotein E receptor 2′, J Biol
antibody binding makes negatively charged phospholipids accessible
to antiphospholipid antibodies, Arthritis Rheum 41:1738–1747,
1998.
endothelial cells by antiphospholipid antibodies, J Clin Invest 96:2211–
2219, 1996.
monoclonal anticardiolipin antibodies in platelet interaction with
subendothelium under flow conditions, Arthritis Rheum 46:3283–3289,
2002.
28. van Lummel M, Pennings MT, Derksen RH, et al: The binding site in
β2-glycoprotein I for ApoER2 on platelets is located in domain V, J
30. Romay-Penabad Z, Montiel-Manzano MG, Shilagard T: Annexin A2
is involved in antiphospholipid antibody-mediated pathogenic effects
NFkappaB is an essential intermediate in the activation of endothelial
cells by anti-beta2-glycoprotein 1 antibodies, Thromb Haemost
32. Pierangeli SS, Vega-Ostertag M, Harris EN: Intracellular signaling
triggered by antiphospholipid antibodies in platelets and endothelial
transducer and activator of transcription, Stat5, and its enhancement
of PRL expression in human endometrial stromal cells in vitro, J Clin
mouse model by IgG, IgM, and IgA immunoglobulins from patients
with the antiphospholipid syndrome, Thromb Haemost 74:1361–1367,
1995.
β2-glycoprotein I-dependent antiphospholipid antibodies in a
photochemically-induced thrombosis model in the hamster, Blood
restore mesenteric ischemia/reperfusion-induced injury in complement receptor 2/complement receptor 1-deficient mice, J Immunol
induces dysregulation of angiogenic factors and causes fetal rejection
39. Di Simone N, Marana R, Castellani R: Decreased expression of
heparin-binding epidermal growth factor-like growth factor as a newly
revised antiphospholipid syndrome (APS) classification criteria
42. Bancsi LF, van der Linden IK, Bertina RM: Beta 2-glycoprotein I
deficiency and the risk of thrombosis, Thromb Haemost 67:649–653,
in beta 2-glycoprotein I null mice, J Biol Chem 276:13817–13821,
44. Kamboh MI, Manzi S, Mehdi H, et al: Genetic variation in apolipoprotein H (beta2-glycoprotein I) affects the occurrence of antiphospholipid antibodies and apolipoprotein H concentrations in systemic

phenotypes of immune-mediated thrombosis, Blood 107:1391–1396,
2006.
accelerated atherosclerosis in systemic lupus erythematosus, N Engl J
1999.
50. Erkan D, Cervera R, Asherson RA: Catastrophic antiphospholipid
antiphospholipid syndrome: a systematic review of the literature, Blood
lupus anticoagulants: an update, Thromb Haemost 74:1185–1190,
1995.
57. Erkan D, Derksen WJ, Kaplan V, et al: Real world experience with
antiphospholipid antibody tests: how stable are results over time? Ann
58. Lockshin MD, Sammaritano LR, Schwartzman S: Brief report: validation of the Sapporo criteria for antiphospholipid antibody syndrome,
genetic thrombophilia in women with complications of pregnancy, N
64. Crowther MA, Ginsberg JS, Julian J, et al: Comparison of two intensities of warfarin for the prevention of recurrent thrombosis in patients
with the antiphospholipid antibody syndrome, N Engl J Med 349:1133–
of high-intensity warfarin vs conventional antithrombotic therapy for
subsequent thrombo-occlusive events in patients with ischemic stroke,
67. Ortel TL, Moll S: Monitoring warfarin therapy in patients with lupus




The references for this chapter can also be found on www.expertconsult.com.
Scleroderma or systemic sclerosis (SSc) is an uncommon disease of unknown cause and complex pathogenesis. The hallmarks of SSc are (1) autoimmunity, (2) inflammation, (3) functional and structural alterations in small blood vessels, and (4) interstitial and vascular fibrosis in the skin and internal organs. This unique constellation of distinct but related pathophysiologic features, illustrated in Figure 83-1, accounts for the characteristic clinical manifestations of SSc. Early-stage disease may be dominated by inflammation and vascular injury, whereas in advanced disease fibrosis and vascular insufficiency are most prominent. However, there is enormous patient-to-patient variability in these features. Recent advances in cell and molecular biology, mouse genetic engineering, functional genomics, and genetic association studies reveal the involvement of a large number of molecules, pathways, and cell types in SSc, yielding an increasing nuanced picture of the pathogenesis. Environmental triggers in a genetically susceptible individual are thought to induce a cascade of events with early vascular injury, immune cell activation, the generation of autoimmunity, and subsequent fibroblast activation and matrix accumulation that results in chronic and progressive tissue damage. Over time, vascular insufficiency and widespread fibrosis cause disruption of vital organs, accounting for the substantial morbidity and mortality of SSc.

ETIOLOGY

Neither the cause of SSc nor the precise contribution of genetic factors is fully understood. Evidence indicates that infectious agents, environmental toxins, and drugs, as well as microchimerism, are potential triggers.

Genetic Risk: Family Studies

Familial clustering of a disease is considered as evidence of inherited disease susceptibility, but such clustering might be explained by shared environmental exposures, shared genetic background, or the interaction between genes and environment. The risk of SSc is increased among first-degree relatives of SSc cases compared with the general population. In one U.S. study, the relative risk of SSc among first-degree relatives of cases was 13, with a rate of 1.6% compared with 0.026% in the general population, identifying a family history of SSc as the strongest known risk factor. The only twin study of SSc to date reported a relatively small disease concordance rate (4.7%), although the concordance rate for positive antinuclear antibody (ANA) was 90% for monozygotic (identical) twins and 40% for dizygotic (fraternal) twins. Raynaud disease and pulmonary fibrosis show increased prevalence in pedigrees of SSc patients. Moreover, autoimmune diseases among first-degree family members of SSc patients have been reported in up to 36% of cases, with hypothyroidism, hyperthyroidism, rheumatoid arthritis, and systemic lupus erythematosus (SLE) most common.
The past decade has witnessed rapid progress in delineating genetic susceptibility factors in SSc. Genetic association studies using candidate gene approaches and, more recently, genome-wide association (GWA) studies have been performed in large multinational patient cohorts. The major histocompatibility complex (MHC) is the dominant genetic region implicated in autoimmune disease, although the role of specific human leukocyte antigen (HLA) alleles in pathogenesis remains unknown. The interpretation of HLA associations is complicated by the extensive linkage disequilibrium of risk haplotypes. Specific HLA alleles have long been known to be associated with SSC and specific autoantibodies. For instance, a case-control study of SSc revealed strong associations with HLA DRB1*1104, DQA1*0501, and DQB1*0301 haplotypes. Candidate gene approaches typically look for changes in single nucleotides (single nucleotide polymorphisms [SNPs]), the most common form of deoxyribonucleic acid (DNA) variation. Non-HLA susceptibility genes associated with SSc include the protein tyrosine phosphatase nonreceptor 22 (PTPN22), which has been associated with SLE, myasthenia gravis, vitiligo, and Addison’s disease; interleukin (IL)-1β and NLRP1, an inflammasome scaffold that promotes pro-IL-1β maturation and processing; and interferon regulatory factor 5 (IRF5), a transcription factor in the Toll-like receptor (TLR) pathway that mediates type I interferon (IFN) induction and is associated with SLE, as well as SSc and interstitial lung disease. The association of IRF5 with SSc is particularly interesting, in light of the role of IFN type I in immune responses. Table 83-1 summarizes SSc susceptibility genes identified to date. In addition to classic SNPs, informative genetic polymorphisms in SSc include variations of copy numbers, rare allelic variants, and epigenetic changes. The identification of these genetic associations will require in-depth analysis using deep-sequencing technologies.

### Other Candidate Genes and Genome-wide Association Studies

Candidate gene approaches focusing on genes involved in vascular homeostasis or matrix remodeling have to date not yielded robust associations with SSc. A British study identified a functional SNP in the connective tissue growth factor (CTGF) promoter region associated with SSc. However, in another study an apparent association with SPARC could not be validated in an independent cohort. Potential explanations for discrepancies among genetic association studies include ethnic differences in the study populations and disease heterogeneity.

Several GWA studies in large and ethnically diverse populations are currently under way. These unbiased approaches have the advantage over candidate gene studies in that they are driven by discovery rather than a priori hypotheses. The first large-scale SSc GWA study analyzed 300,000 SNPs in a cohort of 2296 white cases and 5014 healthy controls. Statistically significant associations were found with SNPs in the HLA region, as well as IRF5, TNFAIP3, and CD247, a gene implicated in T cell signaling and also associated with susceptibility to SLE. Although genetic association studies represent a rapidly evolving area of research, the results to date can be summed up as follows: (1) Genetic variants associated with SSc susceptibility are involved in innate and adaptive immune responses, and (2) they are shared with SLE and other autoimmune diseases. It is worth noting that important associations can be missed by GWA studies using current technologies. Moreover, despite the wealth of emerging information, the genetic associations discovered to date are of relatively modest magnitude, with odds ratios that are generally less than 1.5. This finding points to the potential importance of gene-gene interactions (epistasis), particularly for genes within the same molecular pathways, and gene-environment interactions.

### Table 83-1  Systemic Sclerosis (SSc) Susceptibility Genes

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome</th>
<th>Associated SSc Subset</th>
<th>Potential Pathogenic Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA</td>
<td>6</td>
<td>Various</td>
<td>Antigen presentation</td>
</tr>
<tr>
<td>PTPN22</td>
<td>1p3.2</td>
<td>Topo1+ positive</td>
<td>T and B cell activation</td>
</tr>
<tr>
<td>NLRP1</td>
<td>17p13.2</td>
<td>dcSSc, pulmonary fibrosis</td>
<td>Inflammasome component, IL-1β processing</td>
</tr>
<tr>
<td>IRF5</td>
<td>7q32</td>
<td>dcSSc</td>
<td>Transcription factor required for induction of type I interferon</td>
</tr>
<tr>
<td>STAT4</td>
<td>2q32.3</td>
<td>IcSSc, ACA</td>
<td>Transcription factor for IL-12 and IL-23</td>
</tr>
<tr>
<td>BANK1</td>
<td>4q24</td>
<td>dcSSc</td>
<td>Adaptor involved in B cell signaling</td>
</tr>
<tr>
<td>TNFSF4</td>
<td>1q25</td>
<td>SSc</td>
<td>T cell co-stimulation</td>
</tr>
<tr>
<td>T-bet</td>
<td>17q21.32</td>
<td>SSc</td>
<td>Transcription factor for Th1 T cell polarization</td>
</tr>
</tbody>
</table>

Based on published candidate gene studies and genome-wide association studies as of 2011.

ACA, antcentromere antibody, dcSSc, diffuse cutaneous SSc; IL, interleukin; IcSSc, limited cutaneous SSc; SSc, systemic sclerosis.
interactions. A current challenge in SSc research is how to handle the large volume of new information emerging from GWA studies. Moreover, it will be important to delineate how genes shared among diverse autoimmune conditions contribute to disease-specific phenotypes. Shedding light on this important problem will require large collaborative studies involving phenotypically well-characterized populations of varied ethnic background and meta-analyses.

INFECTIOUS AGENTS: VIRUSES

Along with exposure to certain environmental and occupational agents and drugs, viruses such as human cytomegalovirus (hCMV) and parvovirus B19 have been implicated as potential triggers for SSc. Patients with SSc have anti-hCMV antibodies directed against the UL83 and UL94 protein epitopes on hCMV. Anti-UL94 antibodies can induce endothelial cell apoptosis and fibroblast activation, suggesting a direct role for antiviral antibodies in tissue damage. Anti-topoisomerase-I can cross-react with hCMV-derived proteins, implicating molecular mimicry as a mechanism linking hCMV infection and SSc. Cytomegalovirus is implicated in the pathogenesis of allograft vasculopathy, a complication of organ transplantation characterized by vascular neointima formation and smooth muscle cell proliferation strikingly reminiscent of the obliterator proliferative vasculopathy seen in SSc. In dermal fibroblasts, hCMV stimulates synthesis of the profibrotic growth factor connective tissue growth factor (CTGF, CCN2) in vitro. Evidence of human parvovirus B19 infection has also been reported in patients with SSc.

ENVIRONMENTAL EXPOSURES, DIETARY SUPPLEMENTS, DRUGS AND RADIATION

Although reports of putative geographic clustering suggest a role for shared environmental exposures, careful investigations have generally failed to substantiate apparent clusters of SSc. On the other hand, well-documented epidemics outbreaks of SSc-like illnesses with acute onset and chronic course have been reported. One such illness, called the toxic oil syndrome, occurred in Spain and was linked to the ingestion of contaminated rapeseed cooking oils. In the United States, dietary supplements containing L-tryptophan were implicated in an explosive outbreak of the eosinophilia-myalgia syndrome (EMS) in 1989. The EMS epidemic subsided following the ban on L-tryptophan, but sporadic cases of EMS following ingestion of L-tryptophan and other food supplements continue to be reported. Although scleroderma-like skin fibrosis was a prominent manifestation of these apparently de novo toxic-epidemic syndromes, along with multisystem involvement, chronic course, and autoimmunity, the associated clinical, histopathologic, and laboratory features clearly differentiate them from SSc. The frequency of SSc appears to be increased among men with occupational exposure to silica dust. This was recently confirmed by a meta-analysis of 16 observational studies, with risk estimates as high as 15. Other exposures linked with SSc include polystyrene chloride, toluene, xylene, trichloroethylene, and organic solvents. Additional reports alleged an association between SSc and environmental exposures to pesticides, hair dyes, and industrial fumes. Although exposure to cigarette smoke is known to increase the risk of multiple autoimmune diseases, there is no evidence to date implicating it as a risk factor for SSc.

Certain drugs have been implicated in SSc-like illnesses. The best studied is the anticancer drug bleomycin, which induces skin and lung fibrosis in the mouse (see later). Other potentially implicated drugs include pentazocine, the taxenes docetaxel and paclitaxel, and cocaine. The use of appetite suppressants has been linked to the development of pulmonary arterial hypertension (PAH). The occurrence of SSc in women following cosmetic breast augmentation with silicone implants raised significant concern regarding a possible causal association. Subsequent large-scale epidemiologic surveys and a meta-analysis, however, failed to confirm an increased risk of SSc or of other connective tissue diseases among women with silicone breast implants. Radiation treatment for malignant neoplasms has been linked with the onset of de novo SSc, as well as exacerbation of tissue fibrosis in patients with pre-existing SSc. Some of the environmental agents and drugs that have been linked with the development of SSc are listed in Table 83-2.

<table>
<thead>
<tr>
<th>Table 83-2</th>
<th>Environmental Agents and Drugs Implicated in Scleroderma-like Syndromes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chemicals</strong></td>
<td></td>
</tr>
<tr>
<td>Silica</td>
<td></td>
</tr>
<tr>
<td>Heavy metals</td>
<td></td>
</tr>
<tr>
<td>Mercury</td>
<td></td>
</tr>
<tr>
<td>Organic chemicals</td>
<td></td>
</tr>
<tr>
<td>Vinyl chloride</td>
<td></td>
</tr>
<tr>
<td>Benzene</td>
<td></td>
</tr>
<tr>
<td>Toluene</td>
<td></td>
</tr>
<tr>
<td>Trichloroethylene</td>
<td></td>
</tr>
<tr>
<td><strong>Drugs</strong></td>
<td></td>
</tr>
<tr>
<td>Bleomycin</td>
<td></td>
</tr>
<tr>
<td>Pentazocine</td>
<td></td>
</tr>
<tr>
<td>Taxol</td>
<td></td>
</tr>
<tr>
<td>Cocaine</td>
<td></td>
</tr>
<tr>
<td><strong>Dietary Supplement/Appetite Suppressants</strong></td>
<td></td>
</tr>
<tr>
<td>L-tryptophan (contamination)</td>
<td></td>
</tr>
<tr>
<td>Mazindol</td>
<td></td>
</tr>
<tr>
<td>Fenfluramine</td>
<td></td>
</tr>
<tr>
<td>Diethylpropion</td>
<td></td>
</tr>
</tbody>
</table>

MICROCHIMERISM

Healthy women harbor immunologic stem cells of fetal origin that persist for many years following pregnancy, a condition called microchimerism. Some studies found that the number of circulating fetal cells is elevated in women with SSc compared with healthy women. Moreover, cells with a male chromosome, presumably from a prior pregnancy with a male fetus, have been detected in affected organs from women with SSc. It has been speculated that persistence of fetal cells in SSc may be linked to the development of the disease through a graft-versus-host response triggered by the fetal cells or through a maternal (auto) immune response against the fetal cells.
**PATHOLOGY**

**General Features**

The pathologic hallmarks of SSc are a noninflammatory proliferative/obliterative vasculopathy affecting small arteries and arterioles in multiple vascular beds and interstitial and vascular fibrosis, most prominent in the skin, lungs, and heart.\(^9\) Inflammation is generally absent in long-standing SSc, but in early-stage disease inflammatory cell infiltrates in many organs may be prominent. In the skin, the infiltrates are located predominantly around blood vessels and in the reticular dermis and are composed primarily of CD4\(^+\) T lymphocytes and monocytes, whereas in the lungs, cellular infiltrates consist predominantly of CD8\(^+\) T lymphocytes.

**Vascular Pathology**

Vascular injury and activation are the earliest and possibly primary events in SSc. Histopathologic evidence of vascular damage is present before fibrosis and can be detected in involved and uninvolved skin, indicating a generalized process.\(^9\) Raynaud phenomenon and other vascular manifestations typically precede other manifestations of SSc. Additional signs of vasculopathy include cutaneous telangiectasia; nailfold capillary changes (giant capillaries, hemorrhages, and avascular areas); PAH; digital tip pitting; gastric antral vascular ectasia (also called watermelon stomach); and scleroderma renal crisis. The most characteristic histopathologic finding in the small and medium-sized arteries is bland intimal proliferation (Figure 83-2). Intimal hypertrophy, a finding that SSc shares with chronic allograft arteriopathy, is thought to result from proliferation and migration of myointimal cells and local accumulation of collagen.\(^20\) The vascular basement membranes are thickened and reduplicated. These changes are most prominent in blood vessels of the heart, lungs, kidneys, and gastrointestinal tract. A systematic survey of SSc skin biopsies revealed a marked reduction in the number of capillaries (rarefaction) and loss of vascular endothelial cadherin, a molecule required for vascular tube formation.\(^21\) Remarkably, this study showed that clinical improvement following high-dose immunosuppressive therapy was accompanied by capillary regeneration in the skin.

Impaired fibrinolysis, increased levels of von Willebrand factor, and ongoing platelet aggregation are prominent. Endothelial cell injury itself causes further platelet aggregation, release of platelet-derived growth factor (PDGF) and endothelin-1 (ET-1), and endothelial cell apoptosis.\(^22\) Vascular lesions and immune complex deposition in the vessel walls are uncommon. In late stages of the disease, extensive fibrin deposition and perivascular fibrosis cause progressive luminal occlusion, and eventually there is striking paucity of small blood vessels and capillaries in lesional tissue.\(^23\) Loss of vascular supply leads to chronic tissue hypoxia. Widespread proliferative/obliterative vasculopathy of small and medium-sized arteries and capillary rarefaction are the pathologic hallmarks of all forms of SSc. In patients with SSc-associated PAH, pulmonary arteriolar intimal proliferation and evidence of veno-occlusive disease are prominent, whereas in contrast to idiopathic PAH, plexogenic arteriopathy does not occur.\(^24\)

**Tissue Fibrosis**

Fibrosis is characterized by excessive accumulation of fibrillar collagens, fibronectin, elastin, proteoglycans, cartilage oligomeric matrix protein (COMP), and many other structural extracellular matrix (ECM) molecules. The process causes disruption and, ultimately, complete obliteration of tissue architecture. Most prominently affected are the lungs, gastrointestinal tract, heart, tendon sheath, and perisascular tissue surrounding skeletal muscle. Histopathologic examination of these organs shows homogeneous and relatively acellular connective tissue with thick hyalinized collagen bundles.

**ORGAN-SPECIFIC PATHOLOGIC FINDINGS**

**Skin**

Fibrosis of the skin, the hallmark of SSc, is accompanied by marked expansion of the dermis with obliteration of the hair follicles, sweat glands, and sebaceous glands and other skin appendages. Collagen fiber accumulation is most prominent in the reticular (deep) dermis and progressively involves the subjacent adipose layer with entrapment of fat cells. Early-stage SSc skin biopsies reveal dermal edema and perivascular infiltrates composed of T lymphocytes and monocytes (Figure 83-3). Less commonly, mast cells and eosinophils may also be detected.\(^25,26\) The proportion of α-smooth muscle actin–positive myofibroblasts, a mesenchymal cell type that is intermediate between fibroblasts and contractile smooth muscle cells and plays a major role in fibrogenesis, is increased in the lesional skin.\(^27\) With
CHAPTER 83 | ETIOLOGY AND PATHOGENESIS OF SCLERODERMA

CCN2, IL-13, and Wnt signaling pathways show elevated expression. Furthermore, a number of genes reflecting a bone/cartilage phenotype are elevated in the skin. A particularly intriguing finding from these studies is that skin biopsies from different individuals show marked heterogeneity at the level of their “molecular signature,” with at least five distinct and reproducible patterns identified to date.

Lungs

In early-stage SSc, the lungs show patchy infiltration of the alveolar walls with lymphocytes, plasma cells, macrophages, and eosinophils (Figure 83-4). At this stage, bronchoalveolar lavage fluid contains elevated proportions of inflammatory leukocytes. With progression, interstitial lung fibrosis...
and vascular damage predominate, often coexisting within the same lesions. Intimal thickening of the pulmonary arteries, best seen with elastin stain, underlies PAH and at autopsy is often associated with multiple pulmonary emboli and myocardial fibrosis.

Lung fibrosis in SSc is characterized by expansion of the alveolar interstitium due to accumulation of collagen and other connective tissue proteins. The typical histologic pattern is nonspecific interstitial pneumonitis (NSIP), a form of interstitial lung disease characterized by mild to moderate interstitial inflammation, type II pneumocyte hyperplasia, and fairly uniform distribution of fibrosis. Less commonly, SSC is associated with the usual interstitial pneumonia (UIP) pattern that is characterized by scattered fibroelastic foci and patchy distribution of fibrosis and has a worse prognosis. Progressive alveolar septal thickening ultimately results in obliteration of the air spaces and honeycombing, as well as consequent loss of pulmonary blood vessels. This process impairs gas exchange and contributes to worsening of pulmonary hypertension. Extensive pulmonary fibrosis may also predispose to primary lung carcinoma.

Gastrointestinal Tract

Prominent pathologic changes can occur at any level from the mouth to the rectum. The esophagus is virtually always affected, with fibrosis in the lamina propria, submucosa, and muscular layers and characteristic vascular lesions. Replacement of the normal intestinal architecture results in disordered peristaltic activity, gastroesophageal reflux and small bowel dysmotility, pseudo-obstruction, and bacterial overgrowth. Chronic gastroesophageal reflux is complicated by esophageal inflammation, ulcerations, and stricture formation. Up to one-third of SSc patients with severe gastroesophageal reflux develop Barrett’s esophagus, characterized by metaplasia of the normal squamous lining of the esophagus into columnar epithelium. Because Barrett’s metaplasia is a premalignant lesion associated with a greater than 30-fold increased risk of adenocarcinoma, patients with Barrett’s esophagus need ongoing monitoring for dysplasia and adenocarcinoma.

Kidneys

In the kidneys vascular lesions predominate, and glomerulonephritis is rare except in overlap syndromes. Chronic renal ischemia is associated with shrunken glomeruli and other ischemic changes. Patients with acute scleroderma renal crisis show a thrombotic microangiopathy that is indistinguishable from other forms of malignant hypertension. Histopathologic changes are most prominent in the small interlobular and arcuate renal arteries, which show reduplication of elastic lamina, marked intimal proliferation (onion skinning), and fibrinoid necrosis of the arteriolar walls. Similar pathologic changes have also been reported in SSc patients who do not have renal crisis. Intimal thickening leads to severe narrowing and total obliteration of the lumen, often with microangiopathic hemolysis. Tubular changes such as flattening and degeneration of tubular cells occur secondary to vascular insufficiency. The clinical picture of scleroderma renal crisis may resemble thrombotic thrombocytopenic purpura (TTP). However, reduced to absent levels or activity of von Willebrand factor cleaving protease (ADAMTS13) has not been reported in scleroderma renal crisis. Figure 83-5 shows the characteristic histologic features associated with scleroderma renal crisis.

Heart

At autopsy, evidence of cardiac involvement is found in up to 80% of patients with SSc. Modest pericardial effusions are common; occasionally fibrosis and constrictive pericarditis may occur. A characteristic pathologic finding is myocardial contraction band necrosis, which is thought to reflect repeated episodes of ischemia-reperfusion due to “myocardial Raynaud phenomenon.” Significant interstitial and perivascular fibrosis in the heart may occur in the absence of clinically evident heart involvement. Skeletal muscle myositis in SSc is occasionally accompanied by acute myocarditis.

PATHOLOGIC FINDINGS IN OTHER ORGANS

Fibrosis of the thyroid glands is common. Broad bands of fibrous tissue are seen in the thyroid gland, with atrophy and obliteration of the follicles, in the absence of inflammation. Abnormal thyroid function tests and antithyroid antibodies are common. Erectile dysfunction is frequent and may be a presenting manifestation of the disease in men. Pathologic examination shows extensive proliferative/obliterative changes in the penile blood vessels. Fibrosis of the salivary and lacrimal glands in the absence of inflammation can occur and may be associated with Sjögren’s syndrome. Synovial biopsies show fibrosis and characteristic vascular changes in the small arterioles.

ANIMAL MODELS OF SCLERODERMA

Animal models of human disease are indispensable research tools to facilitate the understanding of complex disease states. Such models are used to identify the cellular and molecular components of pathologic process, discover potential therapeutic targets, and develop and evaluate novel treatment strategies. A large number of putative SSc models have been reported, but to date none fully reproduce each of the four cardinal features of the disease: obliterative and proliferative microangiopathy, autoimmunity, inflammation, and fibrosis. However, as illustrated in Figure 83-6, particular animal models recapitulate selected disease features. Broadly speaking, current mouse models can be divided into four types: (1) naturally occurring disease models, in which spontaneous mutations are associated with a genetically transmitted scleroderma-like phenotype such as tight skin (Tsk1/+ mouse); (2) induced models in which the scleroderma phenotype is elicited by chemical exposures or by manipulation of the immune system (bleomycin-induced skin and lung fibrosis); (3) transplantion of HLA-mismatched bone marrow cells resulting in chronic sclerodermatous graft-versus-host disease; and (4) genetic manipulations giving rise to engineered mouse strains with heritable scleroderma-like traits.
Figure 83-5  Histologic appearance of scleroderma renal crisis. Characteristic histologic findings include interstitial fibrosis (A) and occlusion of intrarenal arteries with neointima formation, fibrinoid necrosis of the vessel wall, and reduplication of the internal elastic lamina (B). The glomeruli are shrunken and lack inflammatory cells or proliferative changes (C). Intravascular thrombosis resembling the changes of thrombotic thrombocytopenic pupura may be present (D).

Figure 83-6  Animal models of systemic sclerosis. Mouse models recapitulate selected pathophysiologic features of systemic sclerosis (SSc). Shown are the principal disease processes of SSc (inflammation, autoimmunity, microangiopathy, and fibrosis) and the extent to which they are represented in commonly used mouse models. CTGF, connective tissue growth factor; GVHD, graft-versus-host disease; tg, transgenic. (Modified from Distler JH, Distler O, Beyer C, Schett G: Animal models of systemic sclerosis: prospects and limitations, Arthritis Rheum 62(10):2831–2844, 2010.)
Scleroderma, Inflammatory Myopathies, and Overlap Syndromes

Heritable Animal Models of Scleroderma

The Tsk1/+ mouse is characterized by diffuse skin thickening and tethering to the underlying subcutaneous tissue. Mice homozygous for the Tsk1 mutation die in utero at 8 to 10 days of gestation. However, heterozygous mice (Tsk1+/+) survive and develop tight skin. In contrast to SSc, Tsk1/+ mice show prominent subcutaneous hyperplasia but relatively unremarkable dermis. Furthermore, the lungs of Tsk1/+ mice show emphysematous changes rather than fibrosis, and microangiopathy has not been reported. Although inflammation is uncommon, Tsk1/+ mice develop autoantibodies directed against topoisomerase-I. The Tsk1 mutation is now known to be an intragenic tandem duplication fibrillin-1 gene. Fibrillin-1 is a large ECM protein that is widely distributed in microfibrils, and in addition to its structural role also modulates the latency and activation of TGF-β. Fibrillin-1 gene mutations are associated with Marfan’s disease with multiple tissues showing activation of TGF-β. The fibrillin-1 duplication mutation in Tsk1/+ mice gives rise to an abnormally large 450-kD protein. No corresponding fibrillin-1 mutations have been demonstrated in patients with SSc. Although it has been hypothesized that accumulation of the abnormally large mutant fibrillin-1 destabilizes the matrix or perturbs the homeostatic control of TGF-β latency, the precise mechanisms linking the Tsk1/+ mutation in fibrillin-1 to the development of cutaneous hyperplasia are unknown.

Another animal model of scleroderma is the Tsk2 mouse that spontaneously develops scleroderma-like skin changes by 3 to 4 weeks of age. In contrast to Tsk1/+ mice, Tsk2/+ mice have fibrotic dermis with extensive infiltration with mononuclear inflammatory cells and show evidence of autoimmunity. The Tsk2 mutation, originally induced in normal mice by exposure to ethylnitrosourea, is located on chromosome 1 and is inherited as an autosomal dominant trait, although the underlying molecular defect has not yet been identified.

Inducible Animal Models of Scleroderma

Chronic skin and lung fibrosis (Figure 83-7) can be induced in BALB/c or C57 mice by subcutaneous injections of bleomycin or oxidative agents. Bleomycin is an antitumor

---

Figure 83-7  Bleomycin-induced mouse model of systemic sclerosis. Mice received subcutaneous injection of PBS (A1, A3) or bleomycin (A2, A4) for 28 days. A1-A4, Skin changes. A1 and A2, H&E stain; A3 and A4, Sirius red stain. Note dermal fibrosis and loss of subcutaneous adipose layer (A2), and increased collagen accumulation (A4).
chemotherapy drug that has long been recognized to be complicated by pulmonary fibrosis. The sequence of bleomycin-induced histopathologic changes in mice closely resembles those seen in SSc: early and self-limited mononuclear cell infiltration and increased expression of cytokines such as TGF-β, IL-4, IL-6, IL-13, and monocyte chemotactic protein-1 (MCP-1), followed by the appearance of dermal fibrosis with excessive collagen deposition and accumulation of α–smooth muscle actin–positive myofibroblasts. Bleomycin-induced fibrosis may be linked to reactive oxygen species (ROS) generation, as well as direct activation of innate immunity via TLR2. In contrast to SSc, bleomycin-induced mouse scleroderma is not associated with either microangiopathic changes or autoantibodies and skin fibrosis is limited in extent and duration. Nevertheless, in light of its reproducibility, relative strain independence, and ease of induction, this mouse model is widely used for investigating specific molecules and
pathways in the development and treatment of fibrosis. Injection of oxidative agents (such as hydroxyl radicals or hypochlorite) into the skin in BALB/c mice induces skin and lung fibrosis, as well as the appearance of SSc-specific autoantibodies. The pathologic changes are linked to the generation of hydrogen peroxide and other ROS. Subcutaneous injection of TGF-β induces granulation tissue and only transient fibrosis, but simultaneous injection of CTGF along with TGF-β results in persistent fibrosis, suggesting that CTGF is required for sustaining the fibrotic response. However, adenovirus-mediated delivery of constitutively active TGF-β receptor I is sufficient to induce local dermal fibrosis. Transplantation of bone marrow or spleen cells into sublethally irradiated minor histocompatibility locus-mismatched recipient mice results in sclerodermatous graft-versus-host disease. This mouse model displays interstitial and perivascular fibrosis in the skin and lung and autoimmunity. Skin fibrosis is preceded by mononuclear cell infiltration with elevated TGF-β and chemokine expression.

### Genetic Manipulations in Mice Giving Rise to Scleroderma-like Phenotypes

Mouse strains with genetic gain-of-function or loss-of-function modifications resulting in spontaneous sclerodermalike phenotypes have been created. These transgenic, knockin, and knockout mice provide robust novel experimental tools in scleroderma research. Mouse strains with constitutive or inducible expression of TGF-β signaling in fibroblasts recapitulate key clinical, histologic, and biochemical features of SSc and provide support for the role of perturbed TGF-β signaling in pathogenesis. Other promising transgenic models of SSc include mice overexpressing CTGF, PDGF receptor (PDGFR)-α, Wnt10b, and Fra-2 and mice null for caveolin-1, relaxin, Fli-1, and fetuin A. Sclerodema-like fibrotic, vascular, and calcific changes in the skin, as well as in some cases the lungs, develop spontaneously in these mice. Other genetically engineered mice show increased sensitivity to the induction of fibrosis. Examples include T-bet null and VEGF null mice. These emerging mouse models provide new exciting research opportunities for the discovery of molecules and pathways underlying the pathogenetic features of SSc.

### PATHOGENESIS

#### Integrated Overview

The pathogenesis of SSc is complex, and existing animal models capture only some of its diverse pathologic and clinical attributes. An integrated view of pathogenesis must integrate the cardinal features of SSc: vascular injury and damage, inflammation with activation of the innate and adaptive arms of the immune system, and fibroblast activation resulting in generalized interstitial and vascular fibrosis. Although evidence for each can be found in each SSc patient, the relative individual contribution of these processes to the clinical manifestations varies from one patient to another. As illustrated in Figure 83-1, complex and dynamic interplay between these distinct processes is thought to be responsible for initiating, amplifying, and sustaining tissue damage in SSc.

### Vasculopathy

Vascular injury is likely to be the initiating and proximal event in SSc (see Figure 83-5). Evidence of vascular involvement is early and widespread, and its progression over time is associated with significant clinical sequelae. The presence of nailfold microvascular changes, detected by capillaroscopy, in an individual with isolated Raynaud phenomenon identifies an elevated risk of progression to SSc, indicating that microvasculopathy precedes other clinical manifestations of the disease.

#### Vascular Injury and the Activated Endothelium

The initial vascular insult is triggered by circulating factors such as (unidentified) cytotoxic molecules or T cell–derived proteolytic granzymes. Other potential causes of vascular injury include antiendothelial cell autoantibodies, vasculotropic viruses, inflammatory cytokines, ROS, and other forms of environmental stress. Vascular injury causes endothelial cell activation, with increased expression of vascular endothelial cell adhesion molecule-1 and E-selectin, altered secretion of vasoactive mediators, platelet activation, and activation of the thrombotic and fibrinolytic cascades. In the injured arterioles and capillaries, activated endothelial cells may undergo transdifferentiation to mesenchymal cells via a process termed endothelial–mesenchymal transition. This process driven by TGF-β and Notch is associated with loss of endothelial markers such as CD31 and progressive acquisition of mesenchymal markers such as α-smooth muscle actin. Although endothelial–mesenchymal transition has been documented in cancer, recent studies identify endothelial cell-derived fibroblasts in cardiac and pulmonary fibrosis, suggesting that this form of endothelial cell plasticity may also play a role in SSc. Platelet activation is a prominent early feature of vascular injury in SSc and is associated with the release of thromboxane A₂, PDGF, and TGF-β, which potentiate vasoconstriction and also contribute to fibroblast activation and myofibroblast transdifferentiation. Pericytes, which are smooth muscle–like structural cells found in the walls of small blood vessels, show marked hyperplasia in lesional skin from patients with early-stage SSc and express the surface marker Thy-1 (CD90) and receptors for PDGF. Functional abnormalities of the vascular endothelium include impaired production of and responsiveness to endothelium-derived vasodilatory factors such as nitric oxide (NO), thrombomodulin, calcitonin gene-related peptide, and prostacyclins. The ensuing imbalance of vaso-dilators and vasoconstrictors impairs blood flow responses. Recurrent episodes of ischemia–reperfusion create oxidative stress with the generation of H₂O₂ and other ROS. Damaged microvessels show increased vascular permeability and enhanced transendothelial leukocyte migration. Platelets are exposed to subendothelial structures, which further aggravates platelet aggregation. Activated endothelial cells release the extremely potent vasoconstrictor ET-1, which promotes leukocyte adhesion, vascular smooth muscle cell proliferation, and fibroblast activation (Figure 83-8). Levels of ET-1 are elevated in the blood and in bronchoalveolar lavage fluids from patients with SSc. Ultimately
vasculogenesis as a potential explanation. A reduction in the number of bone marrow-derived circulating endothelial progenitor cells and their impaired differentiation into mature endothelial cells has been described in SSc patients.67-69 Because CD34+ circulating endothelial progenitor cells are required for physiologic vasculogenesis in ischemic tissues, defective progenitor cell mobilization or function compromises the vascular repair. Whether the reduction in circulating endothelial progenitor cells in SSc is due to “exhaustion” of the bone marrow, destruction in the peripheral circulation or some other process, remains unresolved.

Hypoxia

The widespread microangiopathy and resultant capillary loss in affected tissues leads to decreased blood flow and consequent hypoxia. With the onset of fibrosis, excessive ECM accumulation increases diffusion distance from blood vessels to cells, further aggravating tissue hypoxia.70 Hypoxia is a potent inducer of the basic helix-loop-helix transcription factors HIF-1α and HIF-1β. Under normoxic conditions, cellular HIF-1α is undetectable due to its rapid and efficient proteasomal degradation mediated by the tumor suppressor von Hippel-Lindau (vHL) protein. Because in hypoxic cells pVHL is unable to bind to its target, HIF-1α fibrinolytic and coagulation cascades are activated, culminating in intravascular fibrin deposition and thrombosis.

**Vascular Damage and Vasculogenesis**

In the medium-sized and larger vessels a combination of hypertrophy of the intimal and medial layers and adventitial fibrosis causes progressive luminal narrowing. Together with endothelial cell apoptosis, the process culminates in oblitative vasculopathy and vascular rarefaction, with the characteristic striking decrease of blood vessels seen angiographically in late-stage SSc. Loss of microvasculature results in reduced blood flow and chronic tissue hypoxia, which induces hypoxia-inducible factor-1 (HIF-1) α-dependent genes such as VEGF and its receptors. There is substantial controversy regarding the balance of proangiogenic and antiangiogenic factors in SSc. In some studies, the plasma levels of the angiogenesis inhibitor endostatin, a degradation product of type XVIII collagen, were reported to be increased.64 Other studies, however, found elevated levels of the proangiogenic factors VEGF, fibroblast growth factor, and PDGF. Furthermore, the expression of VEGF and its receptors was also elevated in lesional tissue.65,66 It is currently unclear why, in the face of tissue hypoxia and elevated angiogenic factors, SSc is associated with progressive loss of blood vessels. Recent studies implicate defective vasculogenesis as a potential explanation. A reduction in the number of bone marrow-derived circulating endothelial progenitor cells and their impaired differentiation into mature endothelial cells has been described in SSc patients.67-69 Because CD34+ circulating endothelial progenitor cells are required for physiologic vasculogenesis in ischemic tissues, defective progenitor cell mobilization or function compromises the vascular repair. Whether the reduction in circulating endothelial progenitor cells in SSc is due to “exhaustion” of the bone marrow, destruction in the peripheral circulation or some other process, remains unresolved.

**Figure 83-8** Pathogenesis of vasculopathy. Initial vascular insult results in endothelial cell (EC) injury and activation, with reversible functional changes, increased expression of adhesion molecules, and enhanced leukocyte diapedesis resulting in perivascular inflammation. Damaged endothelium promotes platelet aggregation and thrombin release and shows impaired production of vasodilators such as nitric oxide (NO), increased production of vasoconstrictors such as endothelin-1 (ET-1), and release of reactive oxygen species (ROS). Consequent vasoconstriction and defective vasodilation aggravate vascular damage, leading to irreversible and progressive vascular wall remodeling, luminal occlusion, platelet aggregation, in situ thrombosis, and tissue hypoxia. Loss of blood vessels may be further compounded by insufficient vasculogenesis. ECP, endothelial cell progenitor; PGI, prostaglandin I; SMC, smooth muscle cell.
is protected from degradation and translocates into the nucleus, binds to hypoxia-responsive DNA regulatory sequences, and induces the transcription of hypoxia-regulated genes involved in erythropoiesis, angiogenesis, and glucose metabolism. Hypoxia is also a potent in vitro and in vivo stimulus for ECM remodeling genes such as collagens, prolyl hydroxylases, and lysyl oxidase and promotes epithelial cell differentiation into activated myofibroblasts. These and other hypoxia-induced profibrogenic responses are partially mediated by autocrine TGF-β. A recent paper provides compelling experimental support for the role of hypoxia in exacerbating fibrosis. In this study, mice with myeloid-cell-specific deletion of VEGF showed markedly increased tissue hypoxia and activated Wnt-β-catenin signaling upon bleomycin-induced lung injury, culminating in striking exacerbation of fibrosis.

The presence of severe tissue hypoxia in SSC has been documented noninvasively, and oxygen levels in the skin were shown to inversely correlate with skin thickness. In another study using a needle electrode inserted into the dermis, SSC patients exhibited significantly reduced levels of PO₂ (23.7 mm Hg vs. 33.6 mm Hg in controls). Interestingly, HIF-1α in the lesional skin was undetectable, rather than elevated, as is the case with the hypoxia-inducible VEGF.

**Oxidative Stress and Reactive Oxygen Species**

Oxidative stress and elevated levels of ROS are implicated in the pathogenesis of SSC. ROS might be generated from the endothelium in response to repeated episodes of ischemia reperfusion. Normal fibroblasts exposed to TGF-β or to stimulatory PDGFBR autoantibodies generate ROS, whereas explanted SSC skin fibroblasts spontaneously produce excessive amounts of ROS via the membrane NADPH oxidase Nox4. Hydrogen peroxide and other oxygen free radicals in turn stimulate collagen synthesis, TGF-β secretion, and other fibrotic responses in fibroblasts. A role for ROS in pathogenesis is further supported by a mouse model of scleroderma induced by subcutaneous injections of peroxynitrite or hypochlorite.

**IMMUNE DYSREGULATION**

**Introduction**

A resurgence of interest in understanding the role of immune dysregulation in the pathogenesis of SSC is occurring. Both the innate and adaptive arms of the immune system are activated in early SSC, and autoimmunity manifested by highly specific and mutually exclusive autoantibodies is prominent throughout the course of the disease. Lesional tissues show prominent perivascular accumulation of mononuclear inflammatory cells, whereas circulating leukocytes (including T cells, B cells, monocytes, and dendritic cells) show evidence of activation and polarization and display the “type I interferon signature” associated with SLE and other autoimmune diseases (Figure 83-9). The strong genetic associations between SSC and HLA Class II locus polymorphisms provide support for the immunologic basis of SSC. Nevertheless, whether immune dysregulation is a primary or significant factor in SSC pathogenesis still remains to be established. Current clinical trials are evaluating the efficacy of myeloablative therapy followed by hematopoietic reconstitution as a therapeutic strategy to “reset the immune system” in SSC.

**Cellular Effectors of Immune Dysregulation in Scleroderma: T Cells, B Cells, and Monocytes/Macrophages**

**T Cell Activation**

Activation of T cells is evident in lesional tissues, as well as in peripheral blood, and appears to play a direct role in tissue injury. In early SSC, activated CD4 and CD8 T lymphocytes and monocytes/macrophages, and less commonly B cells, eosinophils, mast cells and NK cells, are observed...
in perivascular regions in the lesional skin, lungs, and other affected organs; these inflammatory cell infiltrates are detectable before the appearance of fibrosis. In situ hybridization studies show prominent procollagen gene expression in early-stage SSc skin in fibroblasts that are adjacent to inflammatory cells, suggesting a role for the inflammatory cells or their soluble products in inducing fibroblast activation. The extent of lymphocytic tissue infiltration correlates with the severity and progression of skin fibrosis.

Tissue-infiltrating T cells are predominantly CD$^+$ and CD$^+$ express activation markers CD69, CD45, HLA-DR, and the IL-2 receptor. These T cells display restricted T cell receptor signatures indicative of oligoclonal T cell expansion in response to an as-yet unidentified antigen. In the lungs, a predominance of CD$^+$ and $\gamma\delta$ T cells is observed. It is not known whether T cells in lesional tissue are activated nonspecifically (by cytokines or chemokines) or specifically in response to an antigen. Evidence of T cell activation in SSc is also detected in the peripheral blood, with elevated serum levels of IL-2 and IL-2 receptor expression, and spontaneous cytokine secretion. Peripheral blood lymphocytes show the "IFN signature" defined by the upregulation of type I IFN-inducible genes. Increased expression of LFA-1 and $\alpha1$ integrins may enable T cells to adhere directly to endothelium and fibroblasts. Circulating CD$^+$ T cells also express elevated chemokine receptors. At the same time, vascular endothelial cells express ICAM-1, E-selectin, and other adhesion molecules that facilitate leukocyte diapedesis.

**Th1/Th2 Cytokine Balance and Polarized Immune Responses**

Evidence implicates an altered balance between T helper (Th)1 and Th2 cytokines in fibrosis. T cells polarized to a Th2 pattern secrete abundant IL-4, IL-5, and IL-13, with only low levels of the hallmark Th1 cytokine IFN-$\gamma$. The Th2 cytokines are considered to be profibrogenic because they can directly stimulate collagen synthesis and myofibroblast transdifferentiation, as well as induce TGF-$\beta$, a powerful modulator of immune regulation and ECM accumulation. In contrast, IFN-$\gamma$ blocks these responses and exerts antifibrotic effects. Thus skewing of the immune response toward a Th2 pattern could create a profibrotic environment.

Animal studies support the significance of a Th2-polarized immune response in fibrosis. For example, T cells that have been Th2 polarized in vitro induce fibrosis when passively transferred in mice. Moreover, mice lacking the transcription factor T-bet, which directs T cell differentiation of T cells toward a Th1-predominant phenotype, spontaneously show a Th2-polarized immune response and develop exaggerated skin fibrosis in response to bleomycin. Patients with SSc display an altered Th1/Th2 cytokine balance with a predominant Th2 profile. For example, both the serum levels of IFN-$\gamma$ and in vitro IFN-$\gamma$ production by peripheral blood monocytes are reduced. Microarray analysis of SSc peripheral blood leukocytes showed elevated expression of GATA3, a transcription factor controlling Th2 polarization. Cloned CD$^+$ T cells from SSc skin biopsies secrete IL-4 but not IFN-$\gamma$. T cell lines generated from SSc skin biopsies show prominent CD8 expression and produce high levels of IL-4. Similarly, alveolar CD8$^+$ T cells show elevated Th2 cytokine production, and the Th2 predominance predicts accelerated decline in lung function. By analyzing the ratio of chemokine receptors associated with Th1 versus Th2 responses, peripheral blood T cells in SSc were shown to have a Th2-predominant pattern that predicted the presence and progression of interstitial lung disease. Proteomic and DNA microarray analysis demonstrated a Th2 cytokine predominance in SSc bronchoalveolar lavage fluids and CD8$^+$ lymphocytes. A longitudinal study of SSc patients showed that skin improvement over time was associated with a decline in serum levels of Th2 cytokines and a concomitant increase in IL-12, a Th1 cytokine.

**Other T Cell Subsets**

The Th17 subset of T lymphocytes is implicated in the pathogenesis of rheumatoid arthritis and other inflammatory diseases. SSc patients have increased numbers of IL-17 producing CD4$^+$ T cells in both the peripheral blood and bronchoalveolar lavage fluid. Patterns of Th1, Th2, and Th17 predominant immune responses might be useful in identifying specific clinical phenotypes in SSc.

Suppressor regulatory T cells (Tregs) are important in controlling autoimmunity. Interestingly, although the frequency of peripheral blood Treg cells was increased in SSc patients, they showed impaired suppressive activity and secretion of TGF-$\beta$ and IL-10. Other studies failed to replicate these findings but demonstrated reduced numbers of Foxp3$^+$ CD4$^+$ T cells in the lesional skin compared with other inflammatory skin diseases.

**Monocytes and Macrophages**

Phagocytic monocytes and macrophages regulate innate immunity and tissue repair. Monocytes are a major source of cytokines and chemokines including IL-1, tumor necrosis factor (TNF), MCP-1, PDGF, and TGF-$\beta$, all of which are important in inflammatory and fibroproliferative responses. In addition, monocytes produce collagenases and other matrix-degrading enzymes that mediate tissue remodeling. Macrophages are prominent among the mononuclear cells infiltrating the lesional skin in early SSc and express Siglec-1 and AIF-1, markers of IFN-induced activation. Macrophages exhibit discrete phenotypes in response to specific stimuli, with alternatively activated (M2) macrophages induced in response to IL-4 and TGF-$\beta$ secreting profibrotic mediators. M2 macrophages accumulate in various organs in animal models of fibrosis, and may be important in pathogenesis. In SSc patients with active lung disease, alveolar macrophages show an M2 phenotype characterized by secretion of profibrotic mediators TGF-$\beta$, PDGF, and IL-13. Evidence of local mast cell and eosinophil activation and degranulation is sometimes prominent in lesional skin.

**Dendritic Cells**

These potent antigen-presenting cells found in the skin and circulation bridge innate and adaptive immunity, and by modulating T cell, B cell, and monocyte/macrophage function shape the immune response (see Figure 83-9).
Dendritic cells are the major source of type I IFN and can also secrete fibrogenic cytokines including TGF-β. Recent studies show that CD11c⁺ dendritic cells accumulate in fibrotic tissue in animal models and in patients with SSc. Activated dendritic cells in the lesional tissue might directly or indirectly modulate resident fibroblast activation. Moreover, monocyte-derived dendritic cells from the peripheral blood in SSc show altered in vitro responses to TLR2/3 ligands and increased secretion of IL-6 and IL-10, which favors a Th2 skewing of the immune response.

**AUTOIMMUNITY**

**Autoantibodies in Scleroderma: Pathogenetic Considerations**

Humoral autoimmunity, manifested by the presence of autoantibodies with multiple antigenic specificities, can be detected in the serum in virtually all patients. The presence of SSc autoantibodies has significant utility in diagnosis and classification, as well as in predicting organ-specific complications and clinical course. In contrast, a direct role of autoantibodies in tissue damage has not been conclusively established. Autoantibodies in SSc tend to be highly specific and mutually exclusive, and they show strong association with individual disease phenotypes and immunogenetic backgrounds. The levels of autoantibodies, particularly antitopoiso- merase I, may correlate with the extent of skin and lung fibrosis and fluctuate with disease activity.

Various hypotheses exist to explain the generation of autoantibodies in SSc. According to one hypothesis, self-antigens such as topoiso- merase I undergo proteolytic cleavage in the presence of ROS, resulting in exposure of normally cryptic epitopes and break in immune tolerance. Other potential mechanisms include molecular mimicry as a consequence of viral infection, chronic B cell hyperreactivity, and increased expression or altered subcellular localization of potential autoantigenic peptides. Alternately, SSc autoantibodies might be generated from B cells activated by endogenous TLR7 ligands such as nucleic acid–containing cellular debris.

Several reports describe biologically active autoantibodies directed against ECM components, cell membrane PDGF receptors, fibroblasts, and endothelial cells in SSc patients. Stimulatory autoantibodies can induce target cell activation or apoptosis in vitro. In one study, autoantibodies against the PDGF receptor were detected in each of the SSc patients evaluated in one study and shown to induce ROS generation and multiple signaling cascades in normal skin fibroblasts, resulting in myofibroblast differentiation. It is not yet known whether these autoantibodies precede, or are a consequence of, fibrosis, and their direct pathogenetic role in SSc remains to be established.

**B Cells in Scleroderma**

Recent studies provide evidence for a potential direct role for B lymphocytes in the pathogenesis of SSc. B cells have multiple immune regulatory functions in addition to the generation of antibodies including antigen presentation, cytokine production, lymphoid organogenesis, and T cell differentiation. Although B cells are not generally prominent in microscopic sections of lesional tissue, DNA microarray analyses demonstrated a molecular signature indicative of activated B cells with increased expression of immunoglobulin genes in some SSc skin biopsies. B cells were also prominent in SSc lung biopsies. Patients with SSc display intrinsic abnormalities of B cells, with elevated numbers of naïve B cells, and reduced numbers of plasma cells. Memory B cells are chronically activated and display increased CD95 and CD86, as well as CD19, a signaling cell surface receptor that regulates intrinsic and antigen receptor–induced B cell responses. Transgenic mice overexpressing CD19 develop spontaneous autoimmunity and high titers of antitopoiso- merase I antibodies. Altered B cell function and chronic B cell activation in SSc may not only account for autoantibody production but also contribute to fibrosis because activated B cells secrete IL-6, which directly stimulates fibroblast activation and collagen synthesis. Patients with SSc also have elevated levels of the potent B cell survival factor B cell activating factor belonging to the TNF family (BAFF) in the serum and in lesional skin, as well as increased BAFF receptor expression on B cells. In light of the emerging role of B cells in the pathogenesis of SSc, therapies targeting B cells are under consideration.

**Type I Interferon Signature and Innate Immune Signaling: Similarities to Systemic Lupus Erythematosus**

Stimulation of dendritic cells and other cell types by ligands of TLR3 results in robust secretion of type I IFN. Accordingly, the presence of IFN or IFN-induced cellular responses suggests TLR-mediated innate immune signaling. Type I IFNs (α and β) are themselves potent modulators of innate immunity. Elevated expression of IFN-regulated genes (the “IFN signature”) was first described in SLE patients. Circulating immune complexes containing antinucleic acid autoantibodies serve as endogenous TLR3 ligands that stimulate type I IFN secretion by dendritic cells and macrophages. Similar to SLE, an IFN signature has been detected in circulating leukocytes from SSc patients. Indeed, a recent study reported that the most highly expressed genes in SSc peripheral blood leukocytes were IFN regulated and similar to those detected in SLE. Moreover, incubation of normal leukocytes with SSc sera induced the secretion of IFN, indicative of TLR3 activation by nucleic acid–containing immune complexes in the serum. Interestingly, serum levels of IFN are only inconsistently elevated in SSc. These observations identify an important role for TLR-mediated innate immune signaling and type I IFNs in SSc and highlight immunopathogenic similarities between SLE and SSc, but they fail to explain why these two related autoimmune diseases display largely nonoverlapping clinical phenotypes. The direct pathogenetic role of type I IFN in the development of microangiopathy and fibrosis in SSc remains to be established.

**FIBROSIS**

**Overview**

Fibrosis, characterized by replacement of normal tissue architecture with stiff paucicellular connective tissue, is
Chapter 83

Etiology and Pathogenesis of Scleroderma

Regulation of Collagen Synthesis

The family of collagens is composed of more than two dozen structural proteins with critical roles in organ development, growth, and differentiation. The ECM of the skin, bones, and tendons is composed largely of type I collagen, with smaller amounts of associated type III collagen. Type II collagen is found mainly in articular cartilage. The fibrillar collagens consist of three \( \alpha \) chains wound into a characteristic triple helix, a structure made possible by the presence of a glycine at every third residue of repeating Gly-X-Y sequence, where X is frequently a proline and Y is frequently a hydroxyproline. During their biosynthesis, fibrillar collagens undergo extensive enzymatic modifications inside the cell and additional processing following their secretion. Covalent cross-linking stabilizes the collagen fiber network in the extracellular space.

Extracellular Matrix

The ECM consists of a cellular compartment of resident fibroblasts and myofibroblasts and infiltrating cells, as well as connective tissue composed of large structural proteins such as collagens, proteoglycans, fibrillins, and adhesion molecules. The ECM also serves as a reservoir for sequestering growth factors and matricellular proteins that, together with the connective tissue compartment, control the differentiation, function, and survival of mesenchymal cells. Excessive connective tissue accumulation results from overproduction by fibroblasts activated by soluble factors, hypoxia, and ROS; signals from the surrounding ECM; or via cell-cell interactions. Impaired matrix degradation and turnover, as well as expansion of the pool of ECM-producing mesenchymal cells, also play roles.

The pathologic hallmark of SSc and can be viewed as a form of aberrant wound healing. Fibrosis represents the end result of a complex series of vascular and immune-mediated responses to injury in a genetically predisposed individual. As illustrated in Figure 83-10, injured or activated vascular and immune cells produce soluble mediators, autoantibodies, and ROS that induce the activation and differentiation of mesenchymal effector cells, culminating in excessive and ultimately irreversible ECM accumulation and remodeling.

**Figure 83-10** Pathogenesis of fibrosis. Fibrosis is the end result of immune dysregulation and vascular damage and hypoxia. Injury causes vascular damage, perivascular inflammation, and innate immune signaling, with oxidative stress, secretion of inflammatory and fibrogenic cytokines and chemokines, autoantibodies, fibroblast activation, and myofibroblast accumulation. Circulating mesenchymal progenitor cells traffic to and accumulate within the lesional tissue and transdifferentiate into fibrotic fibroblasts, accelerating matrix accumulation. Tissue hypoxia, matrix remodeling, and stiffness further contribute to fibroblast activation, which causes disruption of tissue architecture and organ failure. DCs, dendritic cells; ECM, extracellular matrix; HIF-1, hypoxia-inducible factor-1; IFN, interferon; PPAR-\( \gamma \), peroxisome proliferator-activated receptor-\( \gamma \); ROS, reactive oxygen species; TGF-\( \beta \), transforming growth factor-\( \beta \); TLRs, Toll-like receptors.

Regulation of Collagen Synthesis

The family of collagens is composed of more than two dozen structural proteins with critical roles in organ development, growth, and differentiation. The ECM of the skin, bones, and tendons is composed largely of type I collagen, with smaller amounts of associated type III collagen. Type II collagen is found mainly in articular cartilage. The fibrillar collagens consist of three \( \alpha \) chains wound into a characteristic triple helix, a structure made possible by the presence of a glycine at every third residue of repeating Gly-X-Y sequence, where X is frequently a proline and Y is frequently a hydroxyproline. During their biosynthesis, fibrillar collagens undergo extensive enzymatic modifications inside the cell and additional processing following their secretion. Covalent cross-linking stabilizes the collagen fiber network in the extracellular space.

Type I collagen synthesis is regulated by cytokines and other soluble extracellular factors, ROS, hypoxia, and cell-cell and cell-matrix contact (Table 83-3). Environmental cues allow fibroblasts to respond to dynamic tissue requirements during development and tissue repair. The genes encoding the various collagens harbor cis-acting regulatory elements that are specifically recognized by DNA-binding transcription factors. Sp1, Ets1, Smad2/3, Egr-1, and CCAAT-binding factor (CBF) stimulate, and Sp3, C/EBP, YB1, c-Krox, and Fli-1 suppress transcription. These
transcription factors interact with one another and with non-DNA-binding cofactors, scaffold proteins, and chromatin-modifying enzymes such as p300/CBP, PCAF, and histone deacetylases. The activities and interactions of transcription factors and cofactors are controlled by extracellular cues. Enzymes that modify chromatin structure, causing it to unwind, enhance DNA-binding factor access to cognate cis-acting regulatory sequences and facilitate transcription. Alterations in the levels, activities, or interactions among the various transcription factors, co-factors, and chromatin-modifying enzymes contribute to persistent fibroblast activation in SSc.

**Effector Cells of Fibrosis: Fibroblasts**

Fibroblasts are versatile spindle-shaped cells that are capable of both synthesis and degradation of ECM. Although unstimulated fibroblasts are biosynthetically quiescent, under the influence of appropriate extracellular signals they secrete ECM macromolecules and growth factors, cytokines and growth factors; adhere to and contract connective tissue; and transdifferentiate into myofibroblasts. Together, these biosynthetic, proinflammatory, contractile, and adhesive functions enable fibroblasts to execute effective wound healing. Whereas under physiologic conditions the fibroblast repair program is self-limited, pathologic fibrosis is characterized by uncontrolled fibroblast activation that results in exaggerated ECM accumulation and remodeling. Recent DNA microarray studies reveal that fibroblasts explanted from different anatomic locations differ markedly in their pattern of gene expression, suggesting that fibroblasts in different sites in the body could be considered distinct differentiated cell types. The apparent “positional memory” of fibroblasts is governed by genetic imprinting by the homeobox (HOX) family transcription factors.

**Table 83-3 Extracellular Factors Potentially Contributing to the Pathogenesis of Fibrosis in Systemic Sclerosis**

<table>
<thead>
<tr>
<th>Signal</th>
<th>Cellular Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGF-β</td>
<td>Inflammatory cells (macrophages, T cells), platelets, fibroblasts</td>
</tr>
<tr>
<td>PDGF</td>
<td>Platelets, macrophages, fibroblasts, endothelial cells</td>
</tr>
<tr>
<td>CTGF/CCN2</td>
<td>Fibroblasts</td>
</tr>
<tr>
<td>IGF-1</td>
<td>Keratinocytes</td>
</tr>
<tr>
<td>IL-1α</td>
<td>IL-4, IL-13</td>
</tr>
<tr>
<td>IL-6</td>
<td>Neutrophils, epithelial cells, endothelial cells, fibroblasts</td>
</tr>
<tr>
<td>Chemokines (MCP-1,</td>
<td></td>
</tr>
<tr>
<td>MCP-3)</td>
<td>Fibroblasts</td>
</tr>
<tr>
<td>Fibroblast growth</td>
<td>Endothelial cells</td>
</tr>
<tr>
<td>factor</td>
<td>Development pathway aberrantly reactivated</td>
</tr>
<tr>
<td>ET-1</td>
<td>Wnt ligands</td>
</tr>
<tr>
<td>Notch/Jagged</td>
<td>Developmental pathway aberrantly reactivated</td>
</tr>
<tr>
<td>Hypoxia</td>
<td>Hypoxic underperfused tissue</td>
</tr>
<tr>
<td>ROS</td>
<td>Generated from ischemia reperfusion</td>
</tr>
</tbody>
</table>

CTGF, connective tissue growth factor; ET-1, endothelin-1; IGF-1, insulin-like growth factor-1; IL, interleukin; MCP, monocyte chemotactic protein; PDGF, platelet-derived growth factor; ROS, reactive oxygen species; SSc, systemic sclerosis; TGF-β, transforming growth factor-β.

**Effector Cells of Fibrosis: Myofibroblasts, Pericytes, Endothelial Cells, and Cellular Plasticity**

In fibrosis, the tissue pool of activated mesenchymal cells is expanded by not only proliferation of resident fibroblasts but also transdifferentiation of other cell types, as well as the influx of bone marrow–derived mesenchymal progenitor cells. Myofibroblasts characterized by the cytoskeletal protein α-smooth muscle actin are specialized cells that arise from resident fibroblasts, as well as epithelial and endothelial cells and fibrocytes in response to TGF-β. Myofibroblasts synthesize collagens, tissue inhibitor of metalloproteinases (TIMPs), and other ECM components and are a major source of TGF-β during the fibrotic response. During normal wound healing, myofibroblasts are detected transiently in the early granulation tissue and then disappear via apoptosis. The removal of fibroblasts is a crucial step in wound resolution. In pathologic fibrogenesis myofibroblasts persist in lesional tissue, resulting in excessively contracted ECM that is characteristic of chronic scar.

Pericytes are mesenchymal cells in the walls of small blood vessels that are normally in intimate contact with the underlying endothelium and regulate vascular homeostasis. In SSc patients, a marked increase in microvascular pericyte compartment and increased expression of PDGF receptors has been reported. Activated pericytes can transdifferentiate into collagen-producing fibroblasts and myofibroblasts, thus linking microvascular injury and fibrosis.

Under certain conditions, epithelial cells can also undergo transformation to fibroblasts. Epithelial-mesenchymal transition (EMT) plays a vital role during vertebrate embryonic development. On stimulation, epithelial cells lose characteristic markers and cell-cell adhesion and acquire fibroblast markers such as α-smooth muscle actin. Pathologic EMT occurs prominently in cancer and is increasingly implicated in renal fibrosis and idiopathic pulmonary fibrosis. Similar to epithelial cells, vascular endothelial cells can, on injury, undergo a transition to fibroblasts. This process, called EndMT, has been demonstrated in various forms of experimentally induced fibrosis. Both EMT and EndMT are stimulated by TGF-β and Notch, and both processes might be implicated in tissue fibrosis in SSc.

**Fibrocytes and Monocyte-Derived Mesenchymal Progenitor Cells**

Fibrocytes are bone marrow–derived CD34+ mesenchymal progenitor cells normally present in small numbers in the peripheral blood that can present antigen and also synthesize collagen. These bone marrow–derived cells express CD14+ (a monocyte marker), as well as chemokine receptors (CCR3, CCR5, and CXCR4), which allows them to accumulate in specific tissues. The role for circulating fibrocytes and their trafficking into lesional tissue in the pathogenesis of fibrosis was established in animal models using neutralizing antibodies and in CXCR4-deficient mice. Mice with accelerated senescence have an increased number of circulating fibrocytes and show increased sensitivity to bleomycin-induced fibrosis. It is thought that in lesional tissue, fibrocytes undergo differentiation into activated myofibroblasts, losing the CD14 and CD34 markers in
the process, and contribute to the progression of fibrosis. Other studies have identified multipotent monocyte-derived mesenchymal progenitor cells in peripheral blood. However, the role of fibrocytes and other bone marrow-derived progenitor cells in SSc-associated fibrosis remains speculative.

**MOLECULAR DETERMINANTS OF FIBROSIS: TRANSFORMING GROWTH FACTOR-β**

The expression of ECM genes is normally tightly regulated by paracrine/autocrine mediators, cell-cell contact, hypoxia, and contact with the surrounding ECM. Multiple cytokines are implicated in SSc (see Table 83-3). Of these, TGF-β is considered to be the master regulator of both physiologic fibrogenesis (wound healing and tissue repair) and pathologic fibrosis. TGF-β plays essential roles in normal tissue repair, angiogenesis, and immune regulation and is implicated in cancer, fibrosis, and autoimmunity. Most cell types secrete TGF-β as a latent complex that is sequestered within the ECM. Under appropriate conditions, latent TGF-β is converted to its biologically active form that is capable of triggering cellular responses. The activation of latent TGF-β is controlled in part by fibrillin-1 and is mediated by integrins, thrombospondins, αvβ3 integrin, and proteolytic enzymes. Because of its fundamental role in orchestrating fibrotic responses, TGF-β is considered as a potential therapeutic target in SSc. Blocking TGF-β activity via biologic approaches using neutralizing antibodies and blocking intracellular TGF-β signaling using small molecular kinase inhibitors such as the c-Abl blocker imatinib mesylate (Gleevec) are undergoing evaluation. To date, small clinical trials of such interventions have not shown significant efficacy in modifying the course of fibrosis.

**Cellular Signaling by Transforming Growth Factor-β: Canonical Smad Pathways**

A member of a large cytokine superfamily that also includes activin and bone morphogenetic proteins, TGF-β is secreted by platelets, monocytes/macrophages, dendritic cells, and fibroblasts, and most cell types express surface receptors for TGF-β. The type of response elicited by TGF-β is specific for the target cell lineage and is context dependent. In mesenchymal cells, TGF-β acts as a potent inducer of fibrillar collagen synthesis; stimulates fibroblast proliferation, migration, adhesion, and transdifferentiation into myofibroblasts; and suppresses the production of matrix-degrading metalloproteinases (Table 83-4). In endothelial and epithelial cells, TGF-β drives transdifferentiation into fibroblasts. Activated TGF-β binds to the type II TGF-β receptor, triggering an intracellular signal transduction cascade that leads to the induction of target genes. The evolutionarily conserved canonical TGF-β signal transduction pathway involves phosphorylation of the type I TGF-β receptor, a serine-threonine kinase that in turn phosphorylates cytosolic Smads. Ligand-induced phosphorylation of Smad2/3 allows them to form heterocomplexes with Smad4 and translocate into the nucleus, where they bind to a consensus Smad-binding element (SBE) and recruit transcriptional cofactors such as the histone acetylase p300/CBP, resulting in gene transcription. Conserved SBE sequences are found in many TGF-β-inducible genes including type 1 collagens, PAI-1, α-smooth muscle actin, and CTGF. Ligand-induced signal transduction through the Smad pathway is tightly controlled by endogenous inhibitors such as Smad7 and BAMBI.

**Noncanonical Transforming Growth Factor-β Signaling**

Although the Smad pathway is the central mediator of signals from the TGF-β receptors, recent evidence indicates the existence of alternative non-Smad pathways. Non-Smad signaling molecules activated by TGF-β include protein kinases (c-Abl, p38, and JNK, integrin-associated focal adhesion kinase FAK, and TGF-β activated kinase TAK1); lipid kinases such as PI3 kinase and its downstream target Akt; and the calcium-dependent phosphatase, calcineurin. Signaling via c-Abl is particularly relevant to SSc because this nonreceptor tyrosine kinase implicated in chronic myelogenous leukemia (CML) mediates profibrotic signals induced by TGF-β and PDGF and is activated in SSc fibroblasts. Imatinib is an effective small molecule inhibitor of c-Abl that is highly effective for the treatment of CML. In scleroderma skin fibroblasts in culture imatinib reversed the abnormal ECM gene expression, and in mouse models of scleroderma, imatinib treatment prevented the development of skin fibrosis.

**CYTOKINES, GROWTH FACTORS, CHEMOKINES, AND LIPID MEDIATORS**

Multiple cytokines, growth factors, chemokines, and eicosanoids that regulate ECM accumulation and mesenchymal cell function show aberrant expression or activity in SSc. Soluble mediators such as CTGF, PDGF, IL-4, IL-6, IL-13, adenosine, prostaglandin F2α, and lysophosphatidic acid (LPA1) each contribute to the pathogenesis of fibrosis and therefore represent potential therapeutic targets.

---

**Table 83-4** Fibrogenic Activities of Transforming Growth Factor-β (TGF-β) Relevant to Systemic Sclerosis

<table>
<thead>
<tr>
<th>Activity</th>
<th>Molecular Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promotes epithelial-mesenchymal transition, endothelial mesenchymal transition</td>
<td>TGF-β activates Smad1/5, leading to differentiation</td>
</tr>
<tr>
<td>Inhibits fibroblast and myofibroblast apoptosis</td>
<td>TGF-β promotes apoptosis of fibroblasts and myofibroblasts</td>
</tr>
<tr>
<td>PDGF, platelet-derived growth factor</td>
<td>TGF-β promotes PDGF receptor signaling</td>
</tr>
</tbody>
</table>

#### Table 83-3

<table>
<thead>
<tr>
<th>Activity</th>
<th>Molecular Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruits monocytes</td>
<td>TGF-β binds to type II receptor</td>
</tr>
<tr>
<td>Stimulates synthesis of collagens, fibronectin, proteoglycans, elastin, tissue inhibitor of metalloproteinases, inhibits matrix metalloproteinases</td>
<td>TGF-β stimulates synthesis of collagens</td>
</tr>
<tr>
<td>Stimulates fibroblast proliferation, chemotaxis</td>
<td>TGF-β promotes fibroblast migration</td>
</tr>
<tr>
<td>Induces fibrogenic cytokine production: connective tissue growth factor, autoinduction, blocks type II interferon synthesis and activity</td>
<td>TGF-β induces cytokine production</td>
</tr>
<tr>
<td>Stimulates production of endothelin-1</td>
<td>TGF-β stimulates endothelin-1 production</td>
</tr>
<tr>
<td>Stimulates generation of reactive oxygen species</td>
<td>TGF-β promotes reactive oxygen species production</td>
</tr>
<tr>
<td>Stimulates expression of surface receptors for TGF-β, PDGF</td>
<td>TGF-β stimulates surface receptor expression</td>
</tr>
<tr>
<td>Induces fibroblast mitogenic responses to PDGF-AA</td>
<td>TGF-β stimulates fibroblast mitogenic responses</td>
</tr>
<tr>
<td>Promotes fibroblast-myofibroblast differentiation, monocyte-fibrocyte differentiation</td>
<td>TGF-β promotes fibroblast differentiation</td>
</tr>
<tr>
<td>Promotes epithelial-mesenchymal transition, endothelial mesenchymal transition</td>
<td>TGF-β promotes epithelial-mesenchymal transition</td>
</tr>
<tr>
<td>PDGF, platelet-derived growth factor</td>
<td>TGF-β promotes PDGF receptor signaling</td>
</tr>
</tbody>
</table>
Connective Tissue Growth Factor/CCN2

CTGF is a cysteine-rich 40-kD member of the CCN early-response gene family. This matricellular growth factor is implicated in angiogenesis, wound healing, and development. In normal adults, CTGF is undetectable, but its expression is markedly elevated in fibrotic conditions. Serum levels of CTGF correlate with the extent of skin and pulmonary fibrosis in SSc patients. The expression of CTGF can be induced in normal fibroblasts by TGF-β, IL-4, and VEGF, whereas TNF and iloprost block stimulation. Transgenic mice overexpressing CTGF develop scleroderma-like diffuse skin fibrosis and microvascular pathology. In vitro, CTGF exerts multiple profibrotic effects in normal fibroblasts. Because many of these CTGF effects parallel those induced by TGF-β, it has been suggested that TGF-β responses are mediated through endogenous CTGF. The identity of the cellular CTGF receptors and the mechanism of action underlying CTGF profibrotic responses remain incompletely characterized.

Platelet-Derived Growth Factor

PDGFs are disulfide-bonded heterodimeric proteins that act mainly on stromal cells and regulate the wound healing process. Originally isolated from platelets, PDGF isoforms are secreted from macrophages, endothelial cells, and fibroblasts. PDGF, signaling via the α and β transmembrane receptors, acts as a potent mitogen and chemotaxant for fibroblasts. Moreover, PDGF induces ROS generation and stimulates the synthesis of collagen, fibronectin, and proteoglycans, as well as the secretion of TGF-β1, MCP-1, and IL-6. Transgenic mice expressing a constitutively active PDGF-α receptor develop progressive fibrosis in the skin and multiple organs. Lesional skin fibroblasts from SSc patients show elevated PDGF and PDGF-β receptor expression, and PDGF levels are increased in the bronchoalveolar lavage fluid. In SSc patients, serum antibodies to the PDGF receptor induce fibroblast activation and ROS generation in vitro; however, these antibodies are not specific for SSc and have also been detected in patients with graft-versus-host disease.

Developmental Pathways: Wnt and Notch

Wnt and Notch developmental pathways required for embryogenesis appear to be deregulated in fibrosis and SSc. The Wnts comprise a family of poorly soluble glycoproteins with dual roles in cell-cell adhesion and transcriptional regulation. Although Wnts have essential roles in morphogenesis, stem cell homeostasis, and cell fate determination, abnormal Wnt signaling is implicated in colorectal cancer, as well as rheumatoid and osteoarthritis, osteoporosis, PAH, and aging. Intracellular Wnt signaling is mediated via canonical (β-catenin) and noncanonical pathways, and there is extensive cross-talk with TGF-β signaling. Transcription of a large number of genes with diverse biologic functions including multiple genes associated with tissue remodeling and pathologic fibrosis are induced by Wnts through β-catenin. Transgenic mice overexpressing Wnt10b or a constitutively-active mutant β-catenin develop exuberant wound healing, dermal fibrosis, and increased collagen synthesis in the skin. Lungs from patients with idiopathic pulmonary fibrosis show increased nuclear β-catenin accumulation at fibrotic foci. Expression profiling analysis of SSc skin biopsies shows elevated expression of Wnt ligands, Wnt receptors, and Wnt targets. Notch is a transmembrane receptor activated by its ligand, Jagged, which has a fundamental role in embryonic development, wound healing, and tissue repair. Notch signaling regulates endothelial cell and fibroblast responses including myofibroblast differentiation. A mouse model of scleroderma showed markedly activated Notch signaling in the skin and lungs, and the activity of ADAM-17, a protease induced by TGF-β and ROS that initiates Notch signal transduction, was elevated in SSc skin biopsies.

Interleukins

IL-1 is secreted by multiple cell types. A recent study showed that SSc epidermal keratinocytes are activated and secreted IL-1α. Co-culture of SSc keratinocytes with normal skin fibroblasts resulted in fibroblast activation and myofibroblast differentiation that was mediated by IL-1α secreted by keratinocytes. The Th2 cytokine IL-4 stimulates fibroblast proliferation, chemotaxis, collagen synthesis, and production of TGF-β, CTGF, and TIMP. Serum levels of IL-4 are elevated in SSc, and the number of IL-4-producing T lymphocytes is increased in peripheral blood and skin. IL-6, produced by monocytes and T lymphocytes, fibroblasts, and endothelial cells, stimulates collagen and TIMP-1 synthesis and promotes a Th2-polarized immune response. The biologic activities of IL-6 are mediated via the JAK-STAT intracellular signaling pathway shared with other cytokines. Serum levels of IL-6 are elevated in SSc and correlated with the severity of skin involvement. IL-13 is implicated in asthma and other fibrotic conditions. The profibrotic effects of IL-13 involve both indirect mechanisms due to stimulation of TGF-β production by macrophages and direct stimulation of fibroblast proliferation and collagen synthesis. Serum levels of IL-13 are elevated in SSc.

Chemokines

Chemokines represent a superfamily of more than 40 low-molecular-weight soluble mediators originally characterized by their chemotactic effects on leukocytes but now recognized to have a broad range of cellular targets and biologic activities. They play important roles in angiogenesis, wound healing, and fibrosis. The CC chemokine MCP-1 stimulates collagen production directly, as well as through induction of endogenous TGF-β production. Serum levels of MCP-1, along with those of MIP-1α, IL-8, CXCL8, and CCL18, are elevated in SSc and correlate with the severity of skin fibrosis. Mononuclear cells and dermal fibroblasts from SSc patients spontaneously produce these chemokines, and lesional SSc fibroblasts show constitutive upregulation of the MCP-1 receptor CCR2. Because MCP-1 drives a Th2 response, the MCP-1–CCR2 axis is thought to play a major role in the pathogenesis of SSc by amplifying collagen stimulation and promoting Th2 cytokine production. Significantly, MCP-1 null mice are resistant to the development...
Bioactive Lipids

A variety of bioactive lipids are potent modulators of fibroblast function. Although some prostanoids inhibit fibrotic responses through a variety of mechanisms, prostaglandin E2 (PGE2) was shown to be elevated in patients with pulmonary fibrosis and can stimulate collagen production and fibroblast proliferation. Mice with targeted deletion of the PGE2 receptor are protected from bleomycin-induced pulmonary fibrosis. Lysophosphatidic acid (LPA), generated via the hydrolysis of membrane phospholipids, exerts multiple biologic activities via G protein–coupled transmembrane receptors. Recently LPA was shown to induce fibroblast chemotaxis and CTGF production. The levels of LPA are elevated in the lungs of patients with pulmonary fibrosis. Moreover, LPA knockout mice are protected from bleomycin-induced skin and lung fibrosis. A recent study indicates that LPA induces αvβ6 integrin-mediated TGF-β activation in epithelial cells, contributing to sustained autocrine TGF-β signaling.

Regulation of Fibroblast Function via Innate Immune Signaling: Toll-like Receptors and the Inflammasome

The IFN signature observed in SSc leukocytes and genetic association with mediators of innate immunity support the role of TLR-mediated innate immune response in SSc. All cell membrane and endosomal TLRs are expressed on normal fibroblasts. Activation of TLR4 by lipopolysaccharide (LPS) plays a critical role in liver fibrosis, with sensitization to TGF-β as the underlying mechanism. In addition, TLR4 also induced the expression of the profibrotic transcription factors Egr-1 and Egr-2. In SSc, fibroblast TLRs might be activated by endogenous TLR ligands called damage-associated molecular patterns generated by tissue injury, autoimmunity, and oxidative stress. Endogenous TLR ligands that could play a role in SSc belong to three categories: ECM-derived molecules such as hyaluronan and its small-molecular-weight degradation products, tenascin C, alternatively spliced extra domain A (EDA) of fibronectin and biglycan; cellular stress proteins such as HMGBl and Hsp60; and nucleic acids and immune complexes released from damaged or necrotic cells. The expression of both TLR3 and TLR4 is elevated in SSc skin and lung biopsies and is accompanied by substantial increases in hyaluronan, tenascin C, and alternately spliced fibronectin levels. The TLR3 ligand poly(I:C) causes dramatic induction of type I IFN, IL-6, and other inflammatory cytokines, as well as ECM molecules, in normal fibroblasts. These observations suggest that fibroblasts exposed to endogenous TLR ligands during tissue injury switch to an activated phenotype. In this way, fibroblast TLR signaling initiated by damage-associated endogenous TLR ligands in SSc might convert a self-limited regenerative tissue repair into an aberrant and intractable fibrotic scar.

Recently identified innate immune sensors in addition to TLRs include NOD-like receptors (NLRs), RIG-I, and Naip. These intracellular receptors respond to cytosolic nucleic acids, damage-associated endogenous molecules, and environmental signals such as silica, bleomycin, and gadolinium. Once activated, these receptors facilitate inflammasome assembly with activation of caspase-1 and secretion of proIL-1β and IL-18. NLRC4 is a susceptibility gene for SSc and associated pulmonary fibrosis. Inflammasome activation and IL-1β play an important role in experimental mouse fibrosis and appear to be important in fibroblast activation of SSc as well.

Negative Regulation of Extracellular Matrix Accumulation

To prevent excessive matrix accumulation and scarring in response to injury, redundant biologic mechanisms have evolved. Fibroblasts are equipped with endogenous molecules that repress ECM gene expression and TGF-β stimulation. For example, Smad7 is an inhibitory Smad that blocks TGF-β signal transduction by accelerating ubiquitin-dependent TGF-β receptor degradation. Functional impairment of Smad7 was demonstrated in SSc fibroblasts. Other cell-intrinsic endogenous repressors of collagen synthesis include the transcription factors Sp3, Fli-1, p53, Ras, Nrf2, and the nuclear hormone receptor, peroxisome proliferator-activated receptor-γ (PPAR-γ). Impaired expression, induction, or function of these endogenous inhibitors may be responsible for failure to limit fibroblast activation in SSc.

Interferon-γ

IFN-γ, produced primarily by Th1 lymphocytes, is a major negative regulator of collagen gene expression and fibroblast activation. IFN-γ represses collagen gene expression and abrogates stimulation induced by TGF-β. IFN-γ is also a potent inhibitor of fibroblast proliferation, fibroblast-mediated matrix contraction, and myofibroblast transdifferentiation. Significantly, some studies have shown that fibroblasts from patients with SSc are relatively resistant to the inhibitory effects of IFN-γ. Clinical trials of IFN-γ in SSc have demonstrated a modest and inconsistent improvement in skin fibrosis.
Peroxisome Proliferator-Activated Receptor-γ

PPAR-γ is an intracellular molecule that modulates TGF-β signaling and mesenchymal cell plasticity and is functionally associated with fibrosis. Originally identified in adipocytes as a key regulator of adipogenesis and lipid metabolism, PPAR-γ is a dual function molecule acting as both a nuclear hormone receptor and ligand-inducible transcription factor. Multiple lipid moieties and electrophilic prostanooids such as 15d-prostaglandin J2 (15d-PGJ2) serve as endogenous ligands for PPAR-γ. Insulin-sensitizing drugs such as rosiglitazone and pioglitazone are potent pharmacologic PPAR-γ agonists. PPAR-γ modulates vascular and immune processes, and abnormal PPAR-γ function is implicated in lipodystrophy, atherosclerosis, PAH, and inflammatory diseases. Activation of fibroblasts with 15d-PGJ2 or pharmacologic PPAR-γ ligands resulted in a virtual abrogation of TGF-β-induced collagen production, myofibroblast differentiation, EMT, and other Smad3-dependent transcriptional responses. The expression and activity of PPAR-γ are impaired in patients with diffuse cutaneous SSc (dcSSc). Furthermore, PPAR-γ expression shows an inverse relationship with enhanced TGF-β signaling in lesional tissue. Of note, multiple factors implicated in SSc pathogenesis including TGF-β, Wnt ligands, IL-13, hypoxia, LPA, and CTGF potently inhibit PPAR-γ expression, which might account for impaired PPAR-γ expression in SSc. Together, these findings implicate altered PPAR-γ expression and function in SSc.

SCLERODERMA FIBROBLAST

Fibroblasts explanted from lesional skin or fibrotic lungs of patients with SSc display an abnormal activated phenotype that persists during their serial passage in vitro, indicating autonomous alteration in cell function. The “SSc phenotype” is characterized by the following: enhanced ECM synthesis, secretion of profibrotic cytokines and chemokines, and resistance to IFN-γ and other inhibitory signals. Moreover, SSc fibroblasts show features of myofibroblast differentiation, in part because of constitutive activation of the FAK focal adhesion kinase. It remains unsettled whether the activated SSc fibroblast phenotype represents an intrinsic abnormality or activation in response to exogenous stimuli in the fibrotic milieu.

Numerous molecules of signal transduction and transcriptional regulation have been reported to be abnormally expressed or constitutively activated in SSc fibroblasts. The list includes protein kinase C, Smad3, Egr-1, p300, and c-Abl. Elevated expression of the prosurvival factors Bcl-2 and Akt in SSc fibroblasts may play a role in their relative resistance to apoptosis. Because most of the SSc fibroblast characteristics can be induced in normal fibroblasts by treatment with TGF-β, it has been suggested that the SSc phenotype is due to autocrine TGF-β signaling. The levels of TGF-β receptors are elevated on SSc fibroblasts, which might enable these cells to mount a robust response to endogenously produced TGF-β or to low levels of environmental TGF-β. Furthermore, SSc fibroblasts have elevated levels of thrombospondin and αvβ3 integrins, which mediate latent TGF-β activation at the cell surface. Consistent with the autocrine TGF-β hypothesis, SSc fibroblasts show constitutive TGF-β signaling, with elevated Smad3 activation, and constitutive interaction with the histone deacetyltransferase p300/CBP. Other studies demonstrate defective expression or function of endogenous suppressors of TGF-β signaling and ECM production, suggesting that failure to terminate fibroblast activation may represent a fundamental defect in SSc. Endogenous molecules that negatively regulate fibroblast activation include Fli-1, PTEN, PPAR-γ, and Smad7. Autocrine TGF-β activation of fibroblasts cannot fully account for all of the phenotypic hallmarks of SSc fibroblasts such as constitutive CTGF production, indicating that both Smad-independent TGF-β signaling mechanisms, as well as non-TGF-β-mediated activation events, are involved in the induction or maintenance of the SSc phenotype. The autonomous SSc phenotype could also result from abnormal integrin-mediated signaling from the surrounding ECM. Moreover, epigenetic alterations in SSc fibroblasts are associated with persistent and heritable fibroblast dysfunction. For example, silencing the Fli-1 gene, an important endogenous negative regulator of collagen gene expression, by DNA methylation or chromatin histone deacetylation suppresses its expression in fibroblasts from lesional skin, which increases collagen synthesis.

References


The references for this chapter can also be found on www.expertconsult.com.
Clinical Features and Treatment of Scleroderma

FRANCESCO BOIN • FREDRICK M. WIGLEY

Some consider that the first description of systemic sclerosis (scleroderma) was put forth in 1753 by Cario Curzio (Naples, Italy). However, a careful review of the reported case suggests that the diagnosis was really sclerodema because of the distribution of the skin changes and the fact that the 17-year-old female improved with bloodletting, warm milk, and small doses of quicksilver. In 1836, Fantonetti (1791-1877), a Milanese physician, became the first to use the word sclerodermie to designate a skin disease in an adult. However, it is likely that this patient also had sclerodema. The first convincing case of scleroderma was reported in 1842; then several acceptable cases were published before 1847, when Gintrac used the term sclerodermie, establishing this condition as a specific clinical entity. By 1860, numerous cases had been reported, and articles reviewing the disease were published. Maurice Raynaud (1834-1881) described a patient with sclerodermie and cold-induced “asphyxie locale”—this was the first description of Raynaud’s phenomenon in scleroderma. Sir William Osler described scleroderma while at Johns Hopkins Hospital between 1891 and 1897. Osler recognized the systemic nature of the disease and was so impressed with the severity of scleroderma that he wrote:

In its more aggravated forms, diffuse scleroderma is one of the most terrible of all human ills. Like Tithonous, to “whither slowly,” and like him to be “beaten down and marred and wasted” until one is literally a mummy, encased in an ever-shrinking, slowly contracting skin of steel, is a fate not pictured in any tragedy, ancient or modern. Matsui (Japan, 1924) emphasized the importance of visceral involvement as part of scleroderma based on several autopsies. Goetz (Capetown, 1945) further confirmed the multisystem involvement and suggested that the disease be named progressive systemic sclerosis.

The concept of subtypes of scleroderma began in 1964, when Winterbauer reported cases with the CRST (calcinosis, Raynaud’s phenomenon, sclerodactyly, and telangectasia) syndrome. A similar group of patients was reported in 1920 and was named after the authors—the Thiberg-Weissenbach syndrome. Velayos and colleagues recognized that esophageal dysmotility was common in these patients, so it is now called the CREST syndrome. In 1969, 58 autopsy cases of scleroderma were compared with matched controls. The organs found to be frequently and significantly involved by this disease were the skin, gastrointestinal tract, lungs, kidneys, skeletal muscle, and pericardium. This report first described the systemic nature of scleroderma vascular disease with findings of both kidney and lung arterial changes. In 1979, Rodnan introduced a clinical method to evaluate the extent of skin disease. Steen and Medsger with others conducted extensive surveys of large populations of scleroderma patients, defining the clinical course and specific subtypes of disease. In the 1970s, an expert subcommittee established diagnostic criteria, and Leroy and colleagues suggested the classification of two major subsets of disease defined by skin involvement: limited and diffuse. Later, work by several investigators recognized that scleroderma has an autoimmune basis with occurrence of specific autoantibodies associated with subtypes of disease and useful in predicting disease course.

Scientific work in the modern era has revealed details regarding the pathogenesis of the disease and has led to the recognition that scleroderma is a complex polygenetic autoimmune disease associated with a unique disease process involving tissue fibrosis. Although no drug has yet been discovered with clear disease-modifying properties and ability to fully control the underlying disease process, major progress has been made in managing specific organ disease. The discovery that an angiotensin-converting enzyme inhibitor could reverse the scleroderma renal crisis in the 1970s has changed the course of kidney disease in scleroderma and has improved the survival of patients. Current therapies for gastrointestinal, cardiac, pulmonary, vascular, and interstitial lung disease have improved quality of life and survival.

EPIDEMIOLOGY

Incidence and Prevalence

Scleroderma is a rare disease with an estimated incidence of approximately 18 to 20 cases per million population per
year, and a prevalence of 100 to 300 cases per million population. Reported rates vary by method of ascertainment, population under study, and definition of disease. Scleroderma is found in all races and in various geographic areas, but the prevalence and severity of disease vary among different racial and ethnic groups. The prevalence of scleroderma appears to be greater in the United States, where it has been estimated at 24.2 per 100,000, than in Europe, where recent studies in Iceland, England, France, and Greece have estimated rates ranging from 7.1 to 15.8 per 100,000. In Australia, estimates are similar to those in the United States, and Japan has a lower reported prevalence—similar to that in Europe. The highest scleroderma prevalence is reported among Choctaw Native Americans, in whom the disease appears to be more severe. Several surveys in the United States show that African-Americans have a higher age-specific incidence rate and experience more severe disease than whites. Occurrence is not different from urban to rural areas. Occasional reports have described unusually large numbers of cases of scleroderma observed in restricted geographic regions, suggesting nonrandom distribution of the disease. For example, in a rural area in the province of Rome, a geographic cluster of scleroderma and disease with related features was reported, suggesting prevalence 1000 times greater than expected. Likewise, a study in the United Kingdom reported a higher prevalence of scleroderma in three geographic areas clustered near airports compared with other areas of the United Kingdom.

The average age at onset is between 35 and 50 years, and the disease is more common among women (3 to 7:1 female-to-male ratio). Disease onset in the elderly is well described; it is uncommon for the disease to become manifest before the age of 25 years. Older age at onset is associated with increased risk for multisystem disease and, in particular, pulmonary arterial hypertension (PAH). A progressive increase in the incidence of scleroderma has been noted with increasing age. Over the period from 1963 to 1982, in Pittsburgh, Pennsylvania, the highest rates of scleroderma were observed between the ages of 45 and 54 years in black women, and between the ages of 55 and 64 years in white women. Younger age at disease diagnosis among black women compared with white women was reported from Detroit, Michigan.

**Survival**

Mortality among scleroderma patients is high, and most deaths are attributed directly to the disease manifestations. In fact, the prognosis in scleroderma is highly variable, and survival is influenced by disease subtype, degree of internal organ involvement, and comorbid conditions. Factors associated with poor prognosis include diffuse skin disease, the presence of pulmonary disease (particularly PAH), renal involvement, multisystem disease, evidence of heart disease, older age at disease onset, and the presence of anemia. The standardized mortality rate (SMR) is the measure used to assess the relative mortality of a disease in comparison with the general population. Surveys of scleroderma patients report an SMR from 1.46 to 7.1. A survey found that among 284 scleroderma patients who died, the median disease duration as estimated from the onset of Raynaud’s phenomenon until death was 7.1 years for patients with diffuse skin disease, and 15.0 years for those with limited skin disease. Among non-scleroderma-related causes of death are, as expected, infection, malignancy, and cardiovascular events. Although premature atherosclerosis has been implicated as the cause of early death in other inflammatory autoimmune diseases, an increased prevalence of macroscopic coronary artery or cerebrovascular involvement beyond that expected in the general population has not yet been demonstrated in scleroderma.

Several studies suggest that the main cause of death has changed over time. Improved survival in recent years is thought to be secondary to more effective treatments for specific organ-based complications. It seems clear that the natural history of scleroderma renal complications has changed secondary to the use of angiotensin-converting enzyme (ACE) inhibitors for renal crisis. Historically, patients with scleroderma renal crisis had 1-year survival less than 15%. Recent case-control studies conducted after the introduction of ACE inhibitors suggest greater than 85% 1-year survival. Pharmacologic prevention and treatment of interstitial lung disease and PAH are thought to have an impact. Cohort studies have demonstrated improved overall survival among scleroderma patients. At one scleroderma center, 10-year survival from disease diagnosis improved from 54% in the 1972-1981 group to 66% in the 1982-1991 group.

Environmental Exposures

Geographic clustering of scleroderma cases suggests that environmental exposure may be responsible for the disease, but definitive proof is still lacking. Epidemics of scleroderma-like illness following exposure to dietary (L-tryptophan or adulterated rapeseed oil), occupational (silica), or pharmacologic toxins (e.g., bleomycin, gadolinium-based contrast media) support the idea that environmental exposure can trigger a fibrotic disease. Typical scleroderma is found among coal miners and workers exposed to silica, particularly male workers. A meta-analysis of the risk of silica exposure found that the combined estimator of relative risk (CERR) was 1.03 (95% confidence interval [CI], 0.74 to 1.44) in females and 3.02 (95% CI, 1.24 to 7.35) in males. Although occupational exposure to solvents (paint thinner or removers, mineral spirits, trichloroethylene, trichloroethane, perchloroethane, gasoline, aliphatic hydrocarbons, halogenated hydrocarbons, and BTX solvents containing benzene, toluene, or xylene) or polyvinyl chloride is reported to be associated with scleroderma, the role of these agents in causing disease is controversial and remains unproven. Likewise, case reports have suggested that silicone present in breast implants has triggered scleroderma, but large epidemiologic surveys have not found a greater incidence of scleroderma than expected in women with breast implants or exposure to silicone. Drugs implicated as potentially causative for systemic sclerosis (SSc)-like illnesses include bleomycin, pentazocine, and cocaine.
Genetic Factors

Although the absolute risk for family clustering remains low (1%), studies from Australia and the United States suggest that a family history of scleroderma increases the risk of disease, with a relative risk of first-degree relatives around 14 and of siblings in the range of 15 to 19. Genetic factors were investigated by a survey of 42 twin pairs among which at least one twin had scleroderma. The overall concordance rate was low at 4.7%, and was similar between monozygotic and dizygotic twin pairs. Although an association between major histocompatibility complex (MHC) alleles and scleroderma has been reported, this has not been confirmed consistently. This finding argues against strong genetic susceptibility and favors a role for environmental and/or epigenetic events.

OVERVIEW OF CLINICAL FEATURES

Diagnostic Criteria

A subcommittee of the American College of Rheumatology (ACR) established diagnostic criteria based on a consensus of experts who evaluated a multicenter survey of scleroderma patients compared with other patient groups (Table 84-1). The purpose of these criteria was to provide diagnostic certainty and consistency for research and other documentary exercises. The single finding of thickening of the skin typical of scleroderma proximal to the metacarpophalangeal (MCP) joints of the hands is considered a major criterion confirming the diagnosis. This includes changes on the face, arms, legs, or trunk. If skin thickening is found only distal to the MCP joints, then two of three minor criteria (digital pits, sclerodactyly, and pulmonary fibrosis on chest radiograph) must be present to confirm the diagnosis. These criteria are very specific (98%) for making the diagnosis of scleroderma but are considered too restricted in that they exclude patients with early or mild expression of the disease. Many patients with skin thickening limited to the fingers and some with no skin changes are recognized to have systemic disease and yet do not meet the ACR criteria. Large longitudinal surveys confirm that a high percentage of patients with definitive Raynaud’s phenomenon and typical scleroderma nail-fold capillary changes, along with scleroderma-specific autoantibodies, develop scleroderma within a 2- to 4-year period of follow-up. Therefore, although the ACR criteria have a definite role in defining cases for research purposes, the clinician needs to take into account subtle features of the disease to make an early diagnosis for management purposes. Most experts would accept the finding of three of the five features of the CREST syndrome (calcinosis, Raynaud’s phenomenon, esophageal dysfunction, sclerodactyly, telangiectasias) to make a clinical diagnosis of scleroderma. In clinical practice, one needs to recognize that some patients will not develop skin changes and yet they do have systemic disease (systemic sclerosis sine scleroderma), and many patients will present with partial expression of the disease (e.g., Raynaud’s with nail-fold changes only) and later will develop well-defined disease. Experts argue that patients with partial expression, especially when a specific scleroderma-related autoantibody is present, should be considered to have a diagnosis of scleroderma and should be treated accordingly.

Classification and Clinical Subsets

The scleroderma disease process is complex, and its clinical expression is very heterogeneous such that the disease is expressed in several distinct clinical phenotypes. Classifying patients into subtypes is useful both for investigative purposes and for clinical practice. A specific subtype can define increased risk for internal organ involvement and overall prognosis. The traditional way to classify patients is by the degree of skin thickening as determined by physical examination. The skin is scored by pinching a fold and deciding whether it is abnormally thickened because of the scleroderma process. A committee of experts decided by consensus that two major groups of patients can be identified based on the distribution of skin changes and associated clinical and laboratory outcomes as observed (Figure 84-1). Patients are considered to have diffuse skin disease if skin changes are found proximal to the elbows and/or knees or on the trunk, excluding the face. These patients tend to have higher risk of multisystem disease and poor survival. Patients are considered to have limited disease if skin changes occur distal to the elbows and/or knees and not on the trunk. Facial skin thickening can be present in the limited group. Some argue that the term CREST syndrome should be eliminated, and that these patients should be classified in the limited skin group. Others feel that the CREST syndrome is a unique subtype within the broader group of limited skin diseases with a distinct disease course.

Another less popular classification system divides patients into three groups based on skin changes: limited (fingers only), intermediate (skin to the elbows or knees), and diffuse (skin above the elbows and/or knees and/or trunk). Studies using this classification have found that the intermediate group had distinct clinical outcomes, including an intermediate survival statistic between limited (best) and diffuse (worse). A subtype of disease with absence of skin fibrosis has been recognized and is referred to as scleroderma sine scleroderma. Serologic studies have shown that the

<table>
<thead>
<tr>
<th>Table 84-1 Establishing Diagnosis of Scleroderma</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ACR Criteria</strong></td>
</tr>
<tr>
<td>Must have (a) or two of (b), (c), or (d)</td>
</tr>
<tr>
<td>a. Proximal SSc (proximal to MCPs/MTPs)</td>
</tr>
<tr>
<td>b. Digital pits</td>
</tr>
<tr>
<td>c. Sclerodactyly</td>
</tr>
<tr>
<td>d. Pulmonary fibrosis (CXR; HRCT)</td>
</tr>
<tr>
<td><strong>CREST Criteria</strong></td>
</tr>
<tr>
<td>Must have three of the five features</td>
</tr>
<tr>
<td>Calcinois</td>
</tr>
<tr>
<td>Raynaud’s phenomenon</td>
</tr>
<tr>
<td>Esophageal dysmotility</td>
</tr>
<tr>
<td>Sclerodactyly</td>
</tr>
<tr>
<td>Telangiectasias</td>
</tr>
<tr>
<td><strong>Minor Criteria†‡</strong></td>
</tr>
<tr>
<td>Must have all three</td>
</tr>
<tr>
<td>Definite Raynaud’s phenomenon</td>
</tr>
<tr>
<td>Abnormal capillary loops</td>
</tr>
<tr>
<td>Specific scleroderma autoantibody</td>
</tr>
</tbody>
</table>

*Some experts continue to classify patients with minor criteria only as undifferentiated connective disease with scleroderma features.
†Proximal areas include the face. Autoantibody includes anticientromere, antinuclear antibodies, and specific scleroderma autoantibodies.
ACR, American College of Rheumatology; CREST, calcinosis, Raynaud’s phenomenon, esophageal dysfunction, sclerodactyly, telangiectasias; CXR, chest x-ray; HRCT, high-resolution computed tomography; MCPs, metacarpophalangeal joints; MTPs, metatarsophalangeal joints; SSc, systemic sclerosis.
CHAPTER 84 | CLINICAL FEATURES AND TREATMENT OF SCLERODERMA

Scleroderma tends to be a chronic monophasic illness, unlike systemic lupus erythematosus, which is characterized by quiescence interrupted by sudden disease flares. The pattern of the disease and the degree or type of internal organ involvement are variable but generally can be predicted by the extent of skin affected. Internal organ disease is more severe in patients with the diffuse skin type than in those with limited scleroderma. The disease course is usually indolent in patients with sclerodactyly alone. In these patients, years pass from onset of Raynaud’s phenomenon before obvious organ dysfunction manifests, although explosive multisystem disease can emerge rapidly over several months in patients with widespread skin changes. Likewise, when patients with diffuse skin disease experience skin improvement, new internal organ disease is generally uncommon. The degree of skin thickening varies in extent and progression over time, with some cases stopping after several months with relatively mild disease, and others progressing over years to involve most of the body surface.

Natural History of Disease

The natural history of scleroderma varies a great deal depending on the specific subtype of disease established in a patient. Scleroderma tends to be a chronic monophasic illness, unlike systemic lupus erythematosus, which is characterized by quiescence interrupted by sudden disease flares. The pattern of the disease and the degree or type of internal organ involvement are variable but generally can be predicted by the extent of skin affected. Internal organ disease is more severe in patients with the diffuse skin type than in those with limited scleroderma. The disease course is usually indolent in patients with sclerodactyly alone. In these patients, years pass from onset of Raynaud’s phenomenon before obvious organ dysfunction manifests, although explosive multisystem disease can emerge rapidly over several months in patients with widespread skin changes. Likewise, when patients with diffuse skin disease experience skin improvement, new internal organ disease is generally uncommon. The degree of skin thickening varies in extent and progression over time, with some cases stopping after several months with relatively mild disease, and others progressing over years to involve most of the body surface.

Patients classified in the limited scleroderma group may have intermediate skin involvement, with skin extending over the hands and forearms or lower legs but not over the trunk. These patients appear to have an intermediate disease course that is not as severe as that of patients with...
widespread skin disease. The intermediate group has decreased survival compared with patients with only sclerodactyly. Although the degree of skin disease provides a rough guide by which to predict disease course, many exceptions have been noted. Variability is associated with other clinically defined factors, including age, gender, and racial/ethnic background. Late age at onset increases the risks for PAH and poor survival. Males may have a more aggressive disease course than females. African-American or Native American race predicts a worse disease course and worse multisystem disease.9

**GENERAL PRINCIPLES OF DISEASE EVALUATION**

Measuring Disease Activity and Severity

One of the challenges in studying or caring for patients with scleroderma consists of determining and measuring disease activity, severity, and damage. These measures are not independent of each other and thus cannot be considered in isolation. Activity is a measure of ongoing disease that is dynamic and reversible; it is a direct reflection of ongoing biologic processes such as autoimmune-mediated inflammation or progressive tissue fibrosis. Severity is a measure of the impact of damage on the whole person or a specific organ at any given time. In the extreme, severity is a measure of irreversible end-stage damage. It is important to distinguish reversible disease activity from irreversible damage in that both can cause clinical distress, yet intervention differs for the two. Patients with scleroderma accumulate irreversible damage over time, thus the measure of activity is best done in early disease, when disease has the potential of being reversed.

Measurement of changes in skin thickness (modified Rodnan skin score) is used as a surrogate measure of disease severity and a predictor of extent of organ involvement and overall prognosis (see skin manifestation section).26 It is also used as a quantifiable, easily obtainable, and valid score, changing over time in response to therapy. Thus, the skin score is now used as a measure of disease activity and is employed in clinical trials as a primary outcome measure. In patients with diffuse skin disease, improvement in skin score is associated with better clinical outcome. A core set of items covering 11 domains (skin, musculoskeletal, cardiac, pulmonary, renal, gastrointestinal, health-related quality of life measures, global health, Raynaud’s phenomenon, digital ulcers, and the biomarkers erythrocyte sedimentation rate and C-reactive protein) is now considered to provide the minimum clinically relevant treatment effect values, as defined by consensus of experts in the field.27 More detailed scales for specific organs are also in use. For example, a skin assessment tool combines the modified Rodnan skin score with a visual analogue score (VAS) of the patient’s assessment of skin activity, a VAS score of the physician’s assessment of skin activity, and measures obtained by a durometer.28 A gastrointestinal scale has been developed to define mild to severe gastrointestinal disease.29 Experts who reviewed current outcome measures in assessing PAH as related to scleroderma decided that lung vascular and pulmonary arterial pressures and cardiac function as measured by right heart catheterization and echocardiography; exercise testing as measured by 6-minute walk and oxygen saturation with exercise, New York Heart Association functional class, severity of dyspnea on a VAS, and measures of quality of life and function obtained by Short Form 36 and the Health Assessment Questionnaire and Disability Index (HAQ-DI); and physician global scale as measured by survival are all important measures. Other outcome measures that are considered valid include Raynaud’s condition score for Raynaud’s severity, forced vital capacity, and Mahler’s dyspnea index to follow interstitial lung disease, serum creatinine, blood pressure, and complete blood count during a scleroderma renal crisis, as well as serum creatine phosphokinase to follow muscle disease.

A scleroderma severity scale has been developed by Medsger to assess disease severity status in individual patients both at a given time and longitudinally (Table 84-2).32 This scale is based on scoring that ranges from 0 (normal) to 4 (end-stage) for each organ system involved in scleroderma, including a general measure, along with measures of the peripheral vascular system, skin, joints and tendons, gastrointestinal tract, lung, heart, and kidney. This severity scale is now used by many experts to define the status of a patient in clinical trials and in research investigating risk for clinical outcomes.

The Health Assessment Questionnaire and Disability Index (HAQ-DI) is a self-administered tool initially developed for rheumatoid arthritis to measure functional impairment; it has been validated and used in scleroderma as well, where it correlates with objective physical signs and reflects the disease course. Higher degrees of impairment are found by HAQ scores in scleroderma patients with diffuse skin disease, higher skin scores, abnormal hand function, presence of myopathy or friction rubs, and joint pain.33 Change in the HAQ score occurs with changes in skin involvement and progressive organ disease.34 The HAQ is useful in evaluating patients with Raynaud’s and correlates with the severity and presence of digital ulcers. VAS instruments that assess disease in various domains specific for scleroderma are added to the HAQ to focus on scleroderma-specific issues, including measures of Raynaud’s and digital ulcers. In addition, the Short Form 36 survey, a simple self-report tool composed of 36 items in 8 domains, is used to assess health-related quality of life.

Developing an activity scale is much more challenging than establishing severity measures. Attempts include evaluating changes in skin score and evidence of active lung disease and documenting congestive heart failure or a scleroderma renal crisis. A composite score generated as a physician global assessment has been used to measure disease activity. This simple global score has merit but is influenced, besides activity, by organ severity and irreversible damage.35 A European study group developed the Valentini disease activity index, which currently is used in clinical studies.36 This index provides more of an organ-specific assessment than a measure of global activity. The Medsger disease severity scale has the potential to measure activity if used in studies with serial observations, when individual organ scales can be combined into a composite score. Intermediate clinical and serologic biomarkers are under investigation as candidates to define and predict disease activity and prognosis, but none has proved reliable or valid. Measures of acute phase factors (e.g., erythrocyte
scleroderma. The three most frequently observed types of scleroderma-specific autoantibodies are anticientromere, antitopoisomerase I, and anti-RNA polymerase III antibodies. Anticientromere patients tend to be older white women. Skin disease usually is limited and most often involves just the fingers. The disease course is indolent, and often the diagnosis is delayed until anticentromere antibodies are discovered or late organ disease becomes evident. Patients frequently develop features of the CREST syndrome. Subcutaneous calcinosis is a late manifestation, appearing in areas such as small clusters on the fingers and along the forearm and anterior lower leg. Raynaud’s phenomenon can become severe and is associated with increased risk for digital ischemia, PAH, sicca syndrome, calcinosis.

### Autoantibodies

Measurement of the autoantibodies present in scleroderma can be helpful in determining the clinical features and prognosis of the disease, in that specific scleroderma-related autoantibodies have been established as strong predictors of disease outcome and the pattern of organ complications.

#### Table 84-2 Medsger Systemic Sclerosis Severity Scale*

<table>
<thead>
<tr>
<th>Organ System</th>
<th>0 (Normal)</th>
<th>1 (Mild)</th>
<th>2 (Moderate)</th>
<th>3 (Severe)</th>
<th>4 (End Stage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>Wt loss &lt;5%; Ht 37%-45%; Hb 12.3+ g/dL</td>
<td>Wt loss 5%-10%; Ht 33%-37%; Hb 11.0-12.2 g/dL</td>
<td>Wt loss 10%-15%; Ht 29%-33%; Hb 9.7-10.9 g/dL</td>
<td>Wt loss 15%-20%; Ht 25%-29%; Hb 8.3-9.6 g/dL</td>
<td>Wt loss 20+%; Ht 25%; Hb &lt;8.3 g/dL</td>
</tr>
<tr>
<td>Peripheral</td>
<td>No Raynaud’s; Raynaud’s not requiring vasodilators</td>
<td>Raynaud’s requiring vasodilators</td>
<td>Digital pitting scars</td>
<td>Digital tip ulcerations</td>
<td>Digital gangrene</td>
</tr>
<tr>
<td>Skin</td>
<td>TSS 0</td>
<td>FTP 0.9 cm</td>
<td>TSS 1-14</td>
<td>TSS 15-29</td>
<td>TSS 30-39</td>
</tr>
<tr>
<td>Joint/tendon</td>
<td>Normal proximal muscle strength</td>
<td>Normal esophagram; normal small bowel series</td>
<td>Distal esophageal hypoperistalsis; small bowel series abnormal</td>
<td>Antibiotics required for bacterial overgrowth</td>
<td>Malabsorption syndrome; episodes of pseudo-obstruction</td>
</tr>
<tr>
<td>Muscle</td>
<td>Lung</td>
<td>DLCO 80+%; FVC &gt;80%; no fibrosis on radiograph; sPAP &lt;35 mm Hg</td>
<td>DLCO 70%-79%; FVC 70%-79%; basilar rales; fibrosis on radiograph; sPAP 35-49 mm Hg</td>
<td>DLCO 50%-69%; FVC 50%-69%; sPAP 50-64 mm Hg</td>
<td>Oxygen required</td>
</tr>
<tr>
<td>Gastrointestinal tract</td>
<td>ECG normal; LVEF 50+4%</td>
<td>ECG conduction defect; LVEF 45%-49%</td>
<td>ECG arrhythmia; LVEF 40%-44%</td>
<td>ECG arrhythmia requiring therapy; LVEF 30%-40%</td>
<td>CHF; LVEF &lt;30%</td>
</tr>
<tr>
<td>Kidney</td>
<td>No history of sRC with serum creatinine &lt;1.3 mg/dL</td>
<td>History of SRC with serum creatinine &lt;1.5 mg/dL</td>
<td>History of SRC with serum creatinine 1.5-2.4 mg/dL</td>
<td>History of SRC with serum creatinine 2.5-5.0 mg/dL</td>
<td>History of SRC with serum creatinine &gt;5.0 mg/dL or dialysis required</td>
</tr>
</tbody>
</table>

*If two items are included for a severity grade, only one is required for the patient to be scored as having disease of that severity level.


#### Table 84-3 Autoantibodies and Associated Phenotypes in Scleroderma

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Subtype</th>
<th>Clinical Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topoisomerase 1 (Scl70)</td>
<td>Diffuse</td>
<td>Pulmonary fibrosis, cardiac involvement</td>
</tr>
<tr>
<td>Centromere (protein B, C)</td>
<td>Limited</td>
<td>Severe digital ischemia, PAH, sicca syndrome, calcinosis</td>
</tr>
<tr>
<td>RNA polymerase III</td>
<td>Diffuse</td>
<td>Severe skin disease, tendon rashes, renal crisis (sine scleroderma)</td>
</tr>
<tr>
<td>U3-RNP (fibrillarin)</td>
<td>Limited</td>
<td>Primary PAH; esophageal, cardiac, and renal involvement; muscular disease</td>
</tr>
<tr>
<td>Th/To</td>
<td>Diffuse or limited</td>
<td>Pulmonary fibrosis, rare renal crisis, lower GI dysfunction</td>
</tr>
<tr>
<td>B23</td>
<td>Limited</td>
<td>PAH, lung disease</td>
</tr>
<tr>
<td>Cardiolipin, β2GPI</td>
<td>Diffuse or limited</td>
<td>PAH, digital loss</td>
</tr>
<tr>
<td>PM/Scl</td>
<td>Overlap</td>
<td>Myositis, pulmonary fibrosis, acro-osteolysis</td>
</tr>
<tr>
<td>U1-RNP</td>
<td>Overlap</td>
<td>SLE, inflammatory arthritis, pulmonary fibrosis</td>
</tr>
</tbody>
</table>

Gl, gastrointestinal; GPI, glycoprotein I; PAH, pulmonary arterial hypertension; RNA, ribonucleic acid; RNP, ribonucleoprotein particle; SLE, systemic lupus erythematosus.

macroversal disease. Higher risk for digital gangrene and amputation is noted in the anticytomegol-ative-positive group.Interstitial lung disease is less common and is isolated and progressive; pulmonary vascular disease with PAH and right heart failure occur in a significant proportion of patients. Overall, however, anticytomegol antibodies generally are predictors of a favorable prognosis. The presence of anticytomegol antibodies can also be seen in patients with primary Sjögren’s syndrome who do not have scleroderma.

Antitopoisomerase I antibodies are seen in 30% of African-American patients and are correlated with a poor prognosis and higher scleroderma-related mortality. Interstitial lung disease (ILD) is highly associated with antitopoisomerase, independent of the degree of skin disease. Patients usually have diffuse skin disease and are at high risk for rapid skin changes and renal crisis, usually in the first few years from disease onset. Raynaud’s phenomenon may be the first symptom, followed in the first years of disease by skin changes and joint-tendon involvement leading to contractures, particularly in the fingers and elbows. The degree of skin thickening varies a great deal in these patients, but usually the level of severity of skin involvement is established in the first 1 to 3 years of disease.

Anti-RNA polymerase III antibodies are associated with rapid and aggressive diffuse skin disease and renal involvement. These patients have the worst cutaneous involvement with rapid widespread skin disease associated with signs and symptoms of deep tissue fibrosis involving joints, tendons, and muscles. Flexion contractures of fingers, wrist, elbows, shoulders, hips, knees, and ankles are complications that can occur within a few months of onset of disease activity. Friction rubs are common. A “fibrosing” myopathy without inflammation, along with skin and joint disease, leads to loss of strength and flexibility. Significant disability occurs in early disease. Anti-RNA polymerase III–positive patients are not likely to have severe gastrointestinal disease and are relatively protected from developing ILD or pulmonary vascular disease. However, they are at high risk (approximately 25% to 40% of patients) for developing scleroderma renal disease with hypertensive crisis. Risk is greatest in the early years, when skin is rapidly changing. Heart disease can be a late complication, but major disability from skin and deep tissue disease is the most worrisome long-term problem in this subtype of scleroderma.

Other antinucleolar antibodies are found among subgroups of patients with scleroderma and are associated with specific clinical phenotypes and clinical outcomes. Of these, anti-Th/To antibodies and anti-PM/Scl antibodies are associated with limited skin disease, whereas anti-U3-RNP (fibrillarin) antibodies are associated with diffuse disease. Anti-Th/To is an increased risk for development of severe ILD and pulmonary hypertension. Anti-U3-RNP is another predictor of a less favorable prognosis with a higher frequency of internal organ involvement, including interstitial lung disease, PAH, and renal crisis. It is frequently present in African-Americans and is associated with diffuse skin disease and poor prognosis. Hyperpigmented and hypopigmented skin changes are common, and contractures of large joints associated with inflammatory or fibrotic muscle disease lead to significant morbidity. Cardiac disease is often subclinical until late-stage disease, when both right and left heart failure can occur.

Anti-PM/Scl, anti-Ku, and anti-U1-RNP antibodies are seen in patients with overlap syndromes. The presence of anti-Pm/Scl is associated with acute onset of weakness due to inflammatory muscle disease and higher risk for interstitial lung disease. Anti-Ku positivity has been demonstrated to be strongly associated with muscle and joint involvement. Anti-U1-RNP is seen in patients with mixed connective tissue disease and is common among African-Americans with limited scleroderma. Polymyositis, myositis, and lupus with skin or renal involvement are common complications. A subset can develop diffuse skin disease, and risk for late onset of PAH is increased overall. Although an uncommon anti-B23 autoantibody is associated with pulmonary arterial hypertension, up to 11% of patients with scleroderma can test negative for antinuclear antibodies.

**CLINICAL MANIFESTATIONS AND TREATMENT**

**General Principles of Management**

Several basic principles can be of help in managing patients with scleroderma. The first is that scleroderma encompasses subsets of disease, each with a unique clinical phenotype; thus carefully determining the specific subtype of the patient sets the scene for deciding therapy. Clinical problems and disease activity in scleroderma are highly variable in expression, even within a given subtype. This variability is thought to be secondary to different susceptibility to, or risk for, systemic complications. It is also due to the presence of a dynamic biologic process unique to scleroderma. Therefore, patients move through stages of the disease process with disease activity that varies with time and can shift in nature. Signs of classic inflammatory events are often present during the initial phase of the disease, but later, an indolent subclinical fibrotic process is dominant and gradually causes organ damage. Although the skin may be the site of the most dramatic physical findings, scleroderma is a multisystem disease, and the physician must search for early organ-based complications, including cardiopulmonary, gastrointestinal, renal, or musculoskeletal involvement. Understanding the shortcomings of current investigations, not getting locked into “traditional” therapy, focusing on an organ-specific approach, and defining the disease subtype and level of disease activity are most important in establishing optimal management. Although no good studies have been conducted to prove the effectiveness of this approach, it seems reasonable to use combination therapy with rapid control of the inflammatory immune process early on, followed by maintenance immunosuppression for an extended time. The anti-inflammatory program is then coupled with organ-specific therapy (see later sections). The missing component in this comprehensive approach is the availability of a direct antifibrotic agent able to prevent the progression of fibrosis once the inflammatory phase is under control. Obviously, early intervention is most important for success; once irreversible tissue or organ damage is noted, supportive care becomes the main and often the only option.
RAYNAUD’S PHENOMENON

KEY POINTS

Raynaud’s phenomenon (RP) is common in the general population and usually presents with a benign clinical course.

In scleroderma, RP is more clinically symptomatic and can be associated with manifestations of digital ischemia.

Abnormal vasomotor regulation and progressive endothelial and structural vessel disease occur in RP associated with scleroderma.

Treatment for RP must include control of excessive vasoreactivity, modification of structural vascular disease, and prevention of microthrombotic events.

Raynaud’s phenomenon (RP) is an exaggerated vascular response of the digital arterial circulation triggered by cold ambient temperature and emotional stress. The diagnosis of RP is based on a history of excessive cold sensitivity and recurrent events of sharply demarcated pallor and/or cyanosis of the skin of the digits (Figure 84-2). Blanching reflects digital arterial vasospasm, and cyanosis occurs secondary to the deoxygenation of sluggish venous blood flow. Some skin blushing (redness) may follow owing to reactive hyperemia after regular blood flow has been restored. Raynaud’s phenomenon occurs in 3% to 15% of the general population. It is more common among females (3 to 4:1) and is likely to begin before 20 years of age. During cold exposure (particularly during shifting temperatures and winter months), Raynaud’s attacks increase in frequency and intensity. Raynaud’s phenomenon can be categorized clinically into primary and secondary forms (Figure 84-3). Primary RP occurs when no disease process is associated with recurrent vasospastic events. Distinguishing primary RP from that associated with an underlying disorder is frequently challenging. Young age at onset (<20 years), symmetric manifestations of symptoms, mild to moderate severity, no association with digital ulceration or tissue gangrene, normal nail-fold capillary examination, and a negative antinuclear antibody (ANA) titer are all indicative of primary RP. Secondary RP occurs in a variety of settings, including connective tissues disorders and other rheumatic conditions, occupational trauma (e.g., hypothenar hammer syndrome), the use of certain drugs (e.g., antimigraine agents, ergotamine derivatives, bleomycin), increased blood viscosity, and compressive or obstructive vascular disease (e.g., thoracic outlet syndrome, atherosclerosis, thromboangiitis obliterans).

Nail-fold capillaroscopy is the tool most commonly used at the bedside to distinguish patients with primary RP from those with scleroderma or another rheumatic disease. Maricq and associates first described the abnormal pattern of nail-fold capillary vessels seen in scleroderma. The simplest method of detection is to coat the skin of the nail fold with immersion oil and then view the area using a bifocal dissecting microscope or a hand-held device such as an ophthalmoscope set at 20 or 40 diopters. A computerized nail-fold videocapillaroscopy technique is used in specialty centers. This technique provides enhanced digital images that can assess local blood flow and follow the disease course. Although patients with primary RP may have normal, thin, palisading nail-fold capillary loops (Figure 84-4), in secondary RP, capillary loop dilation/enlargements and dropouts represent the salient features. Patterns of capillary abnormalities appear to correlate with the course of systemic disease manifestations. Capillary dilation (giant capillaries), microhemorrhages, and some disorganization of the capillary network are typical of early disease; dropouts, avascular areas, and signs of neoangiogenesis with bizarre architectural distortion manifest at later stages of scleroderma. Approximately 20% to 30% of patients with RP and abnormal nail-fold capillary changes will develop clinical features of scleroderma, usually within a 2- to 3-year period. Patients presenting with RP, nail-fold capillary changes, and the presence of a scleroderma-related autoantibody have a 70% to 80% chance of developing scleroderma within 2 to 3 years from presentation. Therefore, capillaroscopy represents an important standard tool that can be used to examine patients presenting with RP.

In scleroderma, RP and digital ischemia are the clinical manifestations of both fixed structural vascular disease and abnormal regulation of local vasomotor control. Cold-induced vasoconstriction of peripheral blood vessels normally occurs via sympathetic stimulation. Abnormal thermoregulation is associated with a nonsclerotic vasculopathy characterized by endothelial dysfunction and a fibrotic proliferation, which produces an increase in collagen content of the intima of small and medium vessels and causes loss of vessel flexibility and obliteration of their lumina. Digital pitting with loss of fingertip tissue and small, painful superficial ulcerations are very common and usually are noted secondary to disease affecting the small arteries and arterioles of the skin (Figure 84-5). Large, deep ulceration of the distal portion of the finger is a consequence of larger-vessel (e.g., digital artery) occlusion associated with severe vasospasm. The latter event usually presents as a sharp demarcation of the distal digit with intense, localized pain secondary to ischemia. Failure to reverse these events may lead to loss of the whole digit or limb secondary to deep tissue infarction.

Although a number of methods may be used to objectively score attacks of RP, no test is considered practical or reproducible enough to replace the clinical criteria for diagnosis and management. Patients with an active RP attack...
The primary goal in the management of RP is prevention of digital ischemia through the use of nonpharmacologic and pharmacologic measures. The primary and most important nonpharmacologic therapy for prevention is avoidance of cold ambient temperatures, particularly transitioning from a warm or hot environment to a cold one. A warm ambient temperature reduces the frequency and severity of RP. All patients with RP should understand that clothing should be layered and loose-fitting, with the goal of maintaining a warm core body temperature, not just warmth of the affected extremities. Patients who have an acute ischemic event are best treated with rest in a warm environment (home or hospital), insulated from cold temperatures. Other potential therapies include minimizing emotional distress (reducing sympathetic tone) and avoiding aggravating factors such as smoking, sympathomimetic drugs (e.g., preparations for the common cold), migraine headache medications (e.g., serotonin agonists), and nonselective β-blockers (e.g., propranolol). Although behavioral therapies (biofeedback, autogenic training, and classical conditioning) are reported to be helpful, their benefit is controversial, and they play no role in the management of acute ischemia related to scleroderma. In fact, biofeedback alone is not helpful for primary RP.

Drug therapy for RP in scleroderma includes the use of a variety of oral or systemic vasodilators. Calcium channel blockers are considered to be first-line therapeutic agents in the treatment of RP. This class of medication works primarily by inducing arterial vasodilation through direct...
inhibition of contraction of vessel smooth muscle cells; however, these agents provide additional benefits by reducing oxidative and inhibiting platelet activation.\(^{44,45}\) Most published studies evaluating the efficacy of calcium channel blockers in RP have employed dihydropyridines (e.g., nifedipine, amlodipine, isradipine, felodipine), and few have looked at nondihydropyridines, which appear to be more efficacious in primary than in secondary RP.\(^ {46,47}\) A meta-analysis of 17 studies evaluating the efficacy of short- and long-acting formulations of calcium channel blockers reported a 33% reduction in attack severity and a near 50% reduction in the number of attacks per week.\(^ {48}\) The current recommendation is to use an extended-release formulation of a calcium channel blocker for treatment of nonurgent RP. The dose should be titrated to clinical efficacy, but common side effects (hypotension, headache, pedal edema) should be monitored. For urgent cases of RP, a short-acting formulation of the medication is preferred, but titration should be carefully monitored. Although calcium channel blockers are the agents most likely to be effective, a host of other vasodilators are used, including nitrates, phosphodiesterase-5 inhibitors (e.g., sildenafil), intravenous prostaglandins, and sympatholytic agents (e.g., prazosin).

Among other agents being tested in the treatment of RP are drugs that enhance nitric oxide availability—serotonin uptake blockers (e.g., fluoxetine), Rho-kinase inhibitors, local injection of botulinum toxin (Botox), antioxidants (acetylcysteine), angiotensin receptor blockers, and selective \(\alpha_2\)-adrenergic receptor antagonists. A combination of these agents, if tolerated, can be used in refractory cases. Various formulations of topical and oral nitroglycerin are employed.\(^ {49,50}\) Although some degree of efficacy is obtained, use of nitroglycerin is limited by substantial side effects, including headache, dizziness, and local skin irritation. The cyclic guanine monophosphate phosphodiesterase-5 inhibitors are thought to be effective in the treatment of RP by sustaining the vasodilatory effect of nitric oxide.\(^ {51,52}\) Studies using phosphodiesterase inhibitors have reported variable effects, and the ideal drug and dosing have not yet been defined. Despite the theoretical advantage of angiotensin-converting enzyme (ACE) inhibitors, a study of quinapril showed that these medications did not affect the occurrence of digital ulcers or the frequency or severity of RP episodes.\(^ {53}\) Thus, although ACEs are critical for the treatment of a scleroderma renal crisis (see renal section), they do not have a major role in the management of RP.

New emphasis is being placed on prevention of scleroderma vascular disease through the use of immunosuppressive and vasoprotective drugs. HMG-CoA (3-hydroxy-3-methylglutaryl-coenzyme A) reductase inhibitors (statins) can modify progression of vascular injury and can prevent vascular ischemia through different pleiotropic functions, as by improvement of endothelial dysfunction, reduction of clotting, and introduction of some anti-inflammatory effects.\(^ {54,55}\) Data suggest that statins may increase the number of endothelial progenitor cells and may improve vascular remodeling after injury.\(^ {56,57}\) In a randomized, placebo-controlled study of scleroderma patients with secondary RP,
Figure 84-5 Scleroderma and Raynaud’s phenomenon can be associated with digital ulcerations and severe digital ischemia. A. Traumatic ulcers over the proximal interphalangeal joints. B and C. Ischemic digital ulcerations secondary to small arterial disease. D. Digital gangrene secondary to macrovascular disease.

Figure 84-6 Approach to drug treatment of Raynaud’s phenomenon (RP) and acute digital ischemia. ACE, angiotensin-converting enzyme; IV, intravenous; PDE-5, phosphodiesterase-5; SSRI, selective serotonin reuptake inhibitor.
SKIN INVOLVEMENT

The most overt clinical manifestation of scleroderma is skin disease. The degree of involvement can vary among patients, and involvement can change in severity and distribution over time in the same individual. Almost every scleroderma patient presents with skin thickening and hardening due to increased collagen and extracellular matrix deposition in the dermis. The distribution of skin changes is characteristic, with more frequent and intense involvement of fingers, hands, forearms, distal legs, feet, and face, and, to a lesser degree, the proximal limbs and anterior trunk. Sparing of the midback is typical. Scleroderma is classically subdivided into limited and diffuse cutaneous forms (Figure 84-7). Limited scleroderma is defined by skin thickening restricted to the face and limbs distally to the elbows and knees. Commonly, in this form of the disease, only the fingers and the face are involved. In contrast, diffuse cutaneous involvement is characterized by widespread skin thickening, including proximal limbs and truncal areas. Proximal skin involvement defines the diffuse cutaneous subset but may be absent in early stages of disease. Signs of skin thickening are not apparent in the setting of other disease manifestations (e.g., Raynaud’s phenomenon, nail-fold capillary abnormalities), nor is evidence of internal organ involvement. This clinical presentation is referred to as systemic sclerosis sine scleroderma. Traditionally, patients are clustered into limited and diffuse skin subsets, but evidence suggests the existence of an intermediate group of patients. Each subset of disease defined by the degree of skin thickening has a unique pattern of disease manifestation and risk for specific clinical outcomes. Therefore, expression of skin disease can be used as a predictor or a “clinical biomarker” of the disease course. The variable degree of skin involvement can be quantified by physical examination. The most widely accepted scoring system (modified Rodnan skin score) is applied by pinching the skin in 17 different body areas and subjectively averaging the thickness of each specific site from 0 = normal to 3 = very thick (Figure 84-8). The skin score (maximum, 51) is a useful clinical measurement tool that can be used to quantify the severity of skin disease. Although the overall extent of skin involvement as measured by the skin score can predict certain clinical outcomes, it does not define the quality of the skin process nor the level of disease activity at any single point in time. Therefore, it is important to follow the skin score over time and to measure changes sequentially to monitor disease progression.

Cutaneous involvement in scleroderma begins with clinical signs of inflammation. This is called the edematous phase because it is characterized by nonpitting edema of affected body areas. In limited scleroderma, this event is mild and is restricted to the digits; however, in the diffuse form of the disease, cutaneous swelling and edema can be widespread and so impressive in the limbs as to mimic a fluid overload state such as congestive heart failure. Edema can also cause local tissue compression. For example, upon involvement of the wrist area, scleroderma patients are not infrequently diagnosed with carpal tunnel syndrome (especially at disease onset) to explain hand and wrist discomfort. In association with edema, signs and symptoms of inflammation are

statin use significantly decreased the number of digital ulcers formed. Intravenous infusions of vasodilating prostaglandins (alprostadil, epoprostenol, iloprost, treprostinil) reduce the severity and frequency of Raynaud’s attacks and are most helpful during periods of sustained critical ischemia. These prostaglandins are efficacious in the treatment of RP because of their strong vasodilating effect, inhibition of platelet aggregation, and enhancement of vascular function through other mechanisms. Intravenous prostacyclins can be given intermittently to patients with scleroderma during winter months or throughout the year, or they can be used acutely during a vascular crisis. Oral prostaglandins have variable absorption and are not yet available.

Endothelin-1 is strongly implicated in the pathogenesis of RP and the vascular disease of scleroderma. Bosentan, an endothelin-1 receptor blocker, was studied in a randomized, placebo-controlled, double-blind multicenter study that showed efficacy in preventing new digital ulcers, along with improved hand function, but it did not improve healing of existing ulcers compared with placebo.

Sympathectomy is a viable option for patients with RP who are unresponsive to medical therapy; it should be considered as acute intervention during a critical ischemic event. Localized digital sympathectomy with lysis of fibrosis around the vessel is effective for acute ischemia and has mostly replaced cervical sympathectomy. Responses to sympathectomies suggest that the long-term benefit derived from these interventions is limited. In fact, Raynaud’s attacks eventually recur, and medical therapy is needed to prevent new events.

Careful assessment for correctable macrovascular disease should be done when confronted with acute digital ischemia, particularly when the whole finger is demarcated, or when the event involves a lower extremity. In this setting, appropriate studies such as arterial Doppler ultrasound or angiographic imaging are warranted. If macrovascular disease is present, vascular surgery may help to alleviate the occlusive process. Preferential involvement of the ulnar artery has been reported in scleroderma patients.

Patients with critical digital ischemia should be hospitalized to reduce vasospastic activity, maintain warmth, and permit rapid initiation of vasodilator therapy. Intravenous vasodilating prostaglandins can be infused to maximize vasodilation. Antiplatelet therapy with low-dose aspirin may be useful, but its benefit is unproven in the acute setting. Heparinization may be considered during acute ischemic crises of the digits, but chronic anticoagulation in association with edema, signs and symptoms of inflammation are...
common. Erythema of the skin and intense pruritus and pain (see Figure 84-7C) are characteristic of advancing active diffuse skin disease. This pain has a neuropathic quality with a reported “pins and needles” sensation. The disease process leads to loss of skin appendages, as well as decreased hair growth and loss of sweat and exocrine glands; thus the skin surface becomes dry and uncomfortable. Small papules can be seen in areas of trauma as the result of scratching, giving the surface a cobblestone texture. The edematous phase continues for several weeks but eventually gives way to a fibrotic stage, with protracted activity that may last months or years.

During the fibrotic phase, acute inflammation is clinically less obvious, and deposition in the dermis of excessive collagen and other extracellular material thickens the skin, making it inflexible and causing further loss of skin appendages. Fibrosis extends beyond the dermis into the deeper layers with loss of subcutaneous adipose tissue

Figure 84-7 Skin involvement in scleroderma is subdivided into “limited” and “diffuse” cutaneous forms. A, Sclerodactyly in limited cutaneous disease. B, Truncal changes in diffuse cutaneous disease. C, Inflammatory signs in early active skin disease. D, Finger contracture in the chronic fibrotic phase of skin involvement in scleroderma.

Figure 84-8 Method used to semi-quantify skin thickness in scleroderma. The modified Rodnan skin score is obtained by clinical palpation of 17 different body areas (fingers, hands, forearms, upper arms, chest, abdomen, thighs, lower legs, and feet) and subjective averaging of the thickness of each specific site: 0 = normal (A); 1 = mild (B); 2 = moderate (C); and 3 = severe (D). The maximum score is 51.
Telangiectasias are erythematous matted skin lesions of vascular origin; for this reason, they blanch on local pressure. They are composed of vasodilated postcapillary venules without evidence of inflammation and resemble the type of lesions seen in Osler-Weber-Rendu disease (hereditary hemorrhagic telangiectasia). Telangiectasias develop primarily on the fingers, hands, face, and mucous membranes, but they may also be found on the limbs and trunk (Figure 84-9B). They tend to become more numerous over time in both limited and diffuse types of skin disease, and are more obvious in white patients with limited scleroderma. The biologic mechanism leading to the development of telangiectasias in scleroderma is thought to be related to the underlying chronic tissue hypoxia that stimulates abnormal secretion of vascular growth factors (e.g., vascular endothelial growth factor [VEGF]). Thus, the development of telangiectasias may indicate ongoing vascular injury and abnormal vascular repair or angiogenesis. The total burden of telangiectasias is associated with increased risk of pulmonary hypertension, suggesting that they are clinical biomarkers of systemic vascular disease. Nail-fold capillary abnormalities can be observed after immersion oil is applied to the skin surface at the base of the fingernails, and direct visualization is performed using an ophthalmoscope or by videocapillaroscopy. In early scleroderma, dilated capillary loops (giant capillaries) and microhemorrhages may be seen. At later stages, the nail-fold capillaries are attenuated and disorganized (see section on vascular changes).

Despite evidence of active inflammation during the early edematous phase of scleroderma skin involvement, systemic corticosteroids do not appear to be effective in stopping disease progression. A variety of other immunomodulatory drugs and newer agents with potential “disease-modifying properties” (i.e., antifibrotic) have been used to control the skin disease, but to date, none has proved uniformly successful (see treatment section). In the early active stage of diffuse scleroderma, pruritus can be one of the most distressing symptoms. Antihistamines, analgesics, or tricyclic antidepressants (e.g., doxepin) are often used but usually provide only partial benefit. Dryness of the skin surface results from damage to the exocrine structures caused by decreased or absent natural oil (sebum) production. This can worsen...
pruritus, leading to skin trauma due to repetitive scratching. Ulceration and secondary tissue infection may also result. The best approach to treatment is characterized by frequent topical application of an emollient preparation, periodic cleansing with soap and water, and use of topical antibiotics for any traumatic skin ulceration. Physical therapy is most important to prevent severe skin and joint contractures, and to help patients with activities of daily living.

**GASTROINTESTINAL INVOLVEMENT**

**KEY POINTS**

**Manifestations of gut dysmotility are universally present in scleroderma and can affect any segment of the gastrointestinal tract.**

**Involvement of the upper GI tract is more common and can present with severe symptoms.**

**Dysfunction and failure of the lower GI tract are associated with poor prognosis.**

Almost every scleroderma patient has symptoms of gastrointestinal disease, ranging from mild gastroesophageal reflux disease (GERD) to severe bowel dysfunction, which can be life threatening. Virtually any segment of the gastrointestinal tract can be affected (Table 84-4).

**Oropharynx**

Patients may report that chewing is difficult because of decreased facial flexibility caused by skin and deeper tissue fibrosis. Perioral skin tightening can result in a decreased oral aperture and inability to open the mouth wide enough for proper dental care or to bite into large solid foods such as an apple (Figure 84-10). Dry membranes resulting from decreased saliva production can lead to difficult mastication. In some patients, periodontal disease and gum regression cause loosening of teeth, which further affects chewing capacity.

Upper pharyngeal function is usually normal, but a subset of patients may develop a myopathy that causes weakness of pharyngeal muscles, mimicking a neuromuscular disease. This manifestation is characterized by pharyngeal dysfunction with problems initiating swallowing and frequent coughing due to laryngeal penetration and can be associated with increased risk for aspiration of foods or liquids. Although limited pathologic studies of the upper pharyngeal structures have examined scleroderma, both myositis and fibrosis are known to occur.

**Esophagus**

Dysphagia resulting from esophageal disease is the most common gastrointestinal symptom, occurring in approximately 90% of patients. Heartburn, regurgitation, and dysphagia for pills and solids (more than liquids) are the most common symptoms, but atypical retrosternal pain (particularly at night), periodic cough, a sense of food “sticking” in the esophagus, and nausea can result from esophageal dysfunction. Esophageal involvement in scleroderma is characterized by loss of normal smooth muscle function, especially in the lower two-thirds of the esophagus, and hypotonia of the lower esophageal sphincter (Figure 84-11A). Functional studies of esophageal motility suggest that neural dysfunction is common in patients with scleroderma, and that this may precede myopathic dysfunction and histologic changes in smooth muscle layers. Manometric evaluation has identified the presence of esophageal hypomotility in areas that later did not show any histologic

<table>
<thead>
<tr>
<th>Site</th>
<th>Manifestation</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oropharynx</td>
<td>Perioral tight skin</td>
<td>Regular dental and periodontal care</td>
</tr>
<tr>
<td></td>
<td>Decreased oral aperture</td>
<td>Artificial saliva</td>
</tr>
<tr>
<td></td>
<td>Periostitis, gum disease</td>
<td>Targeted swallowing exercises and rehabilitation</td>
</tr>
<tr>
<td></td>
<td>Dry mouth</td>
<td>Lifestyle modifications</td>
</tr>
<tr>
<td></td>
<td>Swallowing difficulties</td>
<td>Proton pump inhibitors</td>
</tr>
<tr>
<td></td>
<td>Coughing, aspiration</td>
<td>Prokinetics</td>
</tr>
<tr>
<td></td>
<td>Acid reflux (heartburn)</td>
<td>Upper gastrointestinal endoscopy</td>
</tr>
<tr>
<td>Esophagus</td>
<td>Dysphagia</td>
<td>Prokinetics</td>
</tr>
<tr>
<td></td>
<td>Strictures</td>
<td>Proton pump inhibitors, iron replacement</td>
</tr>
<tr>
<td></td>
<td>Barrett’s esophagus</td>
<td>Endoscopic laser or cryotherapy</td>
</tr>
<tr>
<td>Stomach</td>
<td>Gastroparesis, dyspepsia</td>
<td>Transfusions</td>
</tr>
<tr>
<td></td>
<td>Gastric antral vascular ectasia</td>
<td>Surgery</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mild laxatives</td>
</tr>
<tr>
<td>Small and large intestine</td>
<td>Hypomotility, constipation</td>
<td>Promotility agents</td>
</tr>
<tr>
<td></td>
<td>Bacterial overgrowth, diarrhea</td>
<td>Rotational antibiotics</td>
</tr>
<tr>
<td></td>
<td>Pseudo-obstruction</td>
<td>Octreotide</td>
</tr>
<tr>
<td></td>
<td>Pneumatosis intestinalis</td>
<td>Avoidance of surgery</td>
</tr>
<tr>
<td></td>
<td>Malabsorption</td>
<td>Enteral or parenteral nutrition support</td>
</tr>
<tr>
<td>Anorectum</td>
<td>Colonic pseudodiverticula</td>
<td>Biofeedback, sacral nerve stimulation, surgery</td>
</tr>
</tbody>
</table>
abnormality of the smooth muscle. Other studies have shown esophageal smooth muscle activity in response to pharmacologic stimulation (methacholine challenge) early in the disease, but not in patients with more advanced scleroderma. The cause of these abnormalities is unknown, but clear evidence indicates that neural dysfunction precedes muscle disease.

Tissue fibrosis is often evoked as the cause of esophageal disease in scleroderma. However, pathologic studies demonstrate atrophy of the smooth muscle layers of the distal esophagus in the absence of significant fibrosis. Ischemia of the esophagus in scleroderma is suggested by functional studies that show impaired esophageal blood flow following cold exposure and rewarming protocols. Noninflammatory intimal layer hyperplasia in arterioles of the gastrointestinal tract has been described. Inflammatory infiltrates usually are not present in the smooth muscle unless severe transmural esophagitis is present. This suggests that inflammation is not a cardinal feature of the pathogenesis of esophageal smooth muscle lesions in scleroderma.

The severity of esophageal symptoms may not accurately reflect the seriousness of the underlying disease. Therefore, every patient with scleroderma should be fully evaluated for esophageal involvement. If untreated, gastrointestinal reflux may lead to esophagitis, bleeding, esophageal strictures, or precancerous lesions such as Barrett’s esophagus. The asymptomatic patient without abnormal laboratory testing (anemia, positive evidence of gastrointestinal bleeding) does not require specific interventions. Patients with mild GERD can be treated empirically with proton pump inhibitors and then can be followed clinically if they become symptom free with normal laboratory data. When more severe symptoms are present (e.g., GERD with dysphagia), or when treatment with proton pump inhibitors fails, an upper gastrointestinal endoscopy is necessary to fully assess the anatomy of the esophagus and the extent of related disease. Direct endoscopy is appropriate to rule out Barrett’s esophagus, stricture, or a site of uncontrolled esophagitis or bleeding. Barium swallow and cine-esophagography are sensitive tests for esophageal strictures. However, direct measurement of esophageal motility via esophageal manometry may be needed if the cause of symptoms such as atypical chest pain is unclear. Although complications from Barrett’s esophagus (e.g., esophageal cancer) are uncommon, periodic re-endoscopy to reassess status is recommended, even if symptoms are under therapeutic control.

It is critical that patients alter eating behavior to complement medication therapy. Patients often do better by eating more frequent smaller meals, avoiding food intake for several hours before bedtime, moving the main meal toward mid-day, taking a walk after meals to help gastroesophageal emptying, and eliminating foods that aggravate symptoms (e.g., spicy sauces, caffeinated or carbonated beverages). Bedtime is often the time of severe reflux. This can be
improved by avoiding filling the stomach at least 2 to 3 hours before bedtime and by elevating the head and trunk during sleep. Treatment of reflux by suppression of gastric acid with antacids or histamine-2 (H$_2$) blockers is helpful but overall disappointing in scleroderma. Conversely, proton pump inhibitors (e.g., omeprazole, lansoprazole) can be very effective. Usually, long-term daily use is required. Higher doses may be used periodically for breakthrough symptoms. If symptoms are not controlled with recommended medication dosages, then a formal 24-hour ambulatory pH study may be necessary to determine whether persistent symptoms are due to uncontrolled acid reflux. A prokinetic drug (e.g., metoclopramide, domperidone, erythromycin) should be added when symptoms of dysphagia or endoscopic findings of esophagitis are present despite the use of effective acid suppression. These medications tend to be effective during early disease but are less likely to help later on, with advanced esophageal dysfunction. Long-term use of prokinetics may be needed. Many patients respond to relatively low doses, for example, at bedtime alone. Given the increased risk of neurologic complications with the use of metoclopramide, many recommend domperidone as a relatively safer alternative when long-term prokinetic treatment is needed.

Figure 84-11 Gastrointestinal manifestations in scleroderma. A, Chest computed tomography (CT) (sagittal view) showing severe esophageal dysmotility with dilation and (arrow) retention of gastric content. B, Upper endoscopy: gastric antral vascular ectasias presenting as “watermelon” stomach. C and D, Plain abdominal x-ray and abdominal CT: small intestinal dysmotility with pseudo-obstruction, pneumatosis cystoides intestinalis.
Stomach

The degree of delayed gastric emptying (gastroparesis) is underappreciated in scleroderma, and gastroparesis often causes early satiety, aggravation of GERD symptoms, anorexia, abdominal pain, a sensation of bloating, or nausea. Frequently, weight loss in scleroderma patients is a consequence of poor caloric intake due to lack of appetite from poor gastric function. Prokinetics are used to improve gastric emptying and related symptoms. Gastritis or gastric ulcer can occur. Dilation of the microvasculature in the gastric mucosa is found in a subset of patients. This manifestation, also called gastric antral vascular ectasia (GAVE), is thought to be caused by an abnormal angiogenesis, leading to bizarre dilation of microvessels and arterial-venous (A-V) malformations similar to the telangiectasias seen in the skin, lips, and oral mucous membranes. In general, these lesions are asymptomatic, but they can be responsible for occult gastrointestinal bleeding. Extensive clusters of A-V malformations lead to the presence of longitudinal red stripes in the inner lining of the stomach, converging to the pylorus and described on endoscopy as “watermelon stomach,” based on their appearance (Figure 84-11B). Endoscopic therapy with laser photocoagulation or cryotherapy can be used to ablate bleeding vessels. Rarely, bleeding associated with GAVE is refractory and requires multiple transfusions, repeated laser or cryotherapy, or even gastric surgery (e.g., antral or gastric resection).

Lower Gastrointestinal Tract

Bloating, abdominal distention, diarrhea, and constipation are common complaints caused by dysmotility of the small and large bowel. Functional and pathologic studies again suggest that microvascular disease, neural dysfunction, smooth muscle atrophy, and tissue fibrosis similar to those seen in the esophagus are causing the bowel disease. Most cases present in similar fashion to irritable bowel syndrome (IBS). Constipation due to sluggish or atonic bowel function can result in repeated bouts of diarrhea with malabsorption, progressive loss of weight, and severe malnutrition. Diarrhea is thought to be caused by bacterial overgrowth as a consequence of bowel dysfunction, but other causes have been considered, including decreased mesenteric blood flow or pancreatic insufficiency. Recurrent episodes of pseudo-obstruction constitute one of the most serious intestinal problems in scleroderma. These episodes are sometimes mistaken for surgical emergencies. Pseudo-obstruction is a manifestation of profound loss of bowel smooth muscle function, causing severe dysmotility and segments of luminal dilation. Pneumatosis cystoides intestinalis sometimes complicates scleroderma of the bowel when gas leaks into the diseased intestinal wall and tracks into the mesentery of the gut or the peritoneal cavity, mimicking a bowel perforation (Figure 84-11C and D). Asymptomatic wide-mouthed diverticula, also resulting from fibrosis and atrophy of the bowel wall, are pathognomonic of scleroderma. Volvulus, strictures, and perforation are uncommon complications of severe bowel involvement.

Management of the lower gastrointestinal tract includes avoiding a constipation-diarrhea cycle through adequate fiber ingestion, use of stool softener (docusate), or, if constipation is severe, periodic doses of osmotic laxatives (polyethylene glycol). Lubiprostone is helpful if chronic constipation predominates. Other prokinetic drugs are less effective for treatment of the lower gastrointestinal tract. Octreotide is reported to help patients with severe lower GI hypomotility and pseudo-obstruction. In cases with bloating, recurrent bouts of diarrhea, or episodes of pseudo-obstruction, the use of cyclic antibiotics and/or probiotics is helpful. Total parenteral nutrition may become necessary in patients with severe scleroderma-related bowel disease without response to other medical therapy. The presence of intestinal failure is a poor prognostic complication. Fecal incontinence can result in scleroderma caused by dysfunction of both internal (atrophy and fibrosis) and external (weakness) anal sphincters. Successful treatment includes firming stool texture, performing exercises to strengthen the pelvic muscles, providing biofeedback methods, and conducting surgical repair of aggravating factors such as rectal prolapse or severe hemorrhoids.

PULMONARY INVOLVEMENT

Pulmonary involvement is found in most scleroderma patients. Interstitial lung disease and pulmonary hypertension are recognized as the most common lung complications and now are regarded as the major cause of death in scleroderma. In combination, it is estimated that they are responsible for 60% of scleroderma-related deaths.18 They are present in many patients at the same time, but one process may dominate over the other. Lung disease can occur without symptoms, or it can cause progressive respiratory failure and severe limitation of the quality of life. Additional pulmonary complications such as chronic aspiration, pleural disease, spontaneous pneumothorax, neuromuscular weakness, drug-induced pneumonitis, and lung cancer need to be considered in the differential diagnosis of a scleroderma patient presenting with respiratory symptoms.

Interstitial Lung Disease

<table>
<thead>
<tr>
<th>KEY POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung disease is a major cause of morbidity and mortality in scleroderma patients.</td>
</tr>
<tr>
<td>Pulmonary fibrosis occurs in both limited and diffuse subsets of scleroderma, with a variable disease course in terms of severity and outcome.</td>
</tr>
<tr>
<td>Nonwhite and antitopoisomerase 1–positive patients generally have the worst prognosis.</td>
</tr>
<tr>
<td>Pulmonary function testing (spirometry and diffusing capacity) is helpful for screening and monitoring of interstitial lung disease (ILD). The degree of lung fibrosis on high-resolution computed tomography (HRCT) predicts outcome.</td>
</tr>
<tr>
<td>Treatment for scleroderma-related ILD is limited to immunosuppression.</td>
</tr>
<tr>
<td>Risk factors for pulmonary arterial hypertension include late onset of scleroderma, limited phenotype, presence of numerous telangiectasias, and positive anticientromere antibodies.</td>
</tr>
</tbody>
</table>
Interstitial lung disease (ILD) is the most common lung manifestation, occurring to some degree in about 80% of patients with diffuse cutaneous scleroderma and in 20% of patients with limited skin disease. Higher risk to develop severe progressive ILD has been observed in patients with diffuse skin involvement, African-Americans, Native Americans, and those positive for antitopoisoerase 1 (Scl-70), anti-U3-RNP, or anti-Th/To antibodies. Data from a recent study of patients with active lung disease show that patients with limited and those with diffuse scleroderma were indistinguishable with regard to their baseline pulmonary functions, although patients with limited skin disease presented with more extensive pulmonary fibrosis, perhaps reflecting a delay in diagnosis and more advanced progression of lung disease before study entry.

ILD typically manifests with declining lung volumes and increased parenchymal fibrosis with reticular interstitial thickening that is greatest at the lung bases. Pathologic studies have shown that the most common histologic pattern of fibrosing alveolitis in scleroderma is nonspecific interstitial pneumonia (NSIP), as opposed to usual interstitial pneumonia (UIP), which is instead the common presentation of idiopathic pulmonary fibrosis. A mixed pattern of NSIP and UIP can also be found. The histopathologic classification does not appear to predict outcome, which is more related to baseline disease severity and functional respiratory measures. For this reason, a surgical lung biopsy is not required in the setting of scleroderma patients who are following a typical disease course. Although ILD is usually characterized by a functional restrictive ventilatory defect with reduced gas exchange, at later stages of the disease some degree of airway obstruction can be found in association with honeycombing of lung parenchyma and bronchiectasis. Pleural reactions are not common in scleroderma. Rare cases of spontaneous pneumothorax, adult respiratory distress syndrome, and pulmonary hemorrhage have been reported. Aspiration secondary to gastroesophageal reflux disease (GERD), secondary infection, and heart failure can complicate the course of scleroderma lung disease. Some authors consider GERD associated with chronic aspiration a potential inciting factor for the development of ILD in scleroderma.

During earlier stages of ILD, the underlying active fibrosing alveolitis may be totally asymptomatic and may not be detected by conventional chest radiographs. The most common symptoms of scleroderma lung disease are dyspnea (initially on exertion) and fatigue. Chest pain is not typical, and nonproductive cough is usually a late complication, particularly in association with the presence of traction bronchiectasis. Coughing in scleroderma often is not a primary lung problem but rather is a manifestation of GERD with associated laryngeal irritation. The most characteristic finding of ILD on physical examination is bilateral fine inspiratory crackles (i.e., “velcro” rales) at the lung bases. Pulmonary function testing (PFT) is the method most commonly used to detect interstitial lung disease (Figure 84-12).

*Figure 84-12 Approach to lung disease in scleroderma. This approach is recommended for all patients with scleroderma with both new-onset and long-standing disease. BAL, bronchoalveolar lavage; CT, computed tomography; DLCO, diffusing capacity; ECHO, echocardiography; FVC, forced vital capacity; PAH, pulmonary arterial hypertension; PFT, pulmonary function test; TLC, total lung capacity.*
(DLCO). Although a decline in DLCO correlates with the severity of pulmonary fibrosis as measured by HRCT and predicts worse outcome, it is not a very specific measure and can be abnormal also in the context of pulmonary vascular disease and underlying chronic obstructive disease/emphysema. A restrictive ventilatory defect is identified by a depression of the forced vital capacity (FVC) and the total lung capacity (TLC). Serial measurements of FVC are the most reliable tool with which to follow ILD. A decline in FVC of more than 10% from baseline is often indicative of disease activity and is associated with increased mortality. Lung function may remain stable after an initial decline at disease onset, or it can progress at 4- to 6-month intervals until severe disease develops after a few years. Patients presenting with a normal FVC during early stages of scleroderma have a better lung prognosis, and an early reduction in FVC is likely to be associated with further progression of lung involvement, usually within the first 2 to 3 years from disease onset. The presence of ILD predicts increased mortality. One study found that patients with severe ILD have a 9-year survival rate of approximately 30%, whereas patients with no severe organ involvement have a 72% survival rate. The most rapid decline in FVC is thought to occur within the initial 3 years from disease onset, but progression of ILD can occur in later years, so it is recommended that respiratory status be monitored on a regular basis, and that pulmonary function be tested yearly.

High-resolution computed tomography (HRCT) is more accurate than chest radiography in evaluating diffuse lung disease; it is a well-established, sensitive, and noninvasive method of detecting and characterizing ILD in scleroderma. The extent of pulmonary fibrosis on HRCT correlates closely with PFT abnormalities and can be used to predict prognosis. Recent evidence shows that the degree of HRCT changes coupled with the level of lung function deficit can be used to generate a simple staging system (limited vs. extensive disease) with good prognostic value. An FVC of less than 70% or a disease extent on HRCT greater than 20% predicts ILD progression and higher mortality. HRCT features of fibrosis are present in 55% to 65% of all patients with scleroderma and in almost every patient with abnormal pulmonary function tests. The earliest and most common abnormality on HRCT is usually found in the posterior and basilar portion of the lungs in the form of increased subpleural lung attenuations in the absence of distinct architectural distortion. With progression of the disease, increased “ground glass” opacities (GGO) are seen, but it has been well established that these changes alone are not always indicative of active inflammation, nor can they predict progression itself. Subsequent detection of a fine reticular pattern often precedes overt pulmonary fibrosis manifested by parenchymal distortion, reticular interlobular interstitial thickening, traction bronchiectasis and bronchiolectasis, and, in the late stages, honeycomb and cystic air spaces (Figure 84-13). When reticular interstitial abnormalities are present on HRCT, regression of disease rarely, if ever, occurs.

Pathologic studies on scleroderma patients with early lung disease have shown that lung fibrosis is preceded by the presence of a mixed interstitial inflammatory infiltrate spilling into the alveolar spaces (alveolitis). It has been assumed that measuring abnormal levels of inflammatory cells in the bronchoalveolar lavage (BAL) would have allowed identification of patients with active lung disease at risk for progression and with potential benefit from immunosuppressive therapy. Published data have shown that an abnormal BAL cell count, in particular with increased granulocytes (neutrophils and eosinophils), is associated with more severe lung disease and mortality in patients with scleroderma ILD. However, the value of BAL cytology in predicting lung disease progression or response to treatment has not been uniformly confirmed. Few observational studies have indicated that the presence of BAL granulocytic alveolitis was associated with significant deterioration of lung function tests over time in patients not receiving immunosuppressive therapy. In contrast, other studies did not find a clear association between BAL cellular abnormalities and clinical outcomes. At present, evidence to recommend routine clinical use of BAL cytology to predict outcome in patients with scleroderma and ILD is still insufficient.

Early detection and prompt therapeutic intervention remain essential in preventing progression of lung disease. Current treatment of scleroderma-related ILD is most often directed against the immune response and the inflammatory process mediating lung injury and tissue fibrosis. A randomized placebo-controlled study comparing 1 year of oral cyclophosphamide versus placebo suggests some modest benefit of the active drug (improvement in FVC). This benefit was sustained at 18 months but was completely lost after 2 years' follow-up, suggesting that a sequential or maintenance immunosuppressive approach may be indicated to retain the clinical response. Other studies suggest benefit using intravenous monthly cyclophosphamide followed by azathioprine as maintenance therapy. Several small retrospective studies have shown moderate benefit from mycophenolate mofetil (MMF) with improvement in or stabilization of lung function, suggesting that this drug may prove to be a safe and effective treatment in scleroderma patients, and a possible alternative to cyclophosphamide. Although more studies are needed to confirm the benefit of immunosuppressive therapy in scleroderma-associated ILD, newer drugs are under investigation, including targeted biologic immunotherapies and antibacterial molecules (see general therapy section). More aggressive immunosuppressive protocols such as autologous or allogeneic hematopoietic stem cell transplantation, preceded by myeloablative or immunoablative conditioning regimens, are under investigation, but their treatment toxicity remains a concern.

It remains very important to pursue careful selection of patients and to treat only those with evidence of active lung disease. In fact, in most cases, no or minimal progression to severe disease is seen despite some evidence of underlying pulmonary fibrosis. This is best done with serial studies and careful assessment of both pulmonary function tests and HRCT scans.

Lung transplantation is an option in patients with severe ILD not responsive to medical therapy. Carefully selected scleroderma patients undergoing lung transplantation have similar morbidity and mortality as patients undergoing transplantation for idiopathic lung disease. A retrospective survey of 47 scleroderma patients who underwent lung transplantation found that 1- and 3-year survival rates were 68% and 46%, respectively.
Figure 84-13  Scleroderma-related interstitial lung disease: high-resolution chest computed tomography scan showing (A) normal lung, (B) active alveolar inflammation (“ground glass” opacification), and (C) end-stage lung disease with honeycombing. D, Gross pathology. E, Histology showing fibrosing alveolitis.

**Pulmonary Hypertension**

Pulmonary arterial vascular disease is a common manifestation of scleroderma; when associated with pulmonary arterial hypertension (PAH), it represents a difficult clinical challenge with life-threatening consequences. The pulmonary vascular process can be indolent and remain clinically silent, or it may lead to respiratory distress secondary to severe PAH and associated right heart failure. Typical symptoms associated with clinically manifested PAH include dyspnea on exertion, fatigue, and, less commonly, chest pain or syncope. Physical examination may be normal during early stages of PAH, but as the disease progresses, a systolic murmur of tricuspid regurgitation, a loud pulmonic component, the S2, an S3 gallop, and signs of right heart failure (right-sided parasternal heave, prominent and elevated jugular venous pulse, hepatomegaly, signs of fluid overload with peripheral edema) are seen. Later in the disease, patients become dyspneic with little activity, have a resting tachycardia, and may appear cyanotic. Sudden syncope or death can occur owing to hypoxia and congestive heart failure.

PAH is usually a late complication of scleroderma, typically manifesting after 10 years from disease onset, particularly in patients with limited cutaneous involvement. Considering the potentially devastating consequences of PAH, early diagnosis is very important. Simple bedside examination is not very sensitive, particularly during early stages; thus it is recommended to screen every scleroderma patient using objective testing such as electrocardiography...
Chapter 84  |  Clinical Features and Treatment of Scleroderma  

PAH is defined as resting mean pulmonary arterial pressure equal to or greater than 25 mm Hg, with normal pulmonary capillary wedge pressure equal to or less than 15 mm Hg. In addition to confirming abnormal hemodynamic values, the RHC can help in differentiating true PAH from pulmonary venous hypertension secondary to left heart failure or pulmonary occlusive disease. The 6-minute walk test can provide baseline functional information but cannot accurately assess the severity of PAH. Because in scleroderma patients, baseline hemodynamic parameters do not correlate well with the clinical course of PAH, new cardiopulmonary dynamic testing such as exercise stress ECHO and exercise during RHC have been introduced, but their role in diagnosing or monitoring PAH has not yet been fully established.

Based on the definition and the tools used to obtain measurements, the prevalence of PAH in scleroderma has been reported with variability. Using ECHO as a diagnostic tool, the prevalence of PAH has been estimated at between 30% and 50% of all studied patients, and investigations using confirmatory RHC have detected a prevalence of 8% to 12%.

Clinically, scleroderma-related pulmonary vascular disease can present in three different ways. In some patients, the ECG can be normal, but as disease progresses, signs of right-sided heart strain with right ventricular hypertrophy and right axis deviation can be observed. Detection by ECHO of increased right ventricular systolic pressure (RVSP) is useful, but values need to be confirmed by right heart catheterization (RHC), given the possibility of false-positive or false-negative results. When the ECHO estimates an RVSP greater than 45 mm Hg, about 95% of the time this will be confirmed by RHC. Some use the tricuspid gradient (TG) as a measure of pulmonary artery pressure. Additional noninvasive screening includes pulmonary function testing. A low diffusing capacity (DLCO) without evidence of obstructive or restrictive lung disease (low forced vital capacity, or FVC) suggests a defect in gas exchange secondary to pulmonary vascular disease or PAH. DLCO may progressively decline for years before the diagnosis of PAH is established.

Increased serum N-terminal pro-brain natriuretic peptide (NT-proBNP) may indicate the presence of PAH and right heart strain, even when levels are modestly elevated (>395 pg/mL). The diagnosis of PAH should be confirmed by RHC, providing direct measurement of pulmonary hemodynamics. PAH is defined as resting mean pulmonary arterial pressure equal to or greater than 25 mm Hg, with normal pulmonary capillary wedge pressure equal to or less than 15 mm Hg. In addition to confirming abnormal hemodynamic values, the RHC can help in differentiating true PAH from pulmonary venous hypertension secondary to left heart failure or pulmonary occlusive disease. The 6-minute walk test can provide baseline functional information but cannot accurately assess the severity of PAH. Because in scleroderma patients, baseline hemodynamic parameters do not correlate well with the clinical course of PAH, new cardiopulmonary dynamic testing such as exercise stress ECHO and exercise during RHC have been introduced, but their role in diagnosing or monitoring PAH has not yet been fully established.

Based on the definition and the tools used to obtain measurements, the prevalence of PAH in scleroderma has been reported with variability. Using ECHO as a diagnostic tool, the prevalence of PAH has been estimated at between 30% and 50% of all studied patients, and investigations using confirmatory RHC have detected a prevalence of 8% to 12%. Clinically, scleroderma-related pulmonary vascular disease can present in three different ways. In some
patients, it manifests and progresses as severe isolated PAH in the absence of other significant lung disease. In others, it can be associated with mild or severe pulmonary fibrosis that is believed to drive vascular disease and PAH by causing chronic hypoxia coupled with destruction of lung parenchyma and underlying vasculature. Association of ILD and PAH usually entails a worse prognosis. In a third group of patients, pulmonary vascular disease presents with dyspnea on exertion, isolated decline of the DLCO with no significant signs of ILD (i.e., low lung volumes), and a normal resting ECHO and/or RHC. These patients are considered to have microvascular lung disease and may be at high risk to develop overt PAH. Increased risk of developing PAH has been associated with limited skin disease, late age at onset of scleroderma, presence of numerous telangiectasias, low DLCO, and presence of anti-U3-RNP antibody. Some studies suggest that the presence of anticentromere, anti-B23, and anti-β2-glycoprotein I antibodies also increases the risk for PAH.

PAH has a major impact on quality of life and survival of affected patients. The natural course of the disease is characterized by progression of hemodynamic impairment, resulting in right heart failure and leading to a poor outcome. Despite the availability of newer targeted therapies, the median survival in scleroderma patients with PAH ranges between 1 and 3 years.80,81 A progressive rapid increase in pulmonary pressure is more likely to occur in patients with limited skin disease and late age at onset of scleroderma.82 This more aggressive course is associated with high mortality. Survival in patients with ILD-associated PAH is significantly worse (3-year survival, 46%) than in those with PAH alone (3-year survival, 64%).83 This difference is possibly explained by the presence in scleroderma of underlying comorbidities, and in particular of biventricular heart (myocardial) involvement. A 10-fold increase in NT-proBNP blood levels is associated with higher mortality.84,85 ECHO findings can help predict prognosis. For example, the degree of tricuspid annular plane systolic excursion (TAPSE) reflects right ventricular function. A TAPSE of less than 1.8 cm is associated with decreased survival in patients with PAH.86 A recent survey identified that pulmonary vascular resistance, stroke volume index, and pulmonary arterial capacitance were strong predictors of survival.87 This suggests that right ventricular dysfunction is associated with poorer outcome. Male gender and higher New York Heart Association functional class also impact prognosis negatively. It is suggested that survival has improved with current modes of therapy. One survey found 47% 2-year survival in historical controls as opposed to 71% among patients treated with modern therapy.88

All patients with PAH are encouraged to continue an active lifestyle as tolerated. Formal educational and exercise programs can improve their quality of life. Conventional therapy includes diuretic therapy (loop and potassium-sparing diuretics) and supplemental oxygen therapy if needed. Anticoagulation is used if there is no defined risk for bleeding. Although calcium channel blockers may help patients with evidence of acute vasodilative response during hemodynamic testing, this is an uncommon finding in scleroderma. In recent years, a number of targeted agents have become available for the treatment of PAH. Many clinical trials testing these drugs in primary PAH have included about 20% of patients with connective tissue disease, primarily scleroderma. Specific therapies include prostaglandin derivatives (epoprostenol, treprostinil, beraprost, iloprost), endothelin receptor antagonists (bosentan, sitaxsentan, ambrisentan), and phosphodiesterase-5 inhibitors (sildenafil, tadalafil, vardenafil). These interventions can improve hemodynamics, exercise tolerance, and quality of life. No evidence from clinical trials suggests that any one of these agents is superior to another. Most believe that early intervention is most important, and that combination therapy may be helpful (see Figure 84-14). Decisions regarding the preferred agents to use and how to deliver them are based mostly on personal preferences and the clinical status of the individual patient. For patients with mild to moderate disease (functional class I or II), a single oral agent is usually started (i.e., sildenafil, bosentan, ambrisentan). Our preference is to use sildenafil, given its favorable side effect profile. Patients presenting with more severe disease (class III or IV) and those failing monotherapy are usually treated with intravenous or inhaled prostaglandins alone or in combination with one or two oral agents. Sildenafil and bosentan can be used in combination, even if concern arises about their interaction and increased hepatotoxicity.89,90 Lung or heart-lung transplantation for PAH remains the treatment of last resort when medical therapy fails. Patients with scleroderma who are carefully selected for transplantation have an outcome that is similar to those with other causes of lung disease. PAH in scleroderma is a progressive disease for which no specific or single therapy may suffice. Therefore, careful longitudinal monitoring is important. Noninvasive testing with periodic clinical assessment with 6-minute walk time, repeat pulmonary function testing, and ECHO studies are helpful. When a major therapeutic decision is needed, such as switching or adding a new therapy, a repeat RHC may be indicated. Further improvement in survival of scleroderma patients with PAH may result from future therapeutic interventions targeting the underlying inflammatory process and the dysregulated endothelial proliferation mediating pulmonary vascular remodeling and consequent failure of the right heart.

CARDIAC INVOLVEMENT

The clinical manifestations of heart disease in scleroderma are highly variable, ranging from clinically silent cardiac involvement to frank heart failure. The reported prevalence of heart disease varies from 10% to more than 50%, depending on the diagnostic method used, but in general it tends to be underestimated. Symptoms such as dyspnea, chest pain, or palpitations are commonly thought to be secondary to more obvious organ involvement such as the lungs or the upper gastrointestinal tract, and the contribution of heart disease often is not appreciated until later stages. Cardiac disease can occur in both diffuse and limited subtypes of scleroderma and can manifest as a primary heart problem or in association with other organ failure. When clinically evident, heart disease entails an overall poor prognosis and predicts shortened survival.91 Along with pulmonary fibrosis and PAH, cardiac disease accounts for the majority of deaths in scleroderma. One study found that 25% of deaths could be directly related to heart disease (mostly heart...
failure and arrhythmias). In a large meta-analysis, cardiac involvement was associated with increased mortality (hazard ratio, 2.8; 95% confidence interval [CI], 2.1 to 3.8) after adjustments for age and gender. A survey comparing 129 scleroderma patients with a left ventricle ejection fraction (LVEF) of less than 55% with 250 subjects with normal LVEF demonstrated that male gender, age, digital ulcerations, myositis, and no use of calcium channel blockers were independent factors associated with left ventricular dysfunction.

Cardiac disease in scleroderma can be characterized by involvement of the endocardium, myocardium, and pericardium, separately or concomitantly. As a consequence, pericardial effusion, auricular and/or ventricular arrhythmias, conduction disease, valvular regurgitation, myocardial ischemia, myocardial hypertrophy, and heart failure are all reported. Clinically overt pericarditis is uncommon, but asymptomatic and hemodynamically benign pericardial effusions are frequently detected by ECHO. In a controlled study, significant pericardial effusion was found in about 15% of patients compared with 4% of controls. Pathologic studies have shown that some degree of pericardial involvement is detectable in 33% to 77% of scleroderma patients, usually with evidence of a fibrinous pericarditis with adhesions and chronic inflammatory cell infiltrates. Hemodynamic compromise secondary to tamponade physiology is rare but may require acute intervention with anti-inflammatory medication, pericardiocentesis, or creation of a pericardial window for slow decompression. The presence of a large pericardial effusion even if asymptomatic is a poor prognostic sign, particularly if associated with PAH or renal disease.

Focal myocardial fibrosis is the hallmark of established primary heart involvement in SSc and usually is not secondary to atherosclerotic coronary artery disease. In fact, pathologic studies and a survey of coronary angiograms have found little involvement of coronary arteries in scleroderma patients. Unlike other rheumatic diseases such as rheumatoid arthritis or SLE, scleroderma does not exhibit unequivocally an increased risk for coronary artery disease independent from usual risk factors. Nevertheless, typical angina symptoms should make one consider atherosclerosis of the coronary vessels or another process. Atypical chest pain can also be caused by musculoskeletal problems, esophageal reflux disease, or PAH mimicking cardiac disease. The fibrotic lesions in the heart of scleroderma patients are patchy, involve the myocardium of both ventricles, and usually are accompanied by evidence of microvascular disease with concentric intimal hypertrophy associated with fibrinoid necrosis of the intramural coronary arteries and arterioles. This results in reduced coronary flow reserve even with normal epicardial coronary arteries and in the absence of clinically manifested cardiac dysfunction. Vasospasm and associated myocardial perfusion defects are demonstrated to occur at rest, with exercise, and following cold exposure. These findings suggest that myocardial fibrosis may be associated with reversible vasospasm of the microvascular coronary circulation, and that vasodilating agents such as calcium channel blockers may have the capacity to improve coronary flow and prevent further cardiac damage. Patients with angina-like chest pain may need to undergo angiographic studies because nuclear perfusion studies are likely to be abnormal as a result of microvascular disease.

Electrocardiography (ECG) and Holter monitoring will often show some heart conduction defect or asymptomatic arrhythmia; these are thought to be consequences of myocardial fibrosis. Premature ventricular contractions are the most common abnormality, but sinus node dysfunction, first-degree heart block, supraventricular arrhythmias (supraventricular tachycardia, atrial fibrillation), and ventricular arrhythmias (ventricular tachycardia) are also observed. Complete heart block is uncommon. Scleroderma-related syncope is an ominous manifestation of late-stage PAH or an important arrhythmia. When an arrhythmia is suspected, a Holter monitoring study should be ordered promptly. More extensive evaluation with electrophysiologic studies is indicated in serious cases because a pacemaker or ablative therapy can help in management of these life-threatening complications.

Trans-thoracic echocardiography (TTE) is a widely used type of noninvasive cardiac testing; it represents a sensitive tool to detect the presence of cardiac disease. TTE can provide estimates of elevated right ventricular (RV) and pulmonary artery pressures, as well as evidence of right heart dysfunction such as atrial and ventricular dilation or septal wall motion abnormalities. A depressed left ventricular (LV) contractility is not frequently detected in scleroderma patients with the use of conventional TTE, but relaxation abnormalities can be found in up to 40% of cases. Left ventricular diastolic dysfunction is detected especially in patients with diffuse scleroderma, and may occur in the absence of systemic hypertension (independent of renal disease) and in association with pulmonary vascular disease. Although TTE is very helpful in monitoring underlying cardiac abnormalities later in the disease process, it lacks some sensitivity during preclinical stages of heart involvement. Novel diagnostic methods such as tissue Doppler echocardiography (TDE), with or without strain imaging, and magnetic resonance imaging (MRI) have been applied more recently to study myocardial disease in scleroderma, showing the ability to detect abnormalities earlier than conventional echocardiography and with greater accuracy. Provocative exercise testing during echocardiography can be used to detect underlying diastolic dysfunction and pulmonary hypertension not previously recognized by resting studies.

Cardiac evaluation has been improved by the systematic measurement of brain natriuretic peptide (BNP) or its prohormone N-terminal pro-BNP (NT-proBNP). Blood levels of NT-proBNP reflect both RV and LV volume or pressure overload. It can be a useful test and a prognostic marker for following right or left heart failure in scleroderma patients.

Patients with diffuse cutaneous scleroderma and skeletal muscle inflammation (polymyositis) are particularly prone to develop severe cardiomyopathy and to have a particularly poor prognosis. Acute myocarditis can also occur in association with inflammatory muscle disease and can present as sudden-onset heart failure. Prompt institution of immuno-suppressive treatment is indicated in these cases. An endomyocardial biopsy may be helpful in differentiating overt myocarditis from the myocardial fibrosis commonly present in scleroderma patients.
RENAL INVOLVEMENT

KEY POINTS

- Scleroderma renal crisis (SRC) is a life-threatening condition that occurs in 5% to 10% of scleroderma patients.
- Risk factors for SRC include early diffuse skin disease, use of corticosteroids, and the presence of anti-RNA polymerase III antibodies.
- Early pharmacologic intervention with angiotensin-converting enzyme inhibitors is crucial to control and possibly reverse the disease process.

Renal involvement in scleroderma is classically characterized by the abrupt onset of very high blood pressure (malignant hypertension), elevated plasma renin, and rising serum creatinine reflective of acute renal failure, along with a constellation of symptoms and clinical manifestations such as headaches, malaise, hypertensive retinopathy, encephalopathy, and pulmonary edema, usually referred to as scleroderma renal crisis (SRC). Renal disease in scleroderma can also be asymptomatic, and despite the high prevalence of histopathologic renovascular lesions typical of scleroderma vasculopathy, clinically important renal dysfunction, independent of scleroderma renal crisis, is seen in only a minority of patients. Although SRC is the most recognized renal complication, abnormal renal function can be explained by factors other than intrinsic scleroderma renal disease such as medication side effects, comorbid conditions, or associated heart, gastrointestinal, or lung disease. One survey suggested that mild unexplained proteinuria without loss of renal function or evidence of glomerular disease is a common sign of renal disease in scleroderma. A superimposed inflammatory kidney disease can also occur in scleroderma. Several cases of pauci-immune necrotizing crescentic glomerulonephritis associated with myeloperoxidase (MPO)-specific antineutrophilic cytoplasmic antibodies (ANCAs) have been reported in scleroderma patients with a presentation mimicking SRC.

SRC needs to be considered in scleroderma patients with sudden and severe elevation of blood pressure, with or without renal failure; in patients with sudden renal failure, with or without hypertension; and in patients with sudden-onset microangiopathic hemolytic anemia, with or without hypertension or renal failure. Nonmalignant hypertension alone without signs of renal dysfunction or mild urine abnormalities is unlikely to be due to SRC. Scleroderma renal crisis occurs in 5% to 10% of patients and mostly in those with diffuse scleroderma. It usually is encountered within the first 2 to 4 years from disease onset. The estimated median duration of scleroderma at the time of SRC diagnosis is 8 months. Occurrence in later disease is uncommon. Patients at greatest risk of developing SRC are those with diffuse cutaneous scleroderma, particularly those with the rapidly progressive skin disease. SRC is a rare complication in patients with limited scleroderma (approximately 1% to 2%). African-Americans and patients with diffuse skin disease who are treated with a high dose of corticosteroid (>40 mg/day of prednisone) or low-dose cyclosporine are at increased risk of developing an SRC. A long-term low dose of prednisone may also be a risk factor for SRC. For this reason, it is recommended to avoid corticosteroids or to use doses less than 10 mg daily if necessary. Renal crisis is also associated with a positive ANA (speckled pattern) anti-U3-RNP and usually is not seen in patients with anticentromere antibodies. Antibodies to RNA polymerase III have been found in about 60% of patients with SRC in one survey. Although antitopoisomerase I antibodies are prevalent in patients with diffuse skin disease, no association between their presence and SRC has been reported. Nonmalignant hypertension, abnormal urinalysis, and isolated increases in plasma renin activity are not predictors of an SRC.

Patients experiencing SRC present with typical signs of malignant hypertension, including headache, altered vision, signs of congestive heart failure, florid pulmonary edema, confusion, or neurologic signs secondary to hypertensive encephalopathy in the setting of an abnormally high blood pressure, usually above 150/90 mm Hg. Hypertensive encephalopathy is characterized by acute or subacute onset of lethargy, fatigue, confusion, headaches, visual disturbances (including blindness from hypertensive retinopathy), seizures, or cerebral hemorrhage. In about 10% of cases, SRC occurs with normal blood pressure. New-onset anemia, asymptomatic pericardial effusion, and cardiac disease may precede an SRC. Laboratory testing at onset may show normal or high creatinine, modest to absent proteinuria, and/or microscopic hematuria. Renal insufficiency usually is not the presenting problem in an SRC but rather is a late complication of untreated disease. However, a rapid rise in serum creatinine can occur over several days with a fairly benign urinalysis—less than 2 g of protein excretion in 24 hours—and a mild hematuria with some granular casts. A microangiopathic hemolytic anemia and thrombocytopenia can accompany or precede an SRC and may be the only finding in a normotensive patient. This presentation with normotensive SRC usually entails a worse prognosis. Distinguishing SRC-associated thrombotic microangiopathy from idiopathic thrombotic thrombocytopenic purpura may be difficult but is important because treatments differ. Accurate assays for ADAMTS (a disintegrin and metalloproteinase [ADAM family] with thrombospondin-1 domains)-13 activity can be helpful because its level should be normal in scleroderma.

Poorer outcome is more frequent among males and among patients presenting with creatinine greater than 3 mg/dL (265 mmol/L). Approximately two-thirds of patients with SRC will require renal support with dialysis, but about half will recover function and be able to discontinue therapy, usually within several weeks, most before 6 months. However, cases of recovery from up to 24 months of dialysis support have been reported, distinguishing SRC from other causes of renal failure. The renal crisis truly mimics malignant hypertension, with rapidly progressive renal failure secondary to microvascular disease, vasospasm, and tissue ischemia. The characteristic lesion of vascular disease such as intimal hyperplasia or luminal narrowing involving in particular the renal arcuate arteries is detected even in patients without SRC. Although the renal pathology is now well described, the pathogenesis of the renal crisis remains poorly understood. It is assumed that intrinsic renal vessel disease is complicated by an intense vasospasm
triggering high levels of plasma renin, which results in malignant hypertension.

Before the discovery of angiotensin-converting enzyme (ACE) inhibitors, hypertensive renal crisis had almost uniformly a fatal outcome. In contrast, patients treated with ACE inhibitors have a good outcome 60% of the time, and death or end-stage renal disease is now much less common. The mortality rate at 1 year associated with SRC has decreased from 76% to less than 15%. However, despite aggressive antihypertensive therapy, the 5-year survival of patients with SRC remains only 65%. Although ACE inhibitors are standard therapy in SRC, it remains unclear whether other related drugs such as angiotensin receptor blockers (ARBs) can be effective in treating or preventing a crisis. Endothelin and other mediators of vascular disease are implicated in the pathogenesis of SRC; thus interference with endothelin function by a specific inhibitor may offer a novel therapeutic approach. A vasodilating prostaglandin (prostacyclin) may also provide rapid control of blood pressure while improving renal blood flow. The overall outcome in SRC remains unsatisfactory in terms of acute disease outcome, morbidity, and need for renal transplantation or long-term renal support with dialysis. Data suggest that use of an ACE inhibitor before a crisis does not prevent an SRC.108

Any newly developed hypertension in a scleroderma patient should be urgently evaluated because the key to successful therapy is early detection and rapid intervention (Figure 84-15). Patients with SRC should be hospitalized and immediately started on an ACE inhibitor. Captopril, a rapid-acting ACE inhibitor, should be started, to allow dose increases until systolic blood pressure is down by 20 mm Hg/24 hr while avoiding hypotension. Treatment continues with an ACE inhibitor even if creatinine continues to rise. When full doses of an ACE inhibitor are not controlling blood pressure, other antihypertensive medications may be added, including calcium channel blockers, endothelin inhibitors, and prostacyclin or an ARB (which is infused). Some patients continue to have progressive renal failure despite attaining control of blood pressure. Other causes of renal disease always need to be considered, especially in patients with limited scleroderma, and in those with a high level of proteinuria or abnormal urinalysis such as red cell casts. Renal biopsy should be done to confirm the diagnosis and to gain insight into the prognosis by determining the degree of renal damage. Detection on kidney biopsy of other forms of glomerular inflammation is important in that it will dictate a different specific treatment approach.

Successful renal transplantation has been performed in scleroderma patients, providing an overall survival benefit over long-term dialysis. Renal allograft survival at 3 years is about 60%, which is comparable with the rate observed in systemic lupus erythematosus. It is not recommended to pursue renal transplantation until it is clear that recovery will not spontaneously occur. This can be determined in part by assessing the degree of damage on renal biopsy and by waiting at least 6 months (some suggest 2 years) after recovery from the acute crisis. Recurrence of SRC is uncommon (5%) after renal transplant.

**MUSCULOSKELETAL INVOLVEMENT**

Musculoskeletal symptoms are almost always present in scleroderma, usually are multifactorial in origin, and often are the initial symptoms of the disease. The most common

---

**Figure 84-15** Management of scleroderma renal crisis. ACEi, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; BP, blood pressure; CCB, calcium channel blocker; ETA, endothelin receptor antagonist; IV, intravenous; SRC, scleroderma renal crisis.
symptoms are nonspecific pain, stiffness, and diffuse muscular discomfort that mimics a flu-like syndrome. Impaired hand function, characterized by decreased hand mobility, reduced dexterity, and decreased grip force, is in particular a major source of difficulty in performing activities of daily living. The degree and type of musculoskeletal disease vary a great deal and are influenced by the duration of disease, the level of overall disease activity, and the subtype of skin involvement. For example, in early diffuse scleroderma, pain is often widespread in areas of skin inflammation and advancing fibrosis that can involve joint structures, tendons, subcutaneous tissue, and underlying muscle. In later-stage diffuse disease, joint contractures and muscle atrophy are often associated with pain and loss of function, causing significant disability. In limited scleroderma patients, puffy fingers and loss of hand function and grip may be the only musculoskeletal symptoms throughout the disease course. Although histologic evidence suggests that synovitis occurs, arthralgias, pain, and stiffness without signs of frank arthritis constitute the usual clinical picture. However, overlap syndromes of rheumatoid-like polyarthritis or inflammatory muscle disease can dominate the clinical picture in an individual patient.

In the early edematous phase of diffuse scleroderma, patients often are diagnosed with a carpal tunnel syndrome due to soft tissue swelling and inflammation in the hand and wrist area. Erosive arthritis with joint space narrowing can be seen, but soft tissue swelling, periarthritis osteopenia, and contractures of the joint are more common. Distal bone resorption, osteotomy, and periarthritis calcification are found in the fingers of patients with later-stage diffuse scleroderma. Contractures of the proximal interphalangeal (PIP) and metacarpalphalangeal (MCP) joints are most common, and rarely of the distal interphalangeal (DIP) joint. Patients with diffuse skin disease develop contractures of large joints, including wrist, elbows, shoulders, hips, knees, and ankles. These contractures are the hallmark of severe scleroderma and are associated with fibrosis and ankylosis of the joint, resulting from disease in the overlying skin, fascia, joint capsule, and tendons.

Tendon friction rubs can be felt as a coarse crepitus over joints or over the forearm or lower leg with adjacent joint movement. These rubs are thought to be secondary to low-grade tenosynovitis, local edema, and fibrosis of the tendon sheath, fascia, and joint structures. Friction rubs are seen primarily in patients with diffuse skin disease; when present, they are an indicator of a poor overall prognosis. They are found in about 15% to 30% of patients and are more common in those with diffuse skin disease and in patients with antitoposomerase, anti-RNA polymerase, or anti-U3-RNP antibodies.

The prevalence of signs of joint inflammation in scleroderma is estimated in relatively small observational studies to be about two-thirds of patients. A large multicenter survey reported the frequencies of synovitis, tendon friction rubs, and joint contractures to be 16%, 11%, and 31%, respectively. Synovitis, tendon friction rubs, and joint contractures are more likely to occur together and are more prevalent in patients with the diffuse cutaneous subset. These manifestations also tend to be associated with severe vascular, muscular, renal, and interstitial lung involvement. Synovitis, joint contracture, and tendon friction rubs are associated with more severe disease and with systemic inflammation.

Erosive arthritis is commonly associated with periarticular calcinosis (Figure 84-16A) and can be seen with significant bone loss or osteolysis. A radiologic survey of 120 patients found abnormalities including erosion (21%), joint space narrowing (28%), arthritis (defined by concomitant erosion and joint space narrowing) (18%), radiologic demineralization (23%), acro-osteolysis (22%), flexion contracture (27%), and calcinosis (23%). Radiologic studies using ultrasonography and MRI confirm that hand involvement is striking, with articular, bone, and soft tissue changes. Resorption of distal phalanges was significantly associated with digital ulcers and extra-articular calcification; flexion contracture was associated with the diffuse cutaneous form and with high Health Assessment Questionnaire (HAQ) disability scores. Calcification is most often seen in patients with digital ulcers but was similarly observed in patients with the diffuse or limited cutaneous subtype. Calcinosis is often seen in the subcutaneous tissue in areas of trauma such as the extensor surfaces of the forearms, elbows, or patellae. Unlike the more diffuse calcinosis seen in myositis patients, calcium deposits in scleroderma tend to be found in clusters around joints and sites of trauma. These deposits can restrict joint motion and may be associated with acute inflammation similar to a gouty arthritis, or the deposits can rupture through the skin, weeping a thick white material. Once the skin is broken, secondary infection can occur.

Erosive polyarthritis, seemingly specific for scleroderma, or serositive rheumatoid arthritis can occur in patients with limited or diffuse skin disease. However, the prevalence of scleroderma with rheumatoid arthritis overlap is unusual (1% to 2%) with an estimated incidence of 5%. This is confirmed by the low frequency of anticyclic citrullinated protein antibodies found in patients with scleroderma and arthritis. When an inflammatory arthritis is present, traditional medications used in rheumatoid arthritis can be helpful. One case series demonstrated that the tumor necrosis factor (TNF) inhibitor etanercept appeared to be efficacious in improving active inflammatory joint disease in a subset of scleroderma patients, and it was generally safe and well tolerated.

A sense of weakness in the muscles of the hands, arms, and legs is very common (80%) and can be subtle or profound. One study found measurable weakness in 30% of patients with proximal more than distal strength compromised. Myopathy in an individual scleroderma is often multifactorial, but patients with myopathy are at increased risk for poor outcome and decreased survival. Weakness is often caused by muscle atrophy secondary to the inflexibility due to joint disease and fibrotic skin with lack of usual mobility and exercise. It can also occur because of malnutrition resulting from scleroderma bowel disease. Muscle weakness in scleroderma may be secondary to direct muscle disease. In diffuse scleroderma, fibrosis can extend into the striated muscle, causing muscle atrophy and clinical weakness. This has been called a fibrosing myopathy. It presents in the setting of severe diffuse skin disease with joint contractures and mild elevation of creatine phosphokinase (CPK). Electromyography (EMG) demonstrates nonirritable myopathy, and muscle biopsy shows little inflammation,
fibrosis, and type 2 muscle atrophy. The fibrosing myopathy does not respond to traditional anti-inflammatory medications, including corticosteroids. This type of muscle disease is associated with cardiomyopathy, heart failure, and arrhythmia that can be severe and irreversible.

Approximately 5% to 10% of scleroderma patients have an inflammatory muscle disease that can follow the same course as polymyositis and other forms of idiopathic inflammatory myopathy. Patients present with rapid proximal muscle weakness associated with high CPK. EMG shows irritable myopathy, and muscle biopsy is typical of an inflammatory myositis. In some cases with severe polymyositis, myocardiitis can occur with heart failure. High levels of anti-PM/Scl-100 antibodies are found in about 4% of patients with SSc, and these patients often present with acute inflammatory muscle disease. Patients with muscle weakness should be fully evaluated for underlying causes with the use of laboratory measures of muscle enzymes, electromyographic studies, magnetic resonance imaging, and muscle biopsy. These studies will help differentiate inflammatory muscle disease from other forms of myopathy.

**ENDOCRINE INVOLVEMENT**

Endocrine dysfunction can complicate the course and management of scleroderma patients. The most common endocrine problem associated with scleroderma is thyroid disease, and the frequency of other endocrinopathies is similar to that of the general population. Thyroid disease is reported in 10% to 15% of patients with scleroderma. A case-control study found a 10 to 14 times increased risk of hypothyroidism in scleroderma. Evidence for both fibrosis of the thyroid and autoimmune injury exists. Patients with features of the CREST syndrome (subcutaneous calcinosis, Raynaud’s phenomenon, esophageal hypomotility, sclerodactyly, and telangiectasia) are more likely to have hypothyroidism than those with diffuse skin disease. An autoimmune basis for thyroid disease is suggested by one series, which found Hashimoto’s thyroiditis in 6% and Graves’ disease in 3% of patients. A survey of 719 scleroderma patients found that 273 (38%) had at least one other autoimmune condition; the most frequent was autoimmune thyroid disease. Thyroid autoantibodies are commonly found in scleroderma patients with hypothyroidism. In addition, the association of CREST syndrome with autoimmune hypothyroidism and primary biliary cirrhosis has been reported, suggesting multiple autoimmune targets in these patients. Thyroid disease may not be recognized in patients with subclinical disease or with symptoms associated with chronic multisystem disease. It is recommended that thyroid function should be monitored periodically, especially in patients with limited scleroderma with the CREST syndrome and in those with long-standing disease.

**OTHER ASSOCIATED MANIFESTATIONS**

It must be recognized that in patients who have scleroderma, several less common and often forgotten complications of this disease may require attention. Awareness of these associated problems can improve quality of life and prevent undue distress. Bone disease can occur for a variety of reasons. Most patients with scleroderma are women who are near or in menopause, and therefore are at risk for osteoporosis and its complications. This risk is increased by...
the chronic inflammatory nature of the disease, disuse, and several commonly used medications. Malabsorption and lack of sun exposure may lead to vitamin D deficiency, and bowel dysfunction can limit calcium intake. Avascular necrosis of carpal bones of the wrist is reported and is thought to be secondary to scleroderma peripheral vascular disease. In addition, osteolysis or bony resorption of the tips or tufts of the fingers, middle phalanges, distal radius, and ulna bones (less commonly, the distal clavicle, ribs, mandible, and distal toes) may occur (Figure 84-16B and C). Acro-osteolysis is fairly common, presenting with shortening of the finger or pain of the fingertip. Osteolysis is thought to be a manifestation of peripheral vascular disease and poor nutritional blood flow on affected bones. Neurologic complications can be a significant unrecognized problem. The central nervous system is generally spared in scleroderma, but unilateral or bilateral trigeminal neuralgia is known to occur in a subset of patients. Peripheral neuropathy occurs in a higher percentage of patients than expected owing to multiple causes, including nutritional deficiency from gastrointestinal disease, entrapment neuropathy, vasculitis, or vasculopathy of small vessels typical of scleroderma vascular disease. Carpal tunnel syndrome can complicate diffuse scleroderma and may occur in patients with associated inflammatory joint disease.

Muscle weakness is very common, and the presence of a myopathy is associated with a poor prognosis. Multiple factors, including inflammatory polymyositis, disuse atrophy, malnutrition, medications, and a fibrosing myopathy associated with diffuse scleroderma skin disease, can cause muscle dysfunction. Autoimmune hepatitis and biliary cirrhosis are reported in patients with the CREST syndrome. It is recommended that all patients with scleroderma be screened for associated autoimmune-mediated liver disease because it usually responds to immunosuppressive therapy.

Dry eyes (keratoconjunctivitis sicca) and/or mucous membranes (xerostomia) occur in about 25% of patients. Dry eyes may occur as the result of severe facial periorbital skin fibrosis inhibiting full closure of the eyelids. Topical artificial tears should be used frequently in these cases. Facial changes alter oral aperture and decrease facial flexibility. Vertical lines around the lips and thinning of the lips along with pinching of the nose affect self-esteem, especially in younger women. A decreased oral aperture prevents easy daily oral hygiene and dental procedures. Loosening of the teeth happens because of loss of normal periodontal ligament attaching the tooth to the alveolar bone. Decreased saliva and difficulty performing usual dental hygiene can lead to significant periodontal disease. Treatment of oral manifestations requires early referral for experienced dental care and frequent cleaning with fluoride treatment. Pilocarpine and cevimeline are agonists for the muscarinic cholinergic receptors that can increase saliva and improve symptoms of xerostomia.

Most data suggest normal fertility in scleroderma, but during pregnancy, risk for hypertension, scleroderma renal crisis, or premature fetal loss is increased. Sexual performance is often affected, particularly in women with diffuse disease. Decreased sexual satisfaction is associated with painful inflexible joints, dry vaginal membranes, and general fatigue. Erectile dysfunction and associated impotence among male scleroderma patients have been reported in more than 80% of men with scleroderma. This results from local fibrosis of the corporeal smooth muscle and scleroderma microvascular disease. Psychogenic and medication-induced sexual dysfunction must be considered.

Several causes of lower extremity lesions can occur, including skin fibrosis, stasis dermatitis, lipodermatosclerosis, panniculitis, subcutaneous calcinosis, and leg ulcers. Leg ulcers may result from traumatic breakdown of atrophic fibrotic skin, ischemic ulcers secondary to scleroderma vascular disease, vasculitis, and occlusive vascular events resulting from an associated hypercoagulable state. Livedoid vasculopathy is a rare complication causing skin ulcerations on the lower leg and foot as the result of small vessel vasculitis or small vessel thrombosis.

**PSYCHOSOCIAL ASPECTS**

Scleroderma is a disease that can alter virtually every aspect of the patient's life. Although scleroderma is often mild in its expression, the patient is now faced with a life crisis and must deal with a rare chronic disease that is complex and is not completely understood by the medical community. Patients learn that scleroderma is potentially a life-threatening disease that alters physical capacity and can be disfiguring. At the onset of the disease, patients are often confused, anxious, and frightened. Fear and misunderstanding become a major source of distress. Studies of psychosocial adjustment support the conclusion that a substantial proportion of patients with scleroderma have difficulty adapting to the disease and have reduced quality of life measures. Although little evidence suggests that scleroderma directly affects the central nervous system with a direct organic cause for altered mental status, a complex array of medications used to treat the disease and its complications may have some influence on mood and sense of well-being.

Psychosocial aspects of the disease do not appear to be consistently influenced by age, gender, ethnicity, education, or marital status. However, some data suggest that being disabled with low financial income leads to higher psychological distress. Evidence suggests that depression in scleroderma patients is common and is related more to the patient's personality, the degree of pain, and the level of social support than to actual disease severity. Disease-related disability, as measured by the Health Assessment Questionnaire Disability Index (HAQ-DI), is a more important predictor of adjustment problems than are disease symptoms related to severity. Pain and fatigue are common in scleroderma. Together with depressive symptoms, they are the most significant determinants of physical functioning and social adjustment—two important components of health-related quality of life—in patients with limited and diffuse disease subtypes.

Personality traits influence the degree of psychological adjustment. Patients with scleroderma who describe themselves as anxious, worried, tense, or detail-oriented are more likely to be depressed than patients who describe themselves as agreeable or extraverted and outgoing. Disease unpredictability and lack of control cause increased distress. High self-reported uncertainty is related to poor adjustment. Thus, increasing the patient's understanding of the disease and expectations through education may reduce
uncertainty and improve quality of life. Body image dissatisfaction is a significant concern. Patients are distressed by disease-related changes in visible parts of their body, especially the fingers and hands. A survey found that tight skin and facial changes were not as distressful as hand deformity with finger contractions and hand dysfunction. Sexual function is affected in both men and women with scleroderma. A significant proportion of men experience erectile dysfunction. Women with scleroderma, particularly those with diffuse skin disease, have high levels of sexual impairment compared with women with other chronic diseases for which sexual function has received greater attention. Both may have a decreased sexual drive during the disease course. The impact of sexual function on quality of life is not well studied, but it exists as a problem and should be appropriately addressed as part of comprehensive care.

Psychological factors with demonstrated relevance to scleroderma include pain, depression, and distress about disfigurement, physical function, and social function. Although these dimensions of quality of life are interrelated, pain, depression, and distress about disfigurement are common and may respond to psychological intervention. This begins with providing compassionate support beyond just ordering medications. It is important to spend time in a comfortable setting educating patients about scleroderma, providing a clear understanding of the degree of their disease and explaining to them what they need to do and what can be done for relief. Having insight into social support and life circumstances, including financial distress, work environment, and family structure, provides a framework for helping to decrease external distress. For example, a family conference held to explain the reality of the health situation or the needs of the patient can reduce home tension and can clarify the patient's capacity in their family role. An adjustment in the work environment (e.g., permission for a space heater) can make a difference. Clearly, effective social support improves quality of life. Follow-up visits that provide time to discuss issues of coping and social support are most important. Recognizing the personality of patients while addressing their specific concepts is helpful. Treating underlying depression and especially providing effective management of disease-related pain help to improve quality of life and reduce social and psychological distress. Body image dissatisfaction is a significant concern in women with scleroderma and should be assessed routinely. Early identification and treatment of body image dissatisfaction may help to prevent depression and psychosocial impairment in this population. The possibility of premature death from the disease is a major cause of fear and needs to be addressed with the patient and the family. Most often, life expectancy is not influenced by the disease, yet patients fear death because they have scleroderma. When a patient is facing death from severe disease, appropriate honest and sensitive support must be provided.

**THERAPEUTIC APPROACH FOR DISEASE MODIFICATION**

**Immunotherapy**

Scleroderma is thought to be an inflammatory connective tissue disorder triggered by an autoimmune process. Nonselective immunosuppressive treatments are usually employed in scleroderma to treat specific organ manifestations such as early progressing skin disease, active interstitial lung disease, and underlying inflammatory joint or muscle disease (Table 84-5). Cyclophosphamide (CYC) has shown some modest efficacy in scleroderma-related ILD in a randomized placebo-controlled trial. Current therapeutic protocols use CYC as a daily oral regimen or as monthly intravenous (IV) therapy, until control of disease is achieved. The CYC is often followed by maintenance treatment with azathioprine (AZA) or mycophenolate mofetil (MMF), agents that have antiproliferative effects on inflammatory cells and in particular on activated T and B lymphocytes. Uncontrolled studies suggest benefit of MMF in both ILD

<table>
<thead>
<tr>
<th>Therapeutic Category</th>
<th>Drug</th>
<th>Clinical Studies (Reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonselective immunotherapy</td>
<td>Cyclophosphamide</td>
<td>Tashkin 2006†</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hoyles 2006††</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nadashkevich 2006††</td>
</tr>
<tr>
<td></td>
<td>Mycophenolate mofetil</td>
<td>Tashkin 2007††</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lioissis 2006††</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nhtyanova 2007††</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gerbino 2008††</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zamora 2008††</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Derk 2009††</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dheda 2004††</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nadashkevich 2006††</td>
</tr>
<tr>
<td></td>
<td>Azathioprine</td>
<td>Paone 2007†‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Berezne 2008†‡</td>
</tr>
<tr>
<td></td>
<td>Methotrexate</td>
<td>van den Hoogen 1996†‡</td>
</tr>
<tr>
<td>T cell–targeted immunotherapy</td>
<td>Cyclosporin A</td>
<td>Pope 2001†§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clements 1993‡‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Filaci 1999‡‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Morton 2000‡‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Matteson 1996†‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stratton 2001†‡</td>
</tr>
<tr>
<td></td>
<td>Antithymocyte globulin</td>
<td>Roek 1993†‡</td>
</tr>
<tr>
<td></td>
<td>Extracorporeal photopheresis</td>
<td>Krasagakis 1998†‡</td>
</tr>
<tr>
<td></td>
<td>Sirolimus (rapamycin)</td>
<td>Knobler 2006†‡</td>
</tr>
<tr>
<td></td>
<td>Rituximab</td>
<td>Su 2009†‡</td>
</tr>
<tr>
<td>B cell–targeted immunotherapy</td>
<td>IVIG</td>
<td>Lafortes 2009†§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smith 2010†§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daoussis 2010†§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Levy 2000†§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Amital 2003‡§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Levy 2004‡§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ihn 2007†§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nacci 2007†§</td>
</tr>
<tr>
<td>Intraeryous immunoglobulins</td>
<td>TNF inhibitors</td>
<td>Lam 2007†§</td>
</tr>
<tr>
<td></td>
<td>Etanercept</td>
<td>Denton 2009†§</td>
</tr>
<tr>
<td></td>
<td>Infliximab</td>
<td></td>
</tr>
<tr>
<td>Biologic immunotherapy</td>
<td>CAT-192 (anti–TGF–β ab)</td>
<td>Denton 2007†§</td>
</tr>
<tr>
<td></td>
<td>Imatinib mesylate</td>
<td>Gordon 2009†§</td>
</tr>
<tr>
<td></td>
<td>Autologous HSCT</td>
<td>Pope 2009†§</td>
</tr>
<tr>
<td>Antifibrotic therapy</td>
<td></td>
<td>Binks 2001‡†</td>
</tr>
<tr>
<td>Cell-based immunotherapy</td>
<td></td>
<td>McSweeney 2002†‡</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Farge 2004†§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nash 2007†§</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oyama 2007†§</td>
</tr>
</tbody>
</table>

HSCT, hematopoietic stem cell transplantation; IVIG, intravenous immunoglobulin; TGF–β, transforming growth factor-beta; TNF, tumor necrosis factor.
and active scleroderma skin disease. Evidence supporting the use of AZA as a primary agent to treat scleroderma-related ILD is weak and remains limited to small retrospective studies. Two open-label trials support its role as maintenance therapy after primary CYC immunosuppression. Methotrexate (MTX) is frequently used in scleroderma to treat associated inflammatory arthritis and myositis. MTX efficacy in skin disease and lung function has been investigated in two randomized placebo-controlled trials showing only a small benefit on skin scores after 6 months (weekly intramuscular injections) or 12 months (weekly oral therapy), respectively. Based on the findings of these studies, many experts use MTX for active skin disease.

Selective immunotherapy has also been used in scleroderma (see Table 84-5). T lymphocyte–directed treatments such as cyclosporin A, sirolimus (rapamycin), and antithymocyte globulin (ATG) have shown some benefit. However, these studies generally were small and were limited by drug toxicity. The use of extracorporeal photopheresis (ECP) has been reported with modest improvement in skin disease but lack of any efficacy on internal organ manifestations. This approach is not currently supported. Preliminary studies of rituximab, a chimeric immunoglobulin (Ig)G1 monoclonal antibody directed against CD20, a surface molecule expressed on early pre-B and mature B cells, have not yet demonstrated solid evidence of clinical benefit. However, therapeutic strategies based on B cell depletion remain an active area of research. Administration of intravenous immunoglobulins (IVIG) to scleroderma patients has been reported in uncontrolled studies with some evidence that it can improve skin fibrosis or joint disease.

Studies of cell-based immunotherapy using autologous hematopoietic stem cell transplantation (HSCT) preceded by myeloablative conditioning regimens are under way for severe cases of scleroderma. Overall, preliminary data suggest that this approach is effective in improving skin fibrosis and providing stabilization of the lung and other organ function. However, significant concern regarding toxicity and treatment-related mortality exists such that this approach is still considered investigational and is used only in patients with life-threatening diffuse scleroderma. Nonmyeloablative (immunoablative) conditioning regimens with or without HSCT have demonstrated similar results.

**Treatment of Fibrosis**

To date, no drugs have proved able to reverse the fibrotic disease process in scleroderma, but several potential agents are being used off-label or are under investigation. Their use is still based on in vitro data, animal studies, or case series because formal well-designed controlled trials are still missing.

D-penicillamine is a chelating agent that blocks collagen cross-linking and thus has the potential to have an antifibrotic effect. Several case reports and retrospective reviews have suggested that D-penicillamine is beneficial. A double-blind, randomized clinical trial comparing the efficacy of low-dose versus high-dose D-penicillamine in early diffuse scleroderma found no significant difference between skin scores of the two groups at 24 months. Given the potential toxicity of D-penicillamine, its lack of effectiveness in the only controlled trial, and the long duration before benefit was achieved in several uncontrolled surveys, the drug has fallen out of favor and generally is not used.

Transforming growth factor-beta (TGF-β) is a cytokine known to promote fibroblast proliferation and differentiation, in addition to upregulation of collagen and extracellular matrix synthesis. Therefore, it represents an appealing target to specifically control progression of collagen and extracellular deposition in tissues. A human recombinant neutralizing TGF-β antibody was studied versus placebo in a relatively small phase II/III trial in patients with early diffuse scleroderma. No clinical benefit was observed when skin score or lung function testing parameters were evaluated. Although the results of this study were negative, it does not negate the idea that another more potent neutralizing antibody might be helpful.

Imatinib, dasatinib, and nilotinib are small molecules that inhibit the tyrosine kinase activity of the abl-kinases and platelet-derived growth factor (PDGF) receptors, thus interfering with important profibrotic pathways activated in scleroderma. In addition, dasatinib inhibits Src kinases, which are also involved in fibroblast differentiation and secretory function. Several uncontrolled trials with tyrosine kinase inhibitors are currently under way, with early data analysis reporting conflicting results. Adverse events were common, including fluid retention, nausea, fatigue, and elevation of creatine kinase. Therefore, although a potent antifibrotic effect has been shown by tyrosine kinase inhibitors in vitro or in animal models, clinical studies to date have not provided solid evidence for their use in treating skin and systemic fibrosis in scleroderma. Formal clinical trials are needed.

Halofuginone, a plant-derived alkaloid with antifibrotic properties via inhibition of TGF-β and T cell activation, is known to reduce collagen production in animal models. Only limited use of topical halofuginone therapy in scleroderma skin disease is reported with some encouraging results. Rosiglitazone is an agonist of peroxisome proliferative-activator receptor gamma (PPAR-γ) and has been shown to suppress in vivo fibroblast production of collagen and to alleviate fibrosis in the bleomycin-induced scleroderma mouse model. Pirfenidone (5-methyl-1-phenylpyridin-2[H]-one) is a novel antifibrotic agent with evidence of benefit in patients with idiopathic pulmonary fibrosis (IPF). Clinical trials in scleroderma have not been performed. Neutralizing antibodies against connective tissue growth factor (CTGF) can effectively suppress the development of skin fibrosis in animal models, and anti-CTGF therapy is now under consideration to treat scleroderma.

**Treatment of Vascular Disease**

Although scleroderma is generally considered a fibrosing disorder, it is well recognized that underlying vasculopathy plays a fundamental role in its pathogenesis and associated tissue injury. In RP, small and medium-sized peripheral blood vessels involved in tissue nutrition and body thermoregulation are affected. The pathologic and clinical consequences of scleroderma vascular disease are not limited to
the skin but are widespread and are found in all involved organs. Currently, no guidelines have been put forth for the treatment of scleroderma vascular disease. Therefore, pharmacologic therapies are focused on the treatment of organ-specific disease such as scleroderma renal crisis, PAH, and RP. Conventional treatment strategies have been limited to the use of nonspecific vasodilator agents (see Raynaud’s phenomenon section). More recently, it has been appreciated that agents targeting specific biologic processes involved in vascular disease (i.e., prostacyclin, endothelin antagonists, and phosphodiesterase inhibitors) may have a broader beneficial effect on scleroderma (see Raynaud’s phenomenon and pulmonary hypertension section).

**SUMMARY OF CURRENT PRACTICAL RECOMMENDATIONS FOR TREATMENT**

The key to current management of scleroderma is to not isolate treatment to one problem that seems to be dominant, but to appreciate the complexity and dynamics of this multisystem disease (Table 84-6). It is clear that use of just one drug is not effective in managing scleroderma, and that long-term therapy is required with possible adjustments based on specific clinical circumstances. A combination therapy approach that attempts to treat the immune response, the vascular disease, and the underlying tissue fibrosis is recommended.

Patients with severe RP should be on vasodilators and antiplatelet therapy with aspirin. Vasodilator treatment should start with a calcium channel blocker, with a second agent added if needed. This can be a phosphodiesterase inhibitor or intermittent intravenous prostacyclin (it is hoped that oral prostacyclin will be available soon). If digital ulcers are recurring, then the addition of an endothelin inhibitor or a statin is suggested.

Patients with hypertensive renal disease should be treated urgently with an ACE inhibitor or, if needed, with additional ARBs or antihypertensives, including calcium channel blockers or a prostacyclin analogue. The use of ACE inhibitors to prevent a scleroderma renal crisis is not recommended.

All patients need periodic professional dental care, and if loss of saliva is noted, the use of pilocarpine or cevimeline is recommended. Patients with upper gastrointestinal disease should be treated with lifestyle changes such as improving their eating habits with proton pump inhibitors, in case of unremitting symptoms with a prokinetic drug (metoclopramide, domperidone). If the lower intestinal tract is involved, then rotating antibiotics can improve episodes of pseudo-obstruction, diarrhea, or malabsorption.

First-line therapy for active skin disease has not yet been established, but we recommend immunosuppressive therapy based on severity and level of disease activity. Although most recommend methotrexate, we prefer mycophenolate for mild disease and cyclophosphamide for severe aggressive disease (daily oral, monthly IV). More innovative therapies are needed and may include antithymocyte globulin, immunosuppression or myeloablation with or without hematopoietic stem cell rescue, and intravenous gamma globulin given alone or in combination with other immunosuppressive agents. Participation of patients in new clinical trials should be pursued.

Active interstitial lung disease should be treated with immunosuppressive therapy. Induction therapy with cyclophosphamide (6 to 12 months; daily oral or monthly IV therapy) followed by maintenance (several years) with mycophenolate or azathioprine is recommended. Mycophenolate or azathioprine alone is a reasonable alternative.

Pulmonary vascular disease and PAH are best treated with vasoactive drugs, including endothelin-1 inhibitors, phosphodiesterase inhibitors, or prostacyclin analogues, given alone or in combination. The role of anticoagulation or immunosuppressive agents is not yet defined. Treatment of associated heart failure is most important.

Cardiac disease should be treated with vasoactive drugs such as calcium channel blockers, other antihypertensive agents, diuretics, and antiarrhythmic medications.

**Table 84-6 Current Recommendations for Treatment of Scleroderma**

<table>
<thead>
<tr>
<th>Manifestation</th>
<th>Primary Therapy</th>
<th>Alternative/Second-Line Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raynaud’s phenomenon</td>
<td>Vasodilators (CCB)</td>
<td>PD5 inhibitors, prostacyclin, endothelin antagonists</td>
</tr>
<tr>
<td>Hypertensive renal disease</td>
<td>ACE inhibitors</td>
<td>ARBs, CCB, prostacyclin, renal transplant (wait at least 24 mo)</td>
</tr>
<tr>
<td>GI involvement</td>
<td><strong>Upper GI</strong></td>
<td>EGD to treat stenosis and/or GAVE</td>
</tr>
<tr>
<td></td>
<td>Dental/periodontal care, lifestyle modifications,</td>
<td>Total parenteral nutrition</td>
</tr>
<tr>
<td></td>
<td>proton pump inhibitors, prokinetics</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Lower GI</strong></td>
<td></td>
</tr>
<tr>
<td>Skin</td>
<td>Probiotics, rotational antibiotics</td>
<td></td>
</tr>
<tr>
<td>Interstitial lung disease</td>
<td>Mycophenolate mofetil, cyclophosphamide</td>
<td>IVIG, ATG, research trial (severe cases)</td>
</tr>
<tr>
<td>Pulmonary arterial</td>
<td>Cyclophosphamide, mycophenolate mofetil, azathioprine</td>
<td>Research trial</td>
</tr>
<tr>
<td>hypertension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac involvement</td>
<td>PDES inhibitors, endothelin antagonists,</td>
<td>Combination therapy, atioseptostomy, lung transplantation,</td>
</tr>
<tr>
<td>Joints</td>
<td>prostacyclin</td>
<td>research trial (myocardial inflammation)</td>
</tr>
<tr>
<td>Muscles</td>
<td>Heart failure therapy, diuretics, CCB</td>
<td>IVIG (if contractures and rashes are present), PT/OT</td>
</tr>
<tr>
<td>Psychosocial</td>
<td>Prednisone, methotrexate, TNF inhibitors</td>
<td>IVIG</td>
</tr>
<tr>
<td></td>
<td>Prednisone, methotrexate, azathioprine</td>
<td>Support group</td>
</tr>
<tr>
<td></td>
<td>Antidepressants, pain control, sleep control</td>
<td></td>
</tr>
</tbody>
</table>

ACE, angiotensin-converting enzyme; ARBs, angiotensin receptor blockers; ATG, antithymocyte globulin; CCB, calcium channel blockers; EGD, esophagogastroduodenoscopy; GAVE, gastric antral vascular ectasia; GI, gastrointestinal; HSCT, hematopoietic stem cell transplantation; IVIG, intravenous immunoglobulin; PDES, phosphodiesterase-5 inhibitor; PT/OT, physical therapy/occupational therapy; TNF, tumor necrosis factor.
Immunosuppression should be used if inflammatory muscle or pericardial disease is present.

Inflammatory arthritis should be treated in a similar fashion to the approach used in treating rheumatoid arthritis. However, a fibrosing process causing a nonerosive arthropathy with friction rubs and joint contractures is best treated with the same medications used for active scleroderma skin disease.

Treatment for muscle disease will vary depending on the presence of inflammatory myositis or a nonirritative fibrosing myopathy. The former is treated with corticosteroids, methotrexate, or other immunosuppressive agents effective in immune-mediated polymyositis. The fibrosing process is best treated with the same approach used for scleroderma skin disease.

All patients need emotional and physical support. Scleroderma is a painful disease, and managing pain is essential for reducing depression and improving quality of life. This may require the use of sleep aids (e.g., lorazepam, zolpidem) and/or intermittent use of narcotics. Physical and occupational therapy is most important early in the course of the disease to reduce pain and improve activities of daily living. Family counseling helps the patient cope by providing a clear understanding of what is needed at home. Adjustment of the work environment (e.g., limiting time at work, adjusting room temperature, changing work assignment) or helping to obtain disability support is also important. Open communication between the physician and the patient provides an essential element of care.

### LOCALIZED SCLERODERMA

Localized scleroderma is a nonsystemic skin disease that is seen primarily in children. It can be divided into five major types: plaque morphea, generalized morphea, bullous morphea, linear morphea, and deep morphea. Mixed forms with different types of localized scleroderma occurring at the same time are observed in about 15% of patients. The most common form of localized scleroderma is an isolated circular patch of thickened skin called plaque morphea (Figure 84-17A and B). The histology is characterized by mononuclear cellular infiltrates (particularly during active phases) and deposition of thick bundles of collagen. It is usually confined to the dermis, but occasionally it can present with local panniculitis or deeper fibrosis. Multiple morphea lesions (generalized morphea) can involve extensive areas of the skin surface and occasionally can coalesce, mimicking the skin changes of systemic sclerosis (Figure 84-17D). In some cases, the morphea lesions may appear nodular and resemble keloids (keloid morphea) or, rarely, may form subepidermal bullae (bullous morphea) (Figure 84-17C). Inactive morphea may appear as flat areas of hyperpigmented.

![Figure 84-17](image.png)

**Figure 84-17** Localized scleroderma. A, Plaque morphea active and (B) inactive. C, Keloid morphea. D, Generalized morphea. E, Linear scleroderma affecting a lower limb and (F) the face (en coup de sabre).
skin. Active plaque morphea lesions present as enlarging geographic lesions with raised violaceous borders and ivory-white sclerotic centers. Localized scleroderma can also present as a linear streak (linear scleroderma) that crosses dermatomes and is associated with tracking of fibrosis from the skin into deeper tissues, including muscle and fascia (Figure 84-17E). In severe cases, linear scleroderma causes dramatic growth deformities of bone and supporting tissues in the affected regions. Linear scleroderma that affects the face and/or scalp associated with atrophy of muscle, underlying bone, and, rarely, brain tissue is called en coup de sabre (“sword stroke”) lesion (Figure 84-17F). Progressive hemifacial atrophy (Parry-Romberg syndrome) presents with atrophy of subcutaneous tissue, muscle, and bone without skin fibrosis. A close relationship among seizures, central nervous system (CNS) abnormalities, and ocular malfunction has been noted among these subtypes. Fasciitis with deep subcutaneous sclerosis can be seen in association with morphea lesions. Pansclerotic morphea is an uncommon but aggressive and disabling type of localized scleroderma that follows a progressive course despite treatment. It involves skin and deep tissues mimicking systemic sclerosis, but typically it spares the distal portion of the extremities, and the distribution of fibrosis acquires a “tank top” pattern on the trunk.

Although localized scleroderma may be disfiguring and disabling, it is generally a self-limited process that is not associated with a systemic illness. Patients with localized scleroderma have antinuclear antibodies most often directed against histones, chromatin, or nucleosomes, suggesting the presence of a specific underlying autoimmune process. Recent studies suggest that a combination of systemic corticosteroids and methotrexate can be effective treatment for localized scleroderma. Uncontrolled trials of topical therapy with ultraviolet (UV) A-1 phototherapy suggest that this treatment can be helpful for morphea lesions.

### MIMICS OF SCLERODERMA

Several conditions that present with various degrees of skin fibrosis can be potentially confused with scleroderma (Table 84-7). These disorders have different origins and exhibit distinct clinical characteristics, skin pathology, and disease associations. They often are detected in the primary care setting, and patients are referred to rheumatologists for further evaluation.

In the early 19th century, many cases of sclerema were mistakenly thought to be scleroderma. Sclerema is characterized by thick, indurated skin secondary to collagen and mucin deposition that normally begins on the trunk, especially over the upper back and shoulders, and can spread to arms, legs, and face. Sclerema occurs in the setting of three associated conditions: It can be a transient clinical manifestation following infection; a more persistent disorder associated with insulin-dependent diabetes; or an idiopathic process associated with a monoclonal gammopathy. It is seen in 2% to 15% of diabetic patients, particularly among those with poorly controlled disease. The course of sclerema varies with the underlying associated condition. It tends to resolve in several months when secondary to an infection, and it is improved with control of the diabetic state. No therapy has proved effective, but UVA-1 treatment or PUVA (phototherapy) is reported to be of benefit.

Eosinophilic fasciitis (EF), also called diffuse fasciitis with eosinophilia or Shulman’s syndrome or fascitis-panniculitis syndrome, can also mimic scleroderma. Eosinophilic fasciitis is more common in males and presents as rapid and progressive stiffening of the arms, legs, and trunk. Inflammation and fibrosis within fascia create puckering of the skin and deep venous tracks (the “groove sign”). Because the inflammatory process is deep to cutaneous tissues, the superficial layers of the skin may be pinched readily in EF, in contrast to the thickened dermis involved in scleroderma. EF spares the fingers, is not associated with RP or nail-fold capillary abnormalities, and is not associated with systemic manifestations. Peripheral eosinophilia is present in about 80% of cases but is not required to make a diagnosis. EF associates with a number of hematologic disorders, including immune-mediated anemia or thrombocytopenia, pure red-cell aplasia, myelodysplastic syndromes, and lymphoproliferative processes (T or B cell lymphoma, multiple myeloma). A diagnosis is made by clinical examination and by full-thickness biopsy that includes the fascia. EF generally responds to corticosteroid therapy.

### Table 84-7  Spectrum of Scleroderma-like Fibrosing Skin Disorders

<table>
<thead>
<tr>
<th>Category</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immune-Mediated/Inflammatory</td>
<td>Eosinophilic fasciitis, Graft-versus-host disease, Lichen sclerosus et atrophicus, POEMS syndrome, Overlap (systemic lupus erythematosus, dermatomyositis)</td>
</tr>
<tr>
<td>Metabolic</td>
<td>Phenylketonuria, Porphyria cutanea tarda, Hypothyroidism (myxedema)</td>
</tr>
<tr>
<td>Deposition</td>
<td>Scleromyxedema, Systemic amyloidosis, Nephrogenic systemic fibrosis (or nephrogenic fibrosing dermopathy), Scleredema adulterum</td>
</tr>
<tr>
<td>Occupational</td>
<td>Polyvinyl chloride, Organic solvents, Silica, Epoxy resins</td>
</tr>
<tr>
<td>Genetic</td>
<td>Progeroid disorders (progeria, acrogeria, Werner’s syndrome), Stiff skin syndrome (or congenital fascial dystrophy)</td>
</tr>
<tr>
<td>Toxic or Iatrogenic</td>
<td>Bleomycin, Pentazocine, Carbiodopa, Eosinophilia-myalgia syndrome (L-tryptophan), Toxic oil syndrome (aniline denaturated rapeseed oil), Postradiation fibrosis</td>
</tr>
</tbody>
</table>

**POEMS, polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes.**

and can completely resolve over several months. However, a small percentage of patients will have progressive, nonresponsive disease.

Scleromyxedema (papular mucinosis) closely mimics the cutaneous manifestations of scleroderma. The skin has a flesh-colored papular eruption, giving it a cobblestone feel to light palpation. Distribution is typical with particular involvement of the glabella, posterior auricular area, and neck. The skin of the trunk and limbs also can become involved. Patients usually are between 30 and 70 years old and have an associated monoclonal gammapathy that usually is of the IgG type with lambda light chains. The presence of RP, scleroderactyly, esophageal dysmotility, pulmonary hypertension, and a myopathy can mimic features of scleroderma. Nail-fold capillary changes do not occur, and neurologic complications with encephalopathy, seizures, coma, and psychosis are reported. Scleromyxedema responds rapidly to treatment with intravenous immunoglobulin, but maintenance therapy must be continued to prevent relapse.155

Nephrogenic systemic fibrosis (NSF) is a fibrosing disorder that occurs in patients with end-stage renal failure usually on chronic dialysis or following recent renal transplantation.56 NSF usually develops over a short time (days to weeks) and then follows a chronic progressive course. This disorder is characterized by thickened, hardened skin with waxy hyperpigmentation and raised plaques, and then loss of limb function due to contractures secondary to subacute and chronic fibrosis of the skin, fasciae, and muscles. The face is usually spared. Deep fibrosis leads to flexion contractions and skin ulcerations. It is now clear that exposure to gadolinium-containing contrast agents in the setting of renal impairment triggers the disease process. Gadolinium deposits are found in involved tissues. By routine microscopy, findings range from a very subtle proliferation of dermal fibroblasts in early lesions to a florid proliferation of fibroblasts and dendritic cells in fully developed cases. Although it has been suggested that improvement can occur with improved renal function, no effective therapy has been confirmed. Reports of benefit with the use of imatinib mesylate are encouraging, but aggressive physical therapy, care to prevent skin ulcerations, pain control, and emotional support are the mainstay of therapy.

Scleroderma-like skin changes have been reported in a number of other disorders, including the carcinoid syndrome, chronic graft-versus-host disease, porphyria cutanea tarda, POEMS syndrome (polyneuropathy, organomegaly, endocrinopathy, monoclonal gammapathy, and scleroderma-like skin changes), bleomycin exposure, Werner's syndrome, and phenylketonuria. Eosinophilia-myalgia syndrome and toxic oil syndrome are conditions of historic interest resulting from ingestion of toxic contaminants and presenting with scleroderma-like features.

Selected References


Full references for this chapter can be found on www.expertconsult.com.
References


CHAPTER 84

94.
95.
96.
97.
98.
99.
100.
101.
102.

103.

104.
105.
106.
107.
108.
109.
110.
111.
112.
113.
114.

115.
116.
117.
118.

Allanore Y, Meune C, Kahan A: Outcome measures for heart
involvement in systemic sclerosis, Rheumatology (Oxford) 47(Suppl
Byers RJ, Marshall DA, Freemont AJ: Pericardial involvement in
coronary artery disease in scleroderma, Rheumatology (Oxford)
James TN: De subitaneis mortibus. VIII. Coronary arteries and conduction system in scleroderma heart disease, Circulation 50:844–856,
1974.
of coronary flow reserve in patients with systemic sclerosis, Ann
Nitenberg A, Foult JM, Kahan A, et al: Reduced coronary flow and
resistance reserve in primary scleroderma myocardial disease, Am
Kahan A, Allanore Y: Primary myocardial involvement in systemic
Meune C, Avouac J, Wahbi K, et al: Cardiac involvement in systemic
sclerosis assessed by tissue-doppler echocardiography during routine
care: a controlled study of 100 consecutive patients, Arthritis Rheum
Hachulla AL, Launay D, Gaxotte V, et al: Cardiac magnetic
resonance imaging in systemic sclerosis: a cross-sectional observational study of 52 patients, Ann Rheum Dis 68:1878–1884,
2009.
Scheja A, Bartosik I, Wuttge DM, et al: Renal function is mostly
preserved in patients with systemic sclerosis, Scand J Rheumatol
crisis in patients with diffuse scleroderma, J Rheumatol 32:649–655,
2005.
Kamen DL, Wigley FM, Brown AN: Antineutrophil cytoplasmic
antibody-positive crescentic glomerulonephritis in scleroderma—a
patient characteristics and long-term outcomes, Q J Med 100:485–
Teixeira L, Mouthon L, Mahr A, et al: Mortality and risk factors of
scleroderma renal crisis: a French retrospective study of 50 patients,
Clements PJ, Lachenbruch PA, Furst DE, et al: Abnormalities of
renal physiology in systemic sclerosis: a prospective study with
Steen VD: Autoantibodies in systemic sclerosis, Semin Arthritis
93, 1968.
Baron M, Lee P, Keystone EC: The articular manifestations of progressive systemic sclerosis (scleroderma), Ann Rheum Dis 41:147–152,
1982.
involvement and relationships with systemic inflammation in systemic sclerosis: results from the EULAR Scleroderma Trial and
Research Group (EUSTAR) database, J Rheumatol 37:1488–1501,
2010.
of the hand and wrist in systemic sclerosis, Rheumatology (Oxford)
Low AH, Lax M, Johnson SR, et al: Magnetic resonance imaging of
arthritis overlap syndrome: a unique combination of features suggests

|

Clinical Features and Treatment of Scleroderma

119.
120.
121.
122.
123.
124.
125.
126.

127.
128.
129.
130.

131.
132.

133.

134.
135.
136.
137.
138.
139.
140.

1403.e3

a distinct genetic, serological and clinical entity, Rheumatology
Jinnin M, Ihn H, Yamane K, et al: Clinical features of patients with
systemic sclerosis accompanied by rheumatoid arthritis, Clin Exp
Follansbee WP, Zerbe TR, Medsger TA Jr: Cardiac and skeletal
Kahl LE, Medsger TA Jr, Klein I: Prospective evaluation of thyroid
and thyroid autoimmune diseases, Clin Rheumatol 25:240–245,
2006.
Hoyles RK, Ellis RW, Wellsbury J, et al: A multicenter, prospective,
randomized, double-blind, placebo-controlled trial of corticosteroids
and intravenous cyclophosphamide followed by oral azathioprine for
the treatment of pulmonary fibrosis in scleroderma, Arthritis Rheum
Derk CT, Grace E, Shenin M, et al: A prospective open-label study
Liossis SN, Bounas A, Andonopoulos AP: Mycophenolate mofetil as
first-line treatment improves clinically evident early scleroderma
in diffuse cutaneous systemic sclerosis—a retrospective analysis,
as maintenance therapy in early diffuse systemic sclerosis patients
treated for 1-year with low dose cyclophosphamide pulse therapy, Clin
in systemic sclerosis associated with interstitial lung disease, Clin
Berezne A, Ranque B, Valeyre D, et al: Therapeutic strategy combining intravenous cyclophosphamide followed by oral azathioprine to
treat worsening interstitial lung disease associated with systemic sclerosis: a retrospective multicenter open-label study, J Rheumatol
of methotrexate with placebo in the treatment of systemic sclerosis:
a 24 week randomized double-blind trial, followed by a 24 week
of methotrexate versus placebo in early diffuse scleroderma, Arthritis
Morton SJ, Powell RJ: Cyclosporin and tacrolimus: their use in a
routine clinical setting for scleroderma, Rheumatology 39:865–869,
2000.
placebo-controlled trial of photopheresis in systemic sclerosis, J Am
in patients with diffuse cutaneous systemic sclerosis, Arthritis Rheum
stem cell transplantation in systemic sclerosis: procedure related
mortality and impact on skin disease, Ann Rheum Dis 60:577–584,




Inflammatory Diseases of Muscle and Other Myopathies

KANNEBOYINA NAGARAJU • INGRID E. LUNDBERG

HISTORY OF INFLAMMATORY MUSCLE DISEASES

Inflammatory muscle diseases are a heterogeneous group of systemic autoimmune rheumatic disorders characterized by chronic muscle weakness, muscle fatigue, and mononuclear cell infiltration into skeletal muscle. These disorders were described in the literature more than a century ago as generalized muscle disorders affecting principally the trunk and proximal limb muscles, with or without skin involvement. It was also recognized that these diseases can range from acute and even fatal to slow, progressive, chronic, insidious conditions, with patterns of relapse and remission. Steiner's summary of myositis cases in 1903 made a clear distinction between idiopathic polymyositis (PM) and other forms of myositis caused by bacteria and parasites, and Stertz in 1916 first reported an association between dermatomyositis (DM) and internal malignancy. At about the same time, Batten described the first case of DM with classic histologic features in a child.

Since the 1940s, it has been recognized that PM may occur in the absence of cutaneous lesions, muscle pain, or constitutional symptoms. It may present in an acute, subacute, or chronic insidious form, with some fraction of cases showing systemic features or involvement of organs and tissues. The differential diagnosis has been described independently by several investigators, and the most chronic form was differentiated from an adult variety of muscular dystrophy. Banker and Victor noted that DM in children was different and involved a greater degree of vascular inflammation and thrombosis (systemic angiopathy). The first, and still widely used, classification scheme and set of diagnostic criteria for myositis were proposed by Bohan and Peter in 1975. They include PM and DM but not the later-described subset known as inclusion body myositis (IBM). IBM was later defined by the presence of distinct histopathologic changes including vacuoles and nuclear and cytoplasmic inclusions, as well as by distinct clinical features including resistance to glucocorticoids. Debate continues whether IBM should be considered an idiopathic inflammatory myopathy (IIM). We have chosen to include information on IBM in this chapter because it is clinically relevant to the differential diagnosis of PM.

EPIDEMIOLOGY

The actual annual incidence of inflammatory myopathy is currently unknown. Because these diseases are so rare, no large-scale epidemiologic studies have been reported; however, several retrospective studies have reported an annual incidence of less than 10 per million individuals (Table 85-1). This may be an overestimate, given that the Peter and Bohan diagnostic criteria used in these studies did not distinguish IBM as a separate disease entity. The prevalence of IBM has been estimated to be 10.7 per million in the United States, 9.3 per million in Australia, and 4.9 per million in the Netherlands. The age-adjusted prevalence of IBM for those older than 50 years was reported as 16 to 35 per million. In some geographic areas, IBM appears to be the most common acquired progressive myopathy, representing 16% to 28% of all inflammatory myopathies. There may be referral biases in these studies. The incidence-prevalence studies need to be interpreted cautiously, given that most have not reported confidence intervals for their rates.

The incidence of the various myopathies varies according to ethnicity, age, and gender. Some studies have reported that the incidence of PM is higher in black patients than in white patients. IIMs can occur in any age group, from early childhood to late in adult life. The onset of PM is usually in the late teens or older, with the mean age at onset being 50 to 60 years; DM shows two peaks—5 to 15 years and 45 to 65 years. IBM is commonly seen in individuals older than 50 years and is rare in younger adults. Some studies have reported gender-specific incidence rates. For example, in the case of PM and DM, females are more commonly affected than males (ratio > 2:1), whereas in IBM, the converse is true (again, >2:1 ratio).

Inflammatory myopathies can occur in association with other autoimmune connective tissue diseases such as scleroderma, systemic lupus erythematosus (SLE), rheumatoid arthritis, Sjögren’s syndrome, polyarteritis nodosa, and sarcoidosis. Significant proportions of all myositis patients (11% to 40%) have an associated connective tissue disease. Several studies have also confirmed an association between malignancies and inflammatory myopathies.
A 2-decade-long retrospective study of myositis patients revealed about 12% (37/309) of the patients are associated with malignant diseases. A majority (81%) (30/37) of these patients had DM, and the remainder (19%) had PM. In about 68% of these cases malignancy and myositis appeared within 1 year. The most frequent malignancies associated with myositis were breast tumors and adenocarcinomas, and successful treatment of the underlying malignant disease improved the clinical disease course of myositis. Overall survival rate was considerably worse for DM compared with other forms of myositis. Similar observations were made in a nationwide cohort study in Taiwan that indicated although the general risk of cancer was increased in IIM patients, cancers of the nasopharynx, lung, and breast tissue were the most likely to be diagnosed in both PM and DM patients. Detection of malignancies most frequently occurred within 1 year of being diagnosed with an IIM, with the likelihood of developing a malignancy decreasing over time. Overall, the frequency of malignancies varies widely (4% to 42%) in different studies, but in general the incidence of malignancy is higher in DM patients than in IBM or PM patients. It is difficult to determine the relative risks for a particular malignancy because a variety of malignancies are associated with myositis, and only small numbers of individual malignancies have been reported in any one study.

### ETIOLOGY OF MYOSITIS

#### Genetic Risk Factors

An association with immune response genes and occasional reports of familial clustering of myositis support the role of genetic factors in these diseases. Polymorphisms in human leukocyte antigen (HLA) class I and II genes are known genetic risk factors for several autoimmune diseases including myositis, but the mechanisms for these associations remain unclear. One possibility is that because the gene products influence T cell repertoire development, tolerance, and immune responses to foreign agents, certain polymorphisms may be selected on the basis of environmental triggers. It appears that haplotypes HLA-DRB1*0301 and HLA-DQA1*0501 are the strongest known genetic risk factors for all forms of myositis in whites; however, different phenotypes have additional HLA risk and protective factors. It has been shown that in African-American patients neither DRB1*0301 nor DQA1*0501 is strongly associated with myositis. Instead, the HLA-DRB1*08 allele shows the highest general risk for developing myositis, whereas the HLA-DRB1*14 allele is strongly protective in African-American patients. The HLA-B8/DR3/DR52/DQ2 haplotype is found in a significant proportion of IBM patients. The risk and protection conferred by HLA associations differ significantly among different ethnic and serologic groups. For example, in some populations (e.g., Koreans, Mesoamericans), there is no association with HLA genes. Further, HLA-DRB1*0301, which is a risk factor in whites, is a protective factor in the Japanese population. The HLA-DRB1*0301, HLA-DQA1*0501, and HLA-DQB1*0201 alleles are strongly associated with myositis-specific antibodies in PM patients. Mechanistic data supporting the role of HLA molecules in disease pathogenesis are, unfortunately, lacking at present. Some studies have reported that maternally derived chimeric cells are present in the peripheral blood and muscle tissues of juvenile DM patients, suggesting that HLA alleles control the occurrence of chimerism and explain the HLA association found in these disorders. Like other autoimmune disease conditions, myositis is a complex multigenic disorder involving other non-HLA immune response genes (e.g., cytokines and receptors including tumor necrosis factor [TNF], interleukin [IL]-1, and tumor necrosis factor receptor [TNFR]-1); complement components (e.g., C4, C2); immunoglobulin heavy-chain allotypes, and T cell receptors. The exact contribution of the genetic component in these disorders is currently unknown, in part because of their rarity, the small number of subjects in any single cohort, and the heterogeneity in disease phenotype. International collaborative efforts are currently under way to address these issues and to identify potential genetic and environmental risk factors in myositis.

#### Environmental Risk Factors

The temporal association of myositis onset and environmental agents in certain individuals suggests that specific exposures in the context of certain genetic backgrounds can initiate muscle inflammation. Common environmental agents implicated in myositis include infectious organisms such as viruses and bacteria and noninfectious agents such as drugs and food supplements (Table 85-2). For example, enteroviruses (influenza, coxsackievirus, echoviruses) and retroviruses (human T-lymphotropic virus-I) are known to induce muscle inflammation. The myositis associated with enteroviruses usually occurs in children and is generally self-limited. A viral cause is strengthened by the presence of high-titer antiviral antibodies and viral particles in patients’ serum and tissue samples, as well as the induction of muscle inflammation by enteroviruses in animal models. Attempts to identify virus in the tissues of IIM patients by sensitive techniques such as polymerase chain reaction have failed, leading to doubts about the viral cause of these diseases and ruling out continual viral infection as a cause of the ongoing muscle inflammation in these patients. However, it is possible that viruses initially trigger the disease process before being eliminated by the host’s immune response, thus explaining the absence of viral genomes in the myositis muscle tissue. Similarly, some microorganisms such as staphylococci, clostridia, and mycobacteria are known to affect skeletal muscle and cause acute muscle inflammation, but there is no evidence that these organisms actually cause chronic, self-sustaining muscle inflammation.
It appears that malignancy is an additional risk factor for the development of myositis, and there is a strong association between DM and malignancies. This early clinical observation has been confirmed in epidemiologic studies.\textsuperscript{62,57} With regard to PM and IBM, the association with malignancy is less convincing. The increased risk of malignancy associated with DM has been established both at the time of DM diagnosis and more than 10 years after diagnosis. The pathophysiologic mechanism for the association between malignancy and DM has not been clarified, but there could be several explanations. The strong association between malignancy and the onset of DM indicates that the latter could be a paramalignant phenomenon; that is, the development of myositis is a consequence of the malignancy (related to autoantigens), or the malignancy and DM share disease mechanisms. Thus the molecular mechanisms underlying this unique association are currently unclear. However, there is some evidence that removal of a tumor sometimes results in amelioration of muscle weakness, and tumor reappearance sometimes coincides with muscle weakness, suggesting that these two are linked.\textsuperscript{28} A recent report has shed some light on this connection by showing that myositis-specific antigens are highly expressed in cancer tissues, as well as in regenerating muscle cells of myositis patients.\textsuperscript{58,59} The authors propose that in cancer-associated myositis, an autoimmune response directed against cancer cross-reacts with regenerating muscle cells, enabling a feed-forward loop of tissue damage and antigen selection.\textsuperscript{60} This association must be explored further because cancer-associated myositis patients almost never develop myositis-specific autoantibodies, which are protective for the development of cancer. For malignancies that develop during established disease, the potential explanations include the presence of chronic inflammation or prolonged immunosuppressive treatment, which could contribute to the development of malignancy.

A recent report noted a novel association of myositis with hypertension, diabetes, and ischemic heart disease. The prevalence of hypertension and diabetes in this population was 62\% and 29\%, respectively, considerably higher than the background prevalence of 9.4\% and 4\%. These authors suggest that hypertension and ischemic heart disease were more likely to be present before the diagnosis of myositis, whereas hypertension and diabetes occurred more frequently following the diagnosis of myositis in DM patients in comparison with PM or IBM patients, suggesting that it is essential to perform a comprehensive assessment of vascular risk factors in these patients.\textsuperscript{61} The same group also reported that patients with IIM are at 75\% increased risk for mortality, and cardiovascular diseases followed by infection and malignancy account for the commonest causes of death.\textsuperscript{61}

### Mimics of Myositis

A variety of insults induce the clinical and pathologic spectrum that mimics myositis in some individuals (see Table 85-2). A number of drugs are known to cause a myopathy that closely mimics myositis. For example, n-penicillamine causes clinically and histologically indistinguishable IIM.\textsuperscript{62} Likewise, commonly used lipid-lowering drugs such as statins (e.g., atorvastatin, lovastatin) can cause a

<table>
<thead>
<tr>
<th>Table 85-2 Possible Environmental Risk Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infectious Agents</strong></td>
</tr>
<tr>
<td>Viruses</td>
</tr>
<tr>
<td>Picornavirus family, enteroviruses</td>
</tr>
<tr>
<td>Polio, coxsackievirus types A and B, echoviruses</td>
</tr>
<tr>
<td>Retroviruses</td>
</tr>
<tr>
<td>HIV-1, HTLV-I</td>
</tr>
<tr>
<td>Parvovirus B19</td>
</tr>
<tr>
<td>Hepatitis C virus</td>
</tr>
<tr>
<td>Hepatitis B virus</td>
</tr>
<tr>
<td>Bacteria</td>
</tr>
<tr>
<td>Staphylococci</td>
</tr>
<tr>
<td>Clostridia</td>
</tr>
<tr>
<td>Mycobacteria</td>
</tr>
<tr>
<td>Parasites</td>
</tr>
<tr>
<td>Toxoplasma gondii</td>
</tr>
<tr>
<td>Trypanosoma cruzi</td>
</tr>
<tr>
<td>Borrelia burgdorferi</td>
</tr>
<tr>
<td><strong>Noninfectious Agents</strong></td>
</tr>
<tr>
<td>Drugs</td>
</tr>
<tr>
<td>n-Penicillamine</td>
</tr>
<tr>
<td>Corticosteroids</td>
</tr>
<tr>
<td>Chloroquine</td>
</tr>
<tr>
<td>Statins (atorvastatin, lovastatin, pravastatin, simvastatin)</td>
</tr>
<tr>
<td>Lipid-lowering fibrates (bezafibrate, clofibrate, gemfibrozil)</td>
</tr>
<tr>
<td>L-Tryptophan</td>
</tr>
<tr>
<td>Biologic agents (e.g., growth hormone, interferon-α, interleukin-2)</td>
</tr>
<tr>
<td>Vaccination for tetanus, BCG, diphtheria, hepatitis B, hepatitis A</td>
</tr>
<tr>
<td>Miscellaneous drugs (e.g., local anesthesia, hydroxyurea, leuprolide acetate)</td>
</tr>
<tr>
<td>Ultraviolet radiation exposure</td>
</tr>
<tr>
<td>Miscellaneous agents (e.g., silicone breast implants, chronic graft-versus-host disease associated with bone marrow transplantation, collagen injection, silicone exposure)</td>
</tr>
</tbody>
</table>

BCG, bacille Calmette-Guérin; HIV, human immunodeficiency virus; HTLV-I, human T-lymphotropic virus I.
myopathy that resembles inflammatory myositis. These agents inhibit 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, a rate-limiting enzyme involved in the conversion of HMG-CoA to mevalonic acid, thereby preventing the synthesis of bioactive sterol and nonsterol metabolic intermediates in the cholesterol synthetic pathway. The mechanism by which these drugs cause myopathy is not clear yet. However, a recent study showed anti-HMGCR antibodies in patients with statin-induced autoimmune myopathy. It has been suggested that statins upregulate HMG-CoA autoantigen in regenerating muscle cells that in turn sustain an autoimmune response even after statin withdrawal, providing a mechanism for statin-induced immune-mediated necrotizing myopathy. Other drugs such as hydroxyurea can cause skin rashes that resemble DM. TNF inhibitors have been associated with the onset of autoimmune disorders such as vasculitis and a lupus-like syndrome. Recent reports implicate that TNF inhibitors in inflammatory arthritis patients may either induce or exacerbate DM or anti-Jo-1 positive PM. Some other reports point to the vaccine adjuvant aluminum hydroxide as a cause of macrophagic myofasciitis. The histology shows infiltration by macrophages and some CD8 T cells into the endomysium, perimysium, and epimysium, together with clinically elevated creatine kinase (CK) levels, muscle weakness, myalgias, fatigue, and arthralgias. Despite some reports of vaccine-induced myositis, systematic investigation has failed to link any vaccine to myositis.

**PATHOGENESIS**

Significant advances have been made in our understanding of the pathogenesis of the human inflammatory myopathies. It is generally thought that IIMs are autoimmune in origin because they are frequently associated with other autoimmune diseases (e.g., Hashimoto’s thyroiditis) and collagen vascular diseases (e.g., scleroderma); many patients exhibit an autoantibody response including the presence of myositis-specific autoantibodies; some studies provide evidence for lymphocyte-mediated muscle fiber injury; and a favorable response to immunosuppressive therapies in some patients supports an autoimmune cause of these disorders.

**Humoral Immune Response**

More than 50% of all IIM patients have uniquely defined autoantibodies—some of which are specific to myositis, and some of which are merely associated with myositis. These are generally referred to as myositis-specific autoantibodies (MSAs) and myositis-associated autoantibodies (MAAs), respectively. MAAs include autoantibodies to various nuclear and cytoplasmic antigens. Antinuclear antibodies (ANAs) present in myositis are not particularly associated with any disease subgroup, whereas MSAs that are directed against antigens of the protein synthesis pathway (e.g., aminoacyl–transfer RNA [tRNA] synthetases and signal recognition particles) and nuclear components (e.g., nuclear helicase [Mi-2]) are often associated with distinct clinical disease groups and subgroups (e.g., tRNA synthetases with interstitial lung disease, Mi-2 with DM) (Table 85-3).

**Table 85-3** Myositis-Specific Antibodies

<table>
<thead>
<tr>
<th>Autoantibodies</th>
<th>Clinical Disease/Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antisynthetase autoantibodies*</td>
<td>More common in polymyositis than dermatomyositis, interstitial lung disease, arthritis, Raynaud’s phenomenon, fevers, mechanic’s hands</td>
</tr>
<tr>
<td>Signal recognition particle (SRP)*</td>
<td>Polymyositis; possible severe disease and cardiac involvement</td>
</tr>
<tr>
<td>Chromodomain helicase DNA binding proteins 3 and 4 (Mi-2x and β)*</td>
<td>Dermatomyositis</td>
</tr>
</tbody>
</table>

*Common antisynthetase antibodies found in myositis are targeted to histidyl-tRNA synthetase (Jo-1), threonyl-tRNA synthetase (PL-7), alanyl-tRNA synthetase (PL-12), isoleucyl-tRNA synthetase (OJ), glycyrl-tRNA synthetase (EJ), and asparaginyl-tRNA synthetase (KS).
†Autoantibodies commonly bind to a 54-kD SRP protein in the U.S. patient population and 72-, 54-, and 9-kD proteins in the Japanese population.
‡Targets a 240-kD helicase protein that is part of the nucleosome remodeling deacetylase complex.

Anti–histidyl-tRNA synthetase antibodies are the most frequent and are present in about 16% to 20% of myositis patients. Antibodies against other aminoacyl-tRNA synthetases such as threonyl-tRNA synthetase (PL-7), alanyl-tRNA synthetase (PL-12), isoleucyl-tRNA synthetase (OJ), glycyrl-tRNA synthetase (EJ), and asparaginyl-tRNA synthetase (KS) are found less frequently (1% to 3%). Anti–Mi-2 antibodies are strongly associated with DM, with prominent features such as Gottron’s papules, heliotrope rash, the V sign, and the shawl sign. An individual usually has only one MSA because MSAs are often mutually exclusive. The MSAs are most common in patients with other autoimmune diseases and are infrequent or absent in IBM patients and those with malignancies, muscular dystrophies, or other myopathies. These antibodies are sometimes present before the onset of clinical disease.

MAAs such as PM-Scl are frequently associated with a characteristic overlap syndrome that includes features of scleroderma. This syndrome is characterized by mild muscle disease, prominent arthritis, and limited skin involvement; it frequently responds to therapy. Some myositis patients also have other MAAs such as anti-snRNP, anti-Ro/SSA, anti-Ku, and anti-PMS1. Antibodies recognizing an uncharacterized 56-kD large nuclear ribonucleoprotein have been found in a majority of myositis patients (86%), and the antibody titer appears to vary with disease activity, suggesting its importance in our understanding of disease pathogenesis and its potential usefulness as a clinical disease marker. Some of the MSAs show strong immunogenetic associations; for example, antibodies against aminoacyl-tRNA synthetases are associated with HLA-DQA1*0501, anti-SRP with DR5, anti–Mi-2 with DR7, and anti–PM-Scl with DR3. Neither the molecular mechanisms that initiate and perpetuate the autoimmune response nor the precise role of these autoantibodies in the pathogenesis of myositis is currently known. However, these antibodies serve as excellent clinical markers and can help diagnose and categorize these heterogeneous disorders into homogeneous subgroups.
Cell-Mediated Immune Response

At the cellular level, there are distinct differences in the distribution and location of the various lymphocyte subsets in the muscle tissues in different IIMs. Two major patterns of inflammatory cell infiltrates are seen in muscle tissue. The first has a predominantly perivascular distribution (Figure 85-1A), often in perimysial areas (Figure 85-1C), and is largely made up of CD4+ T cells, macrophages, and dendritic cells. Occasionally, B cells are present in some patients. This pattern is seen mainly in DM patients with skin rash but occasionally in patients without a rash. The second pattern has a predominantly endomysial distribution (Figure 85-1B), with mononuclear inflammatory cells often surrounding and sometimes invading non-necrotic muscle fibers. These inflammatory cellular infiltrates are comprised primarily of CD8+ T cells and macrophages, but CD4+ T cells and dendritic cells are also present. This pattern is generally seen in patients without skin rashes and often in those classified as having PM or IBM. In some patients, the two patterns of inflammation are seen in the same biopsy. The two distinct locations and the varying compositions of the inflammatory cell populations in the two areas suggest two different pathogenic mechanisms—one that targets the blood vessels and one that targets the muscle fibers. Notable inflammation is also seen in other organs.

The vascular involvement in patients with DM is also manifested in the skin and can be seen clinically in the form of nail-fold changes and changes in the gastrointestinal (GI) tract. The capillaries show clear hyperplasia, vacuolization, and necrosis, contributing to an ischemia that could cause fiber damage. One of the earliest events in the pathogenesis of DM appears to be activation of the complement cascade. This leads to the subsequent deposition of complement components, which in turn results in the deposition of lytic membrane attack complexes in the endothelial cells and the eventual loss of capillaries due to complement-mediated damage. The capillaries are abnormally thickened and enlarged and look like high endothelial venules, which are characteristics of vessels that facilitate lymphocyte trafficking (Figure 85-2). The capillaries also show signs of neovascularization. This loss of capillaries results in some of the histopathologic features characteristic of this disease: capillary necrosis and loss, perivascular inflammation and ischemia (rarely seen), and perifascicular atrophy (a late feature; see Figure 85-1C and D). Recent studies also point out a role for type I interferon (IFN)-inducible genes in the pathogenesis of DM. It has been shown that plasmacytoid dendritic cells produce type I IFN and induce expression of IFN-inducible proteins such as MxA and IFN-inducible gene 15 (ISG15) at perifascicular myofibers and capillaries of DM biopsies, suggesting that injury to muscle fibers and capillaries occurs due to the intracellular overproduction of one or more type I IFN-inducible proteins in DM.

The endomysial inflammatory aggregates contain a high percentage of T cells, particularly activated CD8+ T cells, macrophages, and CD4+ T cells and few natural killer cells. Immunoelectron microscopic studies have provided evidence of the invasion, replacement, and probable destruction of non-necrotic muscle fibers by T cells and macrophages. It is suggested that CD8+ cytotoxic T lymphocytes (CTLs) recognize major histocompatibility
In contrast to previous concepts, recent studies also show accumulations of B cells, plasma cells, myeloid dendritic cells, late-activated macrophages expressing 25F9 marker, as well as CD8$^+$CD28$^-$ and CD4$^+$CD28$^-$ T cells (TCR V$\beta$-expanded T cells) in the skeletal muscle and in the peripheral circulation of PM, DM, and IBM patients. These CD28$^-$ cells and late-activated macrophages expressing 25F9 marker are hypothesized to exhibit cytotoxic potential and produce proinflammatory cytokines in IIM skeletal muscle.

Another study also explored the potential role of FOXP3$^+$ Treg cells in the pathology of myositis. These authors reported that the number of Treg cells correlated with the degree of inflammation in IIMs and suggested that these cells might serve to counterbalance activity of cytotoxic T cells in myositis.

On the basis of the data described, two different pathways have been proposed as major mediators of muscle damage and inflammation: one mediated through T lymphocytes (CTLs) directed against muscle fibers, predominating in PM and IBM, and the other directed against complex (MHC) class I on muscle fibers and may mediate muscle fiber damage. Infiltrating CTLs express perforin-containing granules, which are characteristically oriented toward the target muscle fiber, indicating that muscle fiber injury may be partially mediated by perforin-dependent cytotoxic mechanisms (Figure 85-3B and C). In PM and IBM, there is evidence of clonal proliferation of CD8$^+$ T cells, both within the muscle and in the peripheral circulation. T cell lines from patients demonstrate cytotoxicity against autologous myotubes, suggesting that the muscle fiber injury in PM and IBM is mediated by CTLs. CTLs are known to mediate target cell damage by both perforin–granzyme B and Fas-FasL pathways. The overexpression of antiapoptotic molecules such as Bcl-2, Fas-associated death domain–like IL-1 converting enzyme inhibitory protein (FLIP), and human inhibitor of apoptosis protein–like protein in skeletal muscle of myositis patients suggests that perforin–granzyme B–mediated CTL damage may play a predominant role in muscle fiber injury and dysfunction in myositis.

In contrast to previous concepts, recent studies also show accumulations of B cells, plasma cells, myeloid dendritic cells, late-activated macrophages expressing 25F9 marker, as well as CD8$^+$ CD28$^-$ and CD4$^+$ CD28$^-$ T cells (TCR V$\beta$-expanded T cells) in the skeletal muscle and in the peripheral circulation of PM, DM, and IBM patients. These CD28$^-$ cells and late-activated macrophages expressing 25F9 marker are hypothesized to exhibit cytotoxic potential and produce proinflammatory cytokines in IIM skeletal muscle.

Another study also explored the potential role of FOXP3$^+$ Treg cells in the pathology of myositis. These authors reported that the number of Treg cells correlated with the degree of inflammation in IIMs and suggested that these cells might serve to counterbalance activity of cytotoxic T cells in myositis.

On the basis of the data described, two different pathways have been proposed as major mediators of muscle damage and inflammation: one mediated through T lymphocytes (CTLs) directed against muscle fibers, predominating in PM and IBM, and the other directed against...
vessels, predominating in DM. However, several studies have shown that the degree of inflammation does not consistently correlate with the severity of the structural changes in the muscle fibers or with the severity of the clinical disease, suggesting that nonimmune processes also play a role in disease pathogenesis. A role for nonimmune processes is supported by the following observations: First, marked structural changes in the muscle fibers occur in the absence of any inflammatory cells. Second, there is a lack of correlation between the degree of inflammation and the degree of muscle weakness. Third, some myositis patients do not respond, even to powerful anti-inflammatory therapy. Fourth, steroid treatment may eliminate inflammatory cells in myositis muscle tissue, but this removal alone may not substantially improve the clinical disease, suggesting that immunosuppressive therapies modulate disease activity but do not change other mediators of the disease process. Finally, the clinical disease may progress when identifiable inflammation has subsided, suggesting a role for nonimmune mechanisms in the pathogenesis of myositis. Thus the exact contribution of immune-mediated pathways to muscle damage is currently unknown.

### Class I Major Histocompatibility Complex

Normal skeletal muscle cells do not constitutively express or display MHC class I molecules, although they can be induced to do so by proinflammatory cytokines such as IFN-γ or TNFα, or by the alarmin HMGB1. In contrast, in human IMs, the early and widespread appearance of MHC class I in non-necrotic muscle cells is a striking feature, even in muscle cells distant from the lymphocytic infiltration. MHC class I staining is usually observed on the sarcolemma of muscle fibers, but some fibers also show staining in both the sarcolemma and the sarcoplasm (see Figure 85-3A and B). In some patients, the expression is restricted to a few clusters (often in early disease), whereas in others, almost every fiber is positively stained, particularly in late-phase and treatment-resistant cases. The biologic significance of these observations has been explored by generating a conditional transgenic mouse model overexpressing syngenic mouse MHC class I. The overexpression of MHC class I molecules in the skeletal muscle of mice results in the development of clinical, biochemical, histologic, and immunologic features that resemble human myositis and provides a close model of the human disease. The disease in these mice is inflammatory, limited to skeletal muscles, self-sustaining, more severe in females, and often accompanied by MSAs. Recent studies in this model further suggest that MHC class I overexpression leads to endoplasmic reticulum stress, muscle atrophy, and decrease in force generation capacity of skeletal muscle, implicating the role for MHC class I muscle weakness in myositis.

A number of observations in human myositis patients and in the mouse model of myositis suggest that MHC class I molecules themselves may mediate muscle fiber damage and dysfunction in the absence of lymphocytes. For instance, in human myositis, the induction of MHC class I antigen in muscle fibers occurs early, preceding inflammatory cell infiltration. MHC class I staining of human myositis biopsies shows both a cell surface and a sarcoplasmic reticulum pattern of internal reactivity, demonstrating that some of the MHC class I molecules may be retained in the endoplasmic reticulum (ER) of these fibers. Persistent MHC class I overexpression in muscle fibers may exist in the absence of an inflammatory infiltrate. The controlled induction of MHC class I in the mouse model is followed by muscle weakness before mononuclear cell infiltration. It has recently been shown that in vivo gene transfer of MHC class I plasmids attenuates muscle regeneration and differentiation. Together, these observations, and particularly the obvious retention of MHC class I within the cell in both human and murine disease, indicate that the muscle fiber damage seen in myositis may not be solely mediated by immune attack (e.g., CTLs, autoantibodies); it may also be mediated through nonimmunologic mechanisms such as the ER stress response and hypoxia.

Because MHC class I assembly occurs in the ER and because upregulation in myositis muscle fibers is widespread, even in the absence of visible inflammatory infiltrate, it is likely that ER stress plays a role in the muscle fiber damage and dysfunction associated with human myositis. The ER is intimately involved in the folding, exporting, and processing of newly synthesized proteins. When there is an imbalance between the protein load in the ER and the cell’s ability to process that load, a series of signaling pathways that adapt cells to ER stress is activated. This ER stress response can be provoked by a variety of pathophysiologic conditions including ischemia, hyperhomocysteinemia, viral infections, and mutations that impair protein folding, as well as by excess accumulation of protein in the ER. Cells self-protect against ER stress by initiating at least four functionally distinct responses: (1) upregulation of the nuclear factor κB (NFκB) pathway (ER overload response); (2) upregulation of genes encoding ER chaperone proteins such as Bip/GRP78 and GRP94, as a means of increasing protein folding activity and preventing protein aggregation; (3) translational attenuation to reduce the load of protein synthesis and to prevent the further accumulation of unfolded proteins (unfolded protein response); and (4) cell death, which occurs when the ER’s functions are severely impaired. This cell death event is mediated by transcriptional activation of the gene for CHOP/GADD153, a member of the C/EBP family of transcription factors, and by the activation of ER-associated caspase 12.

In myositis, it appears that overexpression of MHC class I in myofibers initiates a series of cell autonomous changes that contribute to myofiber pathology. Recent investigations have indicated that overexpression of MHC class I on muscle fibers results in activation of the NFκB and ER stress response pathway in human inflammatory myopathies and in the mouse model of myositis. NFκB can be activated within minutes by a variety of stimuli including inflammatory cytokines such as TNF and IL-1, T cell activation signals, and stress inducers. It is likely that in human myositis, NFκB activates both classic (proinflammatory cytokines) and nonclassic (ER stress response) pathways.

Further, there is evidence that downstream target genes (e.g., MHC class I, intercellular adhesion molecule [ICAM], monocyte chemoattractant protein [MCP]-1) regulated by the NFκB pathway are highly upregulated in myositis patients. Recent studies have indicated that NFκB p65 is activated both in human myositis biopsies and in the
Cytokines and Hypoxia

A number of other effector molecules produced in muscle tissue by inflammatory cells, endothelial cells, and muscle fibers are thought to play a role in the pathogenesis of myositis. Most of the data assembled relate to cytokines, but some data related to chemokines are also available. The most consistently demonstrated cytokines in muscle tissue from patients with IIMs are cytokines with proinflammatory properties: IL-1α, IL-1β, TNF, and IFN-α. Recently, IL-10, IL-13, epidermal growth factor (EGF), vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF), CCL3 (macrophage inflammatory protein [MIP]-1α), CCL4 (MIP-1β) and CCL11 (eotaxin), IL-15, and IL-15RAα were also demonstrated to be significantly upregulated and granulocyte colony-stimulating factor (G-CSF) downregulated in patients with IIMs relative to normal subjects. Further, the DNA-binding high mobility group box 1 (HMGB1) was found to exhibit both extranuclear and extracellular patterns in the muscle tissue of patients with PM and DM. Stimulation with IFN-γ showed an increased HMGB1 expression in muscle nuclei and the myoplasm. Exposure to HMGB1 induced a reversible upregulation of MHC class I in the muscle fibers and irreversible decrease in Ca²⁺ release from the sarcoplasmic reticulum during fatigue, implicating a role of HMGB1 and MHC class I early in the pathogenesis of IIMs. In addition to inducing the upregulation of MHC class I and II molecules on muscle fibers, cytokines may have a direct effect on muscle fiber function, as has been demonstrated for TNF. The relative importance of the various cytokines and chemokines in patients with myositis is still uncertain, but these molecules are potential biomarker candidates in this disease as exemplified by a recent study showing serum IL-6 production and the type I IFN gene signature in the peripheral blood correlating with disease activity in DM patients.

Microvessel involvement was first observed in DM but has also become evident in PM. The endothelial cells in both subsets show increased expression of adhesion molecules and proinflammatory cytokines such as IL-1α. This phenotype can be induced by tissue hypoxia, which may result from capillary loss and local tissue inflammation. Muscle tissue hypoxia can contribute to the clinical symptoms and muscle fatigue and might be associated with disease mechanisms in inflammatory myopathies. A recent study reported the expression of VEGF receptor in muscle fibers and HIF-2α reactivity in endothelial cells of PM and IBM patients DM patients showed hypoxia-inducible factor (HIF)-1α and HIF-1β expression in endothelial cells, whereas expression of HIF-2α, erythropoietin receptor, VEGF, and VEGF-R were also observed on muscle fiber. These observations suggest that deprivation of blood supply by immune-mediated mechanisms might trigger the upregulation of hypoxia-related proteins as an adaptive response. The hypoxia hypothesis is further supported by the clinical improvement observed after exercise, but a causal connection still needs to be established. In addition, magnetic resonance spectroscopic analysis, before and after a work load, has demonstrated reduced levels of energy substrates that are important for muscle contraction such as adenosine triphosphate and phosphocreatine, when compared with levels in healthy individuals. This finding supports the hypothesis that an acquired metabolic disturbance occurs in chronic inflammatory myopathies and that this disturbance can contribute to impaired muscle performance.

Proposed Mechanisms of Muscle Damage

Currently available data suggest that both immune (cell-mediated and humoral) and nonimmune (ER stress, hypoxia) mechanisms play a role in muscle fiber damage and dysfunction in myositis. ER stress, hypoxia, and the NFκB pathway are highly active within the skeletal muscle of myositis patients, and the proinflammatory NFκB pathway connects the immune and nonimmune components contributing to muscle damage. The relative contribution of...
each of these pathways to muscle fiber damage is presently unclear (Figure 85-5). Therefore use of specific drugs to inhibit these pathways, either alone or in combination, would help define their roles in myositis and potentially serve as effective therapeutic agents.

**CLINICAL FEATURES**

The inflammatory myopathies may occur as distinct disease entities, or they may coexist with some other rheumatic disease. This observation is true for all three subsets of myositis, but it is most often seen in PM and DM. The rheumatic diseases most often associated with inflammatory myopathies are systemic sclerosis, mixed connective tissue disease, Sjögren’s syndrome, and SLE; however, rheumatoid arthritis may also be associated with inflammatory myopathies. IBM may be associated with Sjögren’s syndrome, SLE, and other autoimmune diseases. Because the clinical features of IBM differ somewhat from those of PM and DM, they are presented separately.

**Polymyositis and Dermatomyositis**

The predominant symptoms in patients with PM or DM are muscle weakness and low muscle endurance. The weakness is most pronounced in proximal muscle groups—typically in the neck, pelvic, thigh, and shoulder muscles—with a symmetric distribution. Patients generally experience more problems performing repetitive movements than with single-strength exercises, and they report difficulty walking uphill or upstairs, working with their arms above their shoulders, or rising from chairs. The onset is often subacute, occurring over a few weeks, or it may be insidious, developing over several months. If untreated, the muscle weakness progresses slowly, and in the most severe cases, patients may become wheelchair dependent. Problems with swallowing and nutrition may occur as a result of impaired contractility of the throat muscles, possibly leading to aspiration pneumonia. In rare cases patients develop difficulty breathing because of weakness of the diaphragm or thoracic muscles, and they may require assisted ventilation. Other striated muscles may be involved, such as in the lower part of the esophagus (causing reflux problems) or the sphincter ani (causing incontinence).

**Skin**

Dermatomyositis is characterized by the presence of certain types of rashes; the same types are often seen in both children and adults. The most specific skin manifestations are Gottron’s papules and the heliotrope rash (Figure 85-6). Gottron’s papules are slightly elevated violaceous, pink, or dusky red papules located over the dorsal side of the metacarpal or interphalangeal joints. These papules may also occur over the extensor side of the wrist, elbow, or knee joints. Gottron’s papules are considered to be pathognomonic of DM. A macular rash (without papules) with the same distribution as Gottron’s papules is called Gottron’s sign (see Figure 85-6C and D). The heliotrope rash is a peri-orbital red or violaceous erythema of one or both eyelids, often with edema (see Figure 85-6B). Linear erythema overlying the extensor surfaces of joints is also relatively specific to DM (Figure 85-7A). Many patients with DM have photosensitive rashes, typically located on the face or scalp or over the neck (the so-called V sign), although this rash is not specific to DM (Figure 85-7B and C). Another common rash in DM is located over the shoulders (shawl sign; Figure 85-7D) or over the hips (holster sign). Pruritus is common. Patients with DM often have skin lesions on their fingers such as periungual erythema, nail-fold telangiectasias, and cuticular overgrowth (Figure 85-8C). Other less common skin manifestations are panniculitis, livedo reticularis, and nonscarring alopecia. Vasculitis may be seen in children with DM but rarely in adults.

In general, the skin rash is moderate, with local erythema. In rare cases, a severe, diffuse erythema (erythroderma) may occur, occasionally with vesiculobullous lesions or ulcers. The skin rash may precede the muscle symptoms by months or even years, and in some patients, the skin manifestations may be the only clinical sign of DM; this condition is often called amyopathic DM or DM sine myositis (see later). The pattern of the rash over the knuckles and dorsum of the hand is distinct, in that the rash generally affects the phalanges but spares knuckles in SLE, and vice versa in DM (Figure 85-8A and B). However, no histopathologic skin features are specific for DM; most of the features are also seen in patients with SLE. Thus skin biopsy is rarely helpful in distinguishing between these two disorders. The cutaneous manifestations may fail to respond to...
immunosuppressive treatment, despite improvement in muscle symptoms. Thus it is possible that different molecular pathways or disease mechanisms cause the skin rash and the muscle inflammation.

Calcinosis, which can be severe, is found mainly in juvenile DM but is occasionally seen in adults. The calcinosis occurs predominantly in sites that have been subject to friction or trauma such as the elbows or knees. Sometimes the calcinosis can be extensive and erupt, leading to ulcers. It is most often localized to the subcutaneous tissue but can also develop in the skin, fascia, or muscle and can be visualized by radiography, computed tomography (CT), or magnetic resonance imaging (MRI). The calcinosis seems to progress as long as there is active inflammatory disease. Also, once it has developed, it is often treatment resistant. Some data, however, suggest that the progress of calcinosis can be inhibited by effectively treating the inflammatory process in the skin and muscle.\(^{146}\)

Another type of skin pathology seen in inflammatory myopathies is called mechanic’s hands. This rash is often associated with the presence of antisynthetase autoantibodies and can be seen in both PM and DM. The rash is a hyperkeratotic, scaling, fissuring of the fingers, particularly on the radial side of the index fingers (Figure 85-9).

---

Figure 85-6 Characteristic features of dermatomyositis skin changes. \(A\), Gottron’s papules. \(B\), Heliotrope rash. \(C\), Gottron’s sign on knee and \(D\), elbow. (Courtesy Dr. Paul Plotz.)

Figure 85-7 Characteristic features of dermatomyositis skin changes. \(A\), Linear erythema. \(B\), Scalp rash. \(C\), V-like sign. \(D\), Shawl sign. (Courtesy Dr. Paul Plotz.)
Lungs

Lung involvement is frequent in PM and DM and is a major risk factor for morbidity and mortality. Clinical symptoms such as dyspnea and cough are common. Lung involvement can be caused by weakness of the respiratory muscles or inflammation of the lung tissue (interstitial lung disease). Weakness of the respiratory muscles may lead to restrictive lung disease, and involvement of the pharyngeal muscles is a risk factor for aspiration pneumonia. Interstitial lung disease, caused by inflammation in the small airways, is common in PM and DM and is often associated with antisynthetase autoantibodies; it may be present in up to 70% of patients when investigated with sensitive techniques such as high-resolution CT and measurement of pulmonary function and diffusion capacity. In most cases, the changes are present at the time of diagnosis of myositis; they rarely develop after immunosuppressive treatment has started. The severity of interstitial lung disease may vary from mild or even asymptomatic to rapidly progressive (Hamman-Rich-like) with a fatal outcome. In most cases, the interstitial lung disease is mild and has a slowly progressive course. In some cases, improvement in lung function is seen with immunosuppressive treatment. The course and outcome vary, depending on the histopathology, suggesting that different disease mechanisms cause interstitial lung disease.

In general, the clinical course and histopathology of interstitial lung disease in myositis are no different from those in idiopathic interstitial lung disease. The most common histopathologic finding is nonspecific interstitial pneumonia, but other entities such as cryptogenic organizing pneumonia, bronchiolitis obliterans organizing pneumonia, diffuse alveolar damage, and usual interstitial pneumonia are also found. Some studies suggest that bronchiolitis obliterans organizing pneumonia responds favorably to corticosteroids, whereas histopathologic changes compatible with diffuse alveolar damage, usual interstitial pneumonia, or acute interstitial pneumonia respond poorly to corticosteroids or other immunosuppressive therapies and have a poor prognosis.

Arthritis

Joint pain and arthritis are common in patients with PM or DM. The most common form of arthritis is a symmetric arthritis of the small joints of the hands and feet. This arthritis is typically nonerosive but can sometimes be erosive and destructive. Most frequently, arthritis is seen in patients with anti-Jo-1 antibodies and other antisynthetase autoantibodies, but it is also seen in patients with overlapping syndromes of other rheumatic diseases.

Heart

Cardiovascular disease is a risk factor for death among patients with PM and DM. However, clinically evident heart involvement is rare, perhaps indicating that cardiac

Figure 85-8 Erythematous rashes on the hand in dermatomyositis and systemic lupus erythematosus. A, Note the changes on the knuckles and dorsum of the hand in dermatomyositis (Gottron’s sign). B, Rash is absent on the knuckles but present on the phalanges in lupus. C, Capillary nail-fold changes in dermatomyositis. (Courtesy Dr. Paul Plotz.)

Figure 85-9 Mechanic’s hands in a white (A) and a black (B) patient. Note the characteristic skin changes on the lateral side of the fingers. (Courtesy Dr. Paul Plotz.)
involvement may be overlooked in these conditions. Subclinical manifestations are frequently discovered when patients with PM or DM are evaluated. The most frequently reported subclinical manifestations are conduction abnormalities and arrhythmias detected by electrocardiogram (ECG). The underlying pathophysiologic mechanisms that may lead to cardiac manifestations in patients with PM or DM are myocarditis and coronary artery disease, as well as involvement of the small vessels of the myocardium.

Examination with ECG is recommended in newly diagnosed patients with PM or DM. Serum tests such as CK-MB to detect cardiac involvement are unreliable in patients with inflammatory myopathies because CK-MB can be released from regenerating skeletal muscle fibers, a common feature in biopsies from patients with PM or DM. The CK-MB/total CK ratio may be greater than 3%, a threshold value that is used to define myocardial damage. A more specific marker for myocardial damage in myositis patients is increased serum levels of cardiac isofrm troponin I. The other cardiac troponin isofoms, troponin C and troponin T, are less specific and are also expressed in adult skeletal muscle; increased serum levels have been reported in various muscle disorders.

**Gastrointestinal Tract**

Difficulty swallowing is frequent in patients with inflammatory myopathies, particularly those with IBM. Muscle weakness occasionally becomes severe and causes problems with nutrition and aspiration pneumonia. The pathophysiology is related to weakness in the tongue, pharyngeal muscles, and sometimes the lower esophagus. Reflux that requires special care is common, occurring in 15% to 50% of patients. Constipation, diarrhea, and stomach pain are common symptoms and may result from disturbed motility of the gut or GI tract inflammation. Vasculitis in the blood vessels of the GI tract is rare but may be complicated by intestinal bleeding.

**Antisynthetase Syndrome**

A new classification system is based on the presence of MSAs, rather than clinical and histopathologic changes. The most common of these antibodies are the antisynthetase autoantibodies directed against aminoacyl-tRNA synthetases. A clinically distinct subset of myositis, often called antisynthetase syndrome, has been identified in patients with antisynthetase autoantibodies.\(^4^,\(^6^\) The most common of the antisynthetase autoantibodies is anti-Jo-1, which is directed against histidyl-tRNA synthetase. This autoantibody is present in approximately 20% of patients with PM or DM but is only rarely found in patients with IBM.\(^7^\) Antisynthetase syndrome is characterized by the presence of antisynthetase autoantibodies and a set of clinical features that includes myositis, interstitial lung disease, Raynaud’s phenomenon, nonerosive symmetric polyarthritis of the small joints, and mechanic’s hands (see Figure 85-9). These patients often have fever at disease onset and during flares of disease. Antisynthetase syndrome can be seen in patients with PM or DM but is more often seen in patients without skin rashes other than mechanic’s hands.

**Amyopathic Dermatomyositis**

A subset of DM is called clinically amyopathic DM. These patients have a skin rash, which is typical of DM, but no clinical signs of muscle involvement.\(^1^\(^4\) The proposed definition is based on a skin biopsy consistent with DM and a duration of 6 months or longer in the absence of clinical or laboratory evidence of myositis. Some of these patients do have subclinical myositis based on MRI or biopsy findings at presentation; others develop clinically overt myositis sometime later. Patients without clinically overt myositis, however, may develop extramuscular manifestations such as interstitial lung disease, which may be severe. Amyopathic DM may be associated with malignancies, as is the case for classic DM. The frequency of this subset is uncertain, but some recent studies suggest that this form of DM may be more common than previously thought.

**Juvenile Dermatomyositis**

The incidence of juvenile dermatomyositis (JDM) is between 1.7 and 3 per million children. The disease onset has two peaks—at age 6 and 11 years. JDM is more common in girls than in boys in Europe and North America; in Japan and Saudi Arabia, this difference is less prominent. The most common clinical manifestations at disease onset are muscle weakness, easy fatigability, skin rash, malaise, and in some cases fever.\(^1^\(^6\) The skin rash is often pathognomonic and similar to adult DM, with the most typical skin manifestation being heliotrope discoloration of the upper eyelids, Gottron’s papules, periungual erythema, and capillary loop abnormalities. Calcinosis, cutaneous ulceration, and lipodystrophy are more common in juvenile cases than in adults. Calcinosis is seen in 30% to 70% of children with JDM. The calcinosis is most often located at sites exposed to trauma and can be seen in the skin, fascia, or muscles. In some children, the calcinosis becomes prominent and causes contractures and ulcerations. Lipodystrophy occasionally develops, and other metabolic abnormalities such as insulin resistance and hepatomegaly are sometimes seen. Vasculopathy that affects the GI tract with ulceration, perforation, or hemorrhage is rare but seems to be more common in children than in adults with DM. Because this can be a serious sign, screening for GI involvement should be included in the evaluation of patients with JDM. Interstitial lung disease is rarely seen in JDM cases.

The overall prognosis is variable, but some patients have a good prognosis. Patients with JDM may go into remission, allowing the discontinuation of immunosuppressive treatment. Side effects of immunosuppressive treatment such as growth failure are common. In many patients, however, the disease remains chronic, with persisting disease activity into adulthood.

**Inclusion Body Myositis**

IBM is distinguished from PM and DM on the basis of both clinical and histopathologic features.\(^1^\(^4\),\(^1^\(^5\) Sporadic IBM is a distinct entity from familial hereditary inclusion body myopathy, which shares some clinical and histopathologic features but lacks signs of inflammation in muscle tissue.
IBM was identified in the 1960s as a subset of inflammatory myopathies, distinct from PM, primarily on the basis of typical histopathologic features that include sarcoplasmic and nuclear inclusions and rimmed vacuoles. A characteristic clinical phenotype was later identified, characterized by an insidious onset of muscle weakness over months to years, muscle weakness localized predominantly to the thigh muscles and finger flexors, and resistance to glucocorticoid treatment. IBM patients often have a history of frequent falling. Sporadic IBM cases are sometimes misdiagnosed as PM because the classic histopathologic changes (rimmed vacuoles and inclusions) may not be evident in early biopsies (Figure 85-10). A slowly progressive clinical course, development of severe muscle atrophy in the thighs and forearms, and resistance to treatment with immunosuppressive drugs should raise the suspicion of IBM, and a second muscle biopsy should be considered.

In contrast to PM and DM, IBM is more frequent in men than in women, and it is seen mainly in individuals older than 50 years. The onset is more insidious than that of PM or DM. Patients with IBM rarely have pain. The most frequent initial symptoms are difficulty climbing stairs and walking uphill and frequent falls as a result of weakness in the knee extensor muscles. Muscle weakness may become prominent, and even walking across a threshold may become a problem. Difficulty swallowing may also be an early clinical feature, reflecting the involvement of the pharyngeal muscles. The course is slowly progressive, leading to muscle atrophy that can be striking, particularly in the thigh and forearm muscles. Severe weakness may develop, and many patients become wheelchair dependent. Extramuscular organ involvement is rare, although a subgroup of patients with IBM has sicca symptoms and may develop a secondary Sjögren’s syndrome. There are also occasional case reports of IBM in patients with other chronic inflammatory diseases such as SLE, systemic sclerosis, and interstitial pneumonitis. Autoantibodies are rarely present in IBM patients.

IBM is usually resistant to treatment with glucocorticoids and other immunosuppressive agents. Because of this resistance to treatment, some have questioned whether IBM is an autoimmune disease or a degenerative muscle disease supported by the abnormal accumulation of proteins such as amyloid-β in muscle fibers. This issue is still under debate and subject to ongoing research in several institutions around the world.

**Myositis Associated with Malignancies**

An association between DM and malignancies was observed in several early case reports. The clinical implications of this association, irrespective of the pathophysiologic mechanisms involved, are that it is imperative to screen for tumors in patients with DM at the time of diagnosis and at relapse, particularly if the symptoms do not respond to conventional immunosuppressive treatment. A myositis-specific autoantibody that is associated with DM and malignancies in adults, p155/140, has been identified, but its sensitivity and specificity in the context of myositis is not known. However, so far this autoantibody can only be detected by immunoprecipitation, which limits its use in clinical practice. The types of malignancies vary and include not only hematologic malignancies such as lymphoma but also solid tumors such as lung, ovarian, breast, and colon cancer. The screening for malignancies should include, at minimum, a careful clinical examination, routine blood tests, and chest radiograph. For women, mammography and a gynecologic examination should be conducted as well. If any abnormalities are found, these should guide a more thorough investigation for malignancies.

**CLASSIFICATION AND DIAGNOSTIC CRITERIA**

At present, there are no prospectively validated diagnostic or classification criteria for myositis. Dividing diseases into homogeneous subsets serves several important functions including allowing us to estimate disease incidence and prevalence, understand disease pathogenesis and natural history, and evaluate the patient’s response to therapy and prognosis. More than 3 decades ago, Bohan and Peter proposed a set of five criteria to facilitate the diagnosis of IIM patients (Table 85-4). They classified IIMs into five groups: primary idiopathic PM, primary idiopathic DM, IIMs associated with malignancy, childhood IIMs associated with vasculitis, and IIMs associated with collagen vascular diseases. Exclusion criteria include signs of central or peripheral neurologic disease; family history of muscle disease (although familial myositis has been reported in dozens of cases); and symptoms and signs suggestive of muscular dystrophy, granulomatous myositis, infections (including trichinosis, schistosomiasis, trypanosomiasis,
Table 85-4 Bohan and Peter Criteria for Polymyositis and Dermatomyositis

<table>
<thead>
<tr>
<th>Clinical Features</th>
<th>Dermatomyositis</th>
<th>Polymyositis</th>
<th>Inclusion Body Myositis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical features</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Children and adults</td>
<td>Adults*</td>
<td>Adults &gt; 50 yr</td>
</tr>
<tr>
<td>Disease onset</td>
<td>Subacute</td>
<td>Subacute</td>
<td>Chronic</td>
</tr>
<tr>
<td>Muscle weakness</td>
<td>Proximal</td>
<td>Proximal</td>
<td>Selective pattern*</td>
</tr>
<tr>
<td>Symmetry</td>
<td>Symmetric</td>
<td>Symmetric</td>
<td>Asymmetric</td>
</tr>
<tr>
<td>Systemic features</td>
<td>Yes*</td>
<td>Yes*</td>
<td>Yes*</td>
</tr>
<tr>
<td>Skin changes</td>
<td>Yes*</td>
<td>Yes*</td>
<td>No</td>
</tr>
<tr>
<td>Calcinosis</td>
<td>Yes</td>
<td>Rarely</td>
<td>No</td>
</tr>
<tr>
<td>Associated connective tissue disease</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes*</td>
</tr>
<tr>
<td>Associated malignancy†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory features†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum enzymes†</td>
<td>Normal to high</td>
<td>Normal to high</td>
<td>Normal to high</td>
</tr>
<tr>
<td>Abnormal EMG†</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Abnormal muscle biopsy†</td>
<td>Perifascicular atrophy, capillary regeneration, patchy MHC class I expression and microinfarcts</td>
<td>CD8⁺ T cell invasion of non-necrotic fibers and MHC class I expression on fibers</td>
<td>CD8⁺ T cell invasion, MHC expression, vacuolated fibers, and tubulofilamentous inclusions in fibers</td>
</tr>
</tbody>
</table>

*Rarely in children.
*Early involvement of finger flexor, wrist flexor or wrist extensor weakness, and involvement of quadriceps femoris.
*Some patients have dysphagia, synovitis, and interstitial lung disease.
*Some patients have dysphagia.
*Gottron’s sign and heliotrope rash.
*Especially in children.
*Overlap with scleroderma, systemic lupus erythematosus, rheumatoid arthritis, Sjögren’s syndrome, and mixed connective tissue disease.
*Associated with Sjögren’s syndrome but less frequently associated with other connective tissue diseases.
*Dermatomyositis is more frequently associated with cancer than are polymyositis and inclusion body myositis and not overrepresented in polymyositis or inclusion body myositis.
*Serum creatine kinase, aspartate transaminase, lactate dehydrogenase, and aldolase vary from normal to very high levels.
*Myopathic motor unit potentials with spontaneous discharges in dermatomyositis, with and without spontaneous discharges in polymyositis, and mixed pattern of short- and long-duration motor unit potentials in inclusion body myositis.

EMG, electromyogram; MHC, major histocompatibility complex.

staphylococcal infection, and toxoplasmosis), drug-induced myopathy, toxic myopathy, rhabdomyolysis, metabolic disorders, endocrinopathies, myasthenia gravis, or myositis after viral infection (influenza or rubella). A weakness of the Bohan and Peter classification is that it overdiagnoses PM patients and loosely defines overlap syndromes.

Despite several drawbacks, these criteria have served well in diagnosing and defining patients for research purposes for the past 3 decades. IBM was later recognized as a separate disease entity characterized by a slow onset and progression involving finger flexors or the quadriceps muscles. It can occur as a stand-alone entity or with other connective tissue diseases, and patients are often resistant to steroid therapy. Other focal and diffuse forms of myositis such as orbital myositis, focal nodular myositis, macrophagic myositis, and eosinophilic myositis are relatively rare.

Since Bohan and Peter proposed their classification criteria, advances in clinical research have led to the identification of certain autoantibodies that are strikingly associated with some clinical phenotypes of myositis (see Table 85-3). The identification of clinical features associated with MSAs and MAAs has led to the proposal of a serologic approach to complement the Bohan and Peter classification system. Others have suggested that the Bohan and Peter criteria be modified to add MSA as a criterion. However, the inclusion of MSA has some limitations: these antibodies are not present in all patients, the immunoprecipitation techniques that are the “gold standard” for identifying these antibodies are available in only a few commercial laboratories, and the enzyme-linked immunosorbent assays often used can give false-positive or false-negative results.

Ongoing debate and dialogue within the scientific community about the nature of the diagnostic and classification criteria that could better define these disorders is extensive. Some emphasize a focus on histopathologic features and others on autoantibody profiles. Certainly, the addition of autoantibody profiles, characteristic histopathologic and immunohistochemical features, and imaging techniques such as MRI would significantly strengthen the current criteria and better define these disorders. The most frequently used subclassification is based on differences in clinical, immunologic, and histopathologic features and identifies three subtypes of inflammatory myopathies: PM, DM, and IBM (Table 85-5).

Table 85-5 Clinical and Laboratory Features of Idiopathic Inflammatory Myopathy Subgroups

<table>
<thead>
<tr>
<th>Diagnostic Features</th>
<th>Dermatomyositis</th>
<th>Polymyositis</th>
<th>Inclusion Body Myositis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical features</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Children and adults</td>
<td>Adults*</td>
<td>Adults &gt; 50 yr</td>
</tr>
<tr>
<td>Disease onset</td>
<td>Subacute</td>
<td>Subacute</td>
<td>Chronic</td>
</tr>
<tr>
<td>Muscle weakness</td>
<td>Proximal</td>
<td>Proximal</td>
<td>Selective pattern*</td>
</tr>
<tr>
<td>Symmetry</td>
<td>Symmetric</td>
<td>Symmetric</td>
<td>Asymmetric</td>
</tr>
<tr>
<td>Systemic features</td>
<td>Yes*</td>
<td>Yes*</td>
<td>Yes*</td>
</tr>
<tr>
<td>Skin changes</td>
<td>Yes</td>
<td>Yes*</td>
<td>No</td>
</tr>
<tr>
<td>Calciosis</td>
<td>Yes</td>
<td>Rarely</td>
<td>No</td>
</tr>
<tr>
<td>Associated connective tissue disease</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes*</td>
</tr>
<tr>
<td>Associated malignancy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory features</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum enzymes†</td>
<td>Normal to high</td>
<td>Normal to high</td>
<td>Normal to high</td>
</tr>
<tr>
<td>Abnormal EMG†</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Abnormal muscle biopsy</td>
<td>Perifascicular atrophy, capillary regeneration, patchy MHC class I expression and microinfarcts</td>
<td>CD8⁺ T cell invasion of non-necrotic fibers and MHC class I expression on fibers</td>
<td>CD8⁺ T cell invasion, MHC expression, vacuolated fibers, and tubulofilamentous inclusions in fibers</td>
</tr>
</tbody>
</table>

*Rarely in children.
*Early involvement of finger flexor, wrist flexor or wrist extensor weakness, and involvement of quadriceps femoris.
*Some patients have dysphagia, synovitis, and interstitial lung disease.
*Some patients have dysphagia.
*Gottron’s sign and heliotrope rash.
*Especially in children.
*Overlap with scleroderma, systemic lupus erythematosus, rheumatoid arthritis, Sjögren’s syndrome, and mixed connective tissue disease.
*Associated with Sjögren’s syndrome but less frequently associated with other connective tissue diseases.
*Dermatomyositis is more frequently associated with cancer than are polymyositis and inclusion body myositis and not overrepresented in polymyositis or inclusion body myositis.
*Serum creatine kinase, aspartate transaminase, lactate dehydrogenase, and aldolase vary from normal to very high levels.
*Myopathic motor unit potentials with spontaneous discharges in dermatomyositis, with and without spontaneous discharges in polymyositis, and mixed pattern of short- and long-duration motor unit potentials in inclusion body myositis.

EMG, electromyogram; MHC, major histocompatibility complex.
PHYSICAL EXAMINATION

Although most patients present with muscle weakness or fatigue, the IIMs are systemic connective tissue diseases and other organs are frequently involved; therefore a full clinical examination should be conducted when patients present with muscle symptoms. This could also be helpful in distinguishing IIMs from noninflammatory myopathies.

The muscle problems reported by many patients consist of not only muscle weakness but also muscle fatigue or reduced muscle function. Thus the evaluation must differentiate between strength and fatigue by evaluating muscle strength and testing repetitive movements for muscle fatigue. In the early phases, atrophy is usually not a pronounced phenomenon in PM or DM. In later phases, a moderate symmetric atrophy of proximal muscles may be present. Asymmetric atrophies indicate conditions other than inflammatory myopathies.

Patients with IBM often develop more severe atrophy of the quadriceps muscles and the flexor muscles of the forearms; they may also develop deformities in the finger joints and experience difficulty making fists.

Muscle strength can be tested in various ways. A quick screening test for weakness in proximal lower leg muscles is to ask the patient to stand up from a sitting or squatting position without support. A more standardized test that is easy to perform in the clinic is the manual muscle test, with grading according to the Medical Research Council (MRC) scale. There are many variants of this test, but a short form, addressing eight muscles on the dominant side, is recommended as part of the disease activity score by the International Myositis Assessment and Clinical Studies (IMACS). In most patients with PM or DM, muscle strength as assessed by the manual muscle test is good in most of the tested muscle groups. Typically, moderate muscle weakness is seen in the neck flexors and hip girdle muscles. Testing that involves a number of repetitions is often a more sensitive method of detecting muscle impairment. The Functional Index in Myositis-2 is a myositis-specific outcome measure that assesses a number of repetitions. With this test, proximal muscle groups are more involved than are distal muscles. This index is often used by physical therapists.

The skin should also be examined to detect changes including those in nail folds and the scalp. Joints can be affected by arthritis, and heart and lung changes should be carefully looked for.

LABORATORY EVALUATION

Laboratory evaluations are critical components of both diagnosis and patient management. Combinations of laboratory tests are generally used during patient evaluations. Because no laboratory test is highly specific for IIMs, the results of these tests are usually interpreted in the clinical context.

Biochemical

Measuring serum levels of muscle enzymes is an important part of the evaluation of myositis patients. Increased levels of muscle-derived serum enzymes reflect ongoing damage to the muscle parenchyma. These measurements help differentiate IIMs from conditions such as steroid myopathy and denervation, in which atrophy is a prominent feature. Measurement of the serum CK level is traditionally the first step in the assessment of patients with IIM. CK exists as MM (skeletal muscle), MB (cardiac muscle), and BB (brain) isoforms in serum. In comparison with other serum muscle enzymes, CK appears to be a relatively specific and sensitive indicator of the degree of muscle fiber injury. However, the range varies significantly among patients, with levels being near normal in some patients and elevated by several hundred-fold in others.

Generally, 80% to 90% of adult myositis patients show an increase in CK during the initial evaluation. However, a certain proportion of patients, especially those in advanced stages of the disease, show normal or relatively modest elevations in CK, in part because of a lack of muscle mass or the presence of inhibitors of CK activity. Normal CK is relatively more common in DM than in PM. In the absence of CK elevation, it is usually easier to diagnose DM than PM because of the presence of skin rashes in the former. It is also known that CK levels are generally lower in IBM patients than in those with PM or DM. Therefore a normal CK level does not exclude a diagnosis of IIM, particularly IBM or JDM.

Constantly elevated levels of CK are often a sign of inflammatory activity. A rise in CK level is generally correlated with overall disease activity over time, but not with strength or functional measures of disease activity. CK measurements are usually not useful for monitoring disease exacerbations, and they should always be evaluated in the clinical context. CK levels may normalize without clinical improvement, or they may increase without clinical worsening; however, increasing levels point to a potential flare and warrant closer clinical evaluation. CK elevations are not specific for myositis because this enzyme is also elevated in other muscle diseases including muscular dystrophies, rhabdomyolysis, hypothyroidism, and many drug-induced myopathies. It is important to note that serum levels of CK-MB can be elevated in patients with myositis as a result of the regeneration of skeletal muscle fibers; they are not specific for heart involvement in these patients. The cardiac isoform troponin I has the highest specificity as an indicator of myocardial involvement and is the most reliable serum marker for detecting myocardial damage in patients with inflammatory muscle disease.

Measurement of other serum muscle enzymes including aldolase, aspartate transaminase (AST), alanine transaminase (ALT), and lactate dehydrogenase (LDH) significantly improves the chance of diagnosing myositis, especially in patients with active disease and normal CK levels. Aldolase, LDH, and AST are better correlated with disease activity in JDM patients. The main disadvantage of these enzymes is that they are also elevated in liver diseases; therefore the muscle source needs to be identified before interpreting the data.

The serum myoglobin level is a sensitive index of muscle fiber membrane integrity and can therefore be used to assess the degree of disease activity. The advantage of the myoglobin assay is that it involves a nonenzymatic immunologic reaction. Disadvantages are that a significant range of serum myoglobin levels occurs in myositis patients because of
Inclusion Body Myositis

Dermatomyositis

A particular subgroup. The general features include necrosis, that are common to all IIMs and specific features unique to it is important to select a muscle that is moderately weak. definitive diagnosis of IIMs. Muscle biopsy is the “gold standard” for the diagnosis of inflammatory myopathies and a critical component of the treatment further supports the diagnosis of IBM in an other specific histopathologic finding, albeit a late sign, is perifascicular atrophy. The infiltrates typically have a perivascular distribution. The inflammatory infiltrates are dominated by a high percentage of CD4 T cells and macrophages at the sites of inflammation, with B cells occasionally in evidence. Although perifascicular atrophy is the hallmark of the histologic changes seen in DM, it may not be visible if a biopsy is acquired early in the course of the disease. In early DM, the MHC class I expression is patchy and the perifascicular areas are usually stained.

The features of PM include the presence of macrophages and activated CD8 T cells in muscle fibers and the expression of MHC class I molecules on muscle fibers. Mononuclear cell invasion around non-necrotic muscle fibers in endomyial areas is a characteristic feature of IIMs. The histologic features of IBM resemble those of PM but also include unique features such as red-rimmed vacuoles and inclusions (cytoplasmic or nuclear) and amyloid deposits. An increased number of cytochrome-c oxidase–negative fibers can also be seen, but this change is not specific for IBM. In IBM, electron microscopy usually demonstrates 15- to 21-nm cytoplasmic and intranuclear tubulofilaments, which are not found in DM or PM. It is not uncommon to find biopsies that are negative for both rimmed vacuoles and tubulofilamentous inclusions. In this situation, if the suspicion for IBM is high, it is best to obtain another sample or to treat the patient with steroids. Nonresponsiveness to treatment further supports the diagnosis of IBM in an otherwise typical patient. Inflammation surrounding necrotic fibers is a feature of some muscular dystrophies (e.g., facioscapulohumeral muscular dystrophy, limb-girdle muscular dystrophy type IIb, Duchenne’s muscular dystrophy), where it is secondary to muscle cell degeneration. Thus the presence of a mononuclear infiltration surrounding

**Immunologic**

The immune response to self-antigens is a common feature of several systemic autoimmune rheumatic diseases including IIMs. ANAs are found in approximately 60% to 70% of myositis patients. The autoantibody response in IIMs is directed to ubiquitous nuclear and cytoplasmic antigens. The presence of these antibodies is usually assessed by an indirect immunofluorescence assay. ANAs are more frequently found in patients with PM and DM, especially those with overlap syndrome. These are less frequently seen in IBM patients or those with malignancy-associated myositis. High-titer ANA is a particularly valuable finding for differentiating IIMs from dystrophies. Many IIM patients show speckled nuclear ANA patterns, and about 10% of IIM patients also show exclusive cytoplasmic patterns in indirect immunofluorescence staining. Many of the MSAs are associated with distinctive clinical features such as skin rash or interstitial lung disease (Table 85-6). Certain MSAs such as anti–Jo-1 are more frequently associated with PM; others such as Mi-2 are more frequent in DM.

**Histologic**

Muscle biopsy is the “gold standard” for the diagnosis of inflammatory myopathies and a critical component of the definitive diagnosis of IIMs. For optimal biopsy results, it is important to select a muscle that is moderately weak. The histologic features can be grouped into general features that are common to all IIMs and specific features unique to a particular subgroup. The general features include necrosis, regeneration, degeneration, variation in fiber diameter, increase in connective tissue, and inflammation. The features specific to DM include loss of capillaries, alterations in the morphology of capillaries, capillary necrosis with the deposition of complement products (e.g., membrane attack complex) on the vessel walls, and, rarely, muscle infarcts. Another specific histopathologic finding, albeit a late sign, is perifascicular atrophy. The infiltrates typically have a perivascular distribution. The inflammatory infiltrates are dominated by a high percentage of CD4 T cells and macrophages at the sites of inflammation, with B cells occasionally in evidence. Although perifascicular atrophy is the hallmark of the histologic changes seen in DM, it may not be visible if a biopsy is acquired early in the course of the disease. In early DM, the MHC class I expression is patchy and the perifascicular areas are usually stained.

The features of PM include the presence of macrophages and activated CD8 T cells in muscle fibers and the expression of MHC class I molecules on muscle fibers. Mononuclear cell invasion around non-necrotic muscle fibers in endomyial areas is a characteristic feature of IIMs. The histologic features of IBM resemble those of PM but also include unique features such as red-rimmed vacuoles and inclusions (cytoplasmic or nuclear) and amyloid deposits. An increased number of cytochrome-c oxidase–negative fibers can also be seen, but this change is not specific for IBM. In IBM, electron microscopy usually demonstrates 15- to 21-nm cytoplasmic and intranuclear tubulofilaments, which are not found in DM or PM. It is not uncommon to find biopsies that are negative for both rimmed vacuoles and tubulofilamentous inclusions. In this situation, if the suspicion for IBM is high, it is best to obtain another sample or to treat the patient with steroids. Nonresponsiveness to treatment further supports the diagnosis of IBM in an otherwise typical patient. Inflammation surrounding necrotic fibers is a feature of some muscular dystrophies (e.g., facioscapulohumeral muscular dystrophy, limb-girdle muscular dystrophy type IIb, Duchenne’s muscular dystrophy), where it is secondary to muscle cell degeneration. Thus the presence of a mononuclear infiltration surrounding

**Table 85-6 Immunologic Features of Idiopathic Inflammatory Myopathies**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Dermatomyositis</th>
<th>Polymyositis</th>
<th>Inclusion Body Myositis</th>
</tr>
</thead>
<tbody>
<tr>
<td>B cell infiltration</td>
<td>+</td>
<td>−/+</td>
<td>−/+</td>
</tr>
<tr>
<td>T cell infiltration</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CD8+ T cell infiltration in non-necrotic fibers</td>
<td>−/+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Vascular membrane attack complex</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Immunoglobulin deposition on blood vessels</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Major histocompatibility complex class I</td>
<td>−/+−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>expression on muscle fibers</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Cytokines and chemokines</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cell adhesion molecules</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Antinuclear antibodies</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Anti-Jo 1 antibodiesa,b</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Anti–signal recognition particle antibodiesa</td>
<td>−/+−</td>
<td>+</td>
<td>−/+</td>
</tr>
<tr>
<td>Anti–Mi-2 antibodiesa</td>
<td>+</td>
<td>−/+</td>
<td>−/+</td>
</tr>
<tr>
<td>Anti–Pm-Scl antibodiesa</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Anti–Pm-Scl antibodiesb</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Anti–Pm-Scl antibodiesc</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Anti–Pm-Scl antibodiesd</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Anti–Pm-Scl antibodiesf</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

*Mostly in perifascicular areas and necrotic fibers.

bLess frequently, but 20% higher than in normal population.

bFrequency varies among ethnicities; more frequent in polymyositis (22%) than dermatomyositis (16%) or inclusion body myositis (5%).

bPresent only in a proportion of polymyositis (14%), dermatomyositis (5%), and inclusion body myositis (3%) patients.

bPresent only in a proportion of polymyositis (9%), dermatomyositis (21%), and inclusion body myositis (8%) patients.

bPresent only in a proportion of polymyositis (7%) and dermatomyositis (6%) patients.
**Table 85-7  Histologic Features of Idiopathic Inflammatory Myopathies**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Dermatomyositis</th>
<th>Polymyositis</th>
<th>Inclusion Body Myositis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Necrosis of muscle fibers</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Variation in fiber diameter</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Regeneration of muscle fibers</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Proliferation of connective tissue</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Infiltration of mononuclear cells*</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Perivascular and perimysial inflammation</td>
<td>+/−/+</td>
<td>+/−/+</td>
<td>+/−/+</td>
</tr>
<tr>
<td>Endomysial inflammation</td>
<td>+/−/+</td>
<td>+/−/+</td>
<td>+/−/+</td>
</tr>
<tr>
<td>Perifascicular atrophy</td>
<td>+</td>
<td>−/+</td>
<td>−/−</td>
</tr>
<tr>
<td>Abnormally dilated capillaries</td>
<td>+</td>
<td>−/+</td>
<td>−</td>
</tr>
<tr>
<td>Reduced capillary density</td>
<td>+</td>
<td>−/+</td>
<td>−</td>
</tr>
<tr>
<td>Deposition of complement on vessel walls</td>
<td>+</td>
<td>−/+</td>
<td>−</td>
</tr>
<tr>
<td>Microinfarcts</td>
<td>+</td>
<td>−/+</td>
<td>−</td>
</tr>
<tr>
<td>Invasion of non-necrotic fibers by cytotoxic T lymphocytes and macrophages</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rimmed vacuoles with amyloid deposits and tubulofilaments†</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Angulated or atrophic and hypertrophic fibers</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Ragged red or cytochrome oxidase−negative fibers</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

*Inflammation is absent in a small proportion of polymyositis and dermatomyositis biopsies.
† Also seen in chronic neurogenic conditions and distal myopathies.

non-necrotic muscle fibers confirms the diagnosis of IBM or PM. The common and unique immunologic and histologic features of the various subgroups are listed in Tables 85-6 and 85-7, respectively.

**Molecular**

One of the more tangible deliverables of the human genome product has been the development and use of microarrays for messenger RNA (mRNA) profiling. In the commonly used form of microarray, about 1 million DNA probes are placed on 1 cm² glass slides, allowing the query of each gene of the genome. Although all genes are shared among all cells, only certain genes are expressed (turned on) in any specific cell at any specific time. Messenger RNA expression profiling using microarrays allows genome-wide assessment of the response of each gene, with a comparison of normal and pathologic states.

Muscle is routinely biopsied as part of the clinical workup of muscle disease, and muscle histopathology is an important part of the diagnosis of inflammatory myopathies. Diagnostic muscle biopsies have been used for mRNA expression profiling in a series of studies, with comparisons between inflammatory myopathies (JDM, PM, IBM) and dystrophic myopathies (Duchenne’s muscular dystrophy). These comparisons are important because they differentiate between inflammation associated with downstream myofiber degeneration or regeneration (dystrophies) and inflammatory processes that may initiate the inflammatory myopathies.

Microarrays have provided considerable new insights into DM. Early on, mRNA profiling in DM showed a predominance of type 1 IFN-responsive pathways, suggesting the possible persistence of an antiviral response. Particularly prevalent was the dramatic expression of the IFN-inducible MxA gene. This signature was confirmed and extended, supporting an important role of the innate immune response, with prominent plasmacytoid dendritic cell infiltration in DM compared with the other inflammatory myopathies. The beneficial effect of intravenous immunoglobulin (IVIG) in DM has been queried by microarrays to define the drug-responsive genes, then compared with the lack of response in IBM. This study suggested that IVIG suppressed a relatively small subset of inflammatory responses in DM muscle including complement (C1q) and inflammatory cell migration proteins (ICAM-1). Microarrays have helped elucidate an unexpectedly important role for innate immunity in DM. They have also provided new insights into the humoral immunity in PM and IBM. Specifically, mRNA expression profiling showed a high proportion of immunoglobulin transcripts as differentially expressed (59% of all detected genes in IBM and 33% in PM). Plasma cells, defined as those terminally differentiated B cells expressing CD138 but not CD19 or CD20, were then shown to be a key differentiating cell type within IBM and PM muscle.

**IMAGING**

**Muscles**

Ultrasonography, CT, and MRI are the three general imaging techniques used to evaluate skeletal muscle. MRI has emerged as the method of choice for the examination of soft tissue muscle abnormalities because it efficiently visualizes and quantifies inflammation, fat infiltration, calcification, and alterations in muscle size and localizes pathologic changes in specific muscle groups (Figure 85-11). MRI examinations can be done on large volumes and can be helpful guides for muscle biopsy sampling. MRI is a potential outcome tool to be used in the longitudinal analysis of responses to therapy and in clinical trials, although its sensitivity to changes has not been validated.

Ultrasonography is useful for detecting abnormal vascularity, and rates of blood flow can be monitored effectively with color Doppler imaging. The main disadvantage of ultrasonography is its inability to visualize deep-seated muscles in cross-section. Moreover, image analyses are more subjective than with MRI and depend to a greater degree
CHAPTER 85 | INFLAMMATORY DISEASES OF MUSCLE AND OTHER MYOPATHIES

Inflammatory Diseases of Muscle and Other Myopathies

and are a useful measure when serum levels and muscle strength are uninterpretable. The inflammation in IIMs is often patchy, and EMG is useful in determining which muscle should be sampled for biopsy. Because EMG can cause histopathologic changes that complicate the interpretation of the biopsy, it is best to perform EMG on one side and obtain the muscle biopsy from the same muscle on the contralateral side. Even for IBM, in which the disease is often asymmetric, this technique is useful as long as the contralateral muscle is also weak.

LUNG FUNCTION TESTS

Pulmonary function tests are an important means of obtaining an objective assessment of respiratory involvement. Typically, patients demonstrate a restrictive ventilatory impairment, with decreased total lung capacity, functional residual capacity, residual volume, forced expiratory volume in 1 second (FEV₁), and forced vital capacity (FVC), but with a normal or elevated FEV₁/FVC ratio and reduced diffusing capacity for carbon monoxide. Pulmonary function tests are also important in estimating disease severity and response to therapy, in concert with radiographic examination.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of IIMs and other myopathies is important because the clinical associations and response to therapeutic interventions differ significantly. A variety of myopathies closely mimic IIMs (Table 85-8).

**Dystrophic Myopathies**

**Dysferlinopathy**

Genetic defects in the dysferlin gene result in limb-girdle muscular dystrophy type 2B and distal muscular dystrophy of the Miyoshi type. These diseases can appear in the late teens or early 20s, and the weakness in the limb-girdle type 2B phenotype first assumes a pelvifemoral distribution: the quadriceps muscle is affected first, followed by weakness in the arms in the later stages of the disease. A relatively acute

---

**Figure 85-11** Magnetic resonance images (short tau inversion recovery) of the thigh. Note the symmetric inflammation in the affected muscle, seen as bright areas relative to unaffected muscle.
onset with elevated levels of serum muscle enzymes points to PM as a differential diagnosis. The weakness in the Miyoshi phenotype occurs predominantly in the gastrocnemius and soleus muscles, thereby affecting the ability to walk on the toes. The weakness is slowly progressive, with loss of ambulation generally occurring in the fourth decade, but earlier in some cases. Serum CK levels are high during the active phase of the disease. In general, the muscle biopsy is dystrophic, with significant mononuclear cell infiltration and small sarcolemmal defects with thickened basal lamina structures over the defects.

### Facioscapulohumeral Muscular Dystrophy

A partial deletion of the D4Z4 repeats near the chromosome 4q telomere at 4q35 leads to the facioscapulohumeral muscular dystrophy.
phenotype. The initial weakness usually affects the facial muscles, and the onset is insidious. Shoulder weakness is commonly seen because of the weakness of the scapular fixator muscles. The weakness is generally slowly progressive, with typical myopathic changes on the EMG such as brief, small-amplitude, polyphonic voluntary motor unit potentials. The presence of perivascular, endomysial, and perimysial inflammation is a common feature. Serum CK levels are elevated and vary with age and sex.

**Dystrophinopathies**

These X-linked recessive disorders are caused by mutations in the dystrophin gene. Milder forms of Becker's muscular dystrophy manifest as myalgias, muscle cramps, exercise intolerance, mild limb-girdle weakness, and quadriceps myopathy. The severe Becker's phenotype that presents before age 8 years is indistinguishable from the Duchenne's phenotype. An elevation in serum CK levels is seen in asymptomatic patients. The mean age at loss of ambulation is usually in the fourth decade. Histologic features include variation in fiber size, central nuclei, regeneration, necrosis, hypercontracted fibers, and endomysial fibrosis. In Becker's dystrophy the number of necrotic and regenerating fibers is decreased compared with the Duchenne phenotype, and the incidence of hypercontracted and central nuclei increases with age. A plasma membrane defect in non-necrotic fibers and endomysial inflammation with macrophage, T cell, mast cell, and eosinophil infiltration are also characteristic features of this disease.

**Proximal Myotonic Myopathy**

CCTG expansion in intron 1 of the zinc finger transcription factor (ZNF9) gene results in type II myotonic dystrophy. The myotonia is usually absent or minimal but is detectable by EMG. The weakness is mainly proximal, with minimal or no facial involvement. Smooth muscle, cardiac, and diaphragmatic involvement is common in this disease. First-degree heart block is the most common abnormality, and sudden death is well documented. Muscle biopsies show nonspecific features such as central nuclei, sarcoplasmic masses, and atrophy of type I fibers.

**Sarcoglycanopathy**

Mutations in sarcoglycans (α, β, γ, and δ) result in limb-girdle muscular dystrophy types IIC to IIF. Sarcoglycancopathies often start in childhood, with a median age of onset of 6 to 8 years. These diseases present initially as pelvic muscle weakness including a waddling gait and difficulty performing common tasks such as getting up from the floor, climbing the stairs, and running. The trunk muscles are prominently affected, and upper extremity involvement usually follows lower extremity involvement. Distal muscles are generally spared until later in the disease process. These progressive disorders result in high levels of serum CK early in the disease; the levels decrease when patients become wheelchair bound by 12 to 16 years of age. Dilated cardiomyopathy is often seen in these disorders, and muscle biopsies show marked regeneration and necrosis.

**Neuromuscular Disorders**

**Motoneuron Diseases**

These diseases including amyotrophic lateral sclerosis (ALS) are progressive, degenerative disorders of motoneurons in the spinal cord, brain stem, and cerebral motor cortex that manifest clinically as amyotrophy and exaggerated reflexes. These diseases are characterized by a selective loss of function of upper or lower motoneurons, finally leading to a progressive loss of both types of motoneurons over time. EMG shows fibrillation and fasciculation potentials in the muscles of the lower and upper limbs or in the bulbar muscles. Muscle biopsies show the presence of denervation atrophy and secondary myopathic changes in chronically denervated muscles. IBM is the primary muscle disorder most likely to be confused with ALS, and muscle biopsy helps differentiate the two. Serum CK levels are slightly elevated, particularly in the early stages of the disease and in men who are physically active.

**Spinal Muscular Atrophy**

Late-onset forms of spinal muscular atrophy (SMA) are characterized by progressive muscle weakness and atrophy and reduced tendon reflexes. EMG and muscle testing reveal neurogenic changes in the muscle. Typical muscle biopsy findings include small and large groups of atrophic fibers in the chronic and severe forms of SMA, respectively. Histochemical changes show fiber type grouping, indicating reinnervation. Serum CK levels are slightly increased in juvenile-onset cases and normal in other forms of SMA. EMG shows abnormal spontaneous electrical activity (fibrillations, positive sharp waves, fasciculations), suggesting ongoing denervation.

**Myasthenia Gravis**

The clinical manifestations of myasthenia gravis include abnormal weakness that is worsened by repeated or sustained exertion and fatigability. Proximal muscles are usually more severely affected than distal muscles. This is a generalized disease that exhibits external ocular muscle involvement, positive anticholinesterase drug tests, and a decremental EMG response. Patients are often positive for acetylcholine receptor antibodies.

**Metabolic Myopathies**

**Acid Maltase Deficiency**

This autosomal recessive glycogen storage disease is caused by acid maltase gene mutations. The disease has infantile, childhood, and adult variants. The infantile form manifests in the first few months after birth as rapidly progressive weakness and hypotonia, with death occurring as a result of cardiorespiratory failure. The childhood form manifests as a myopathy in which the weakness is usually greater in the proximal than in the distal muscles; the disease progresses relatively slowly, and patients die of respiratory failure. The adult form presents in the 20s as a progressive myopathy that resembles PM or limb-girdle muscular dystrophy, with
additional respiratory symptoms. Serum muscle enzymes (CK, AST, and LDH) are increased in all three forms of the disease, and EMG indicates myopathy in all three cases. Histologic examination reveals a vacuolar myopathy, with the vacuoles displaying a high glycogen content and strongly positive staining for acid phosphatase; necrotic and regenerating fibers are uncommon.

**McArdle’s Disease**

McArdle’s disease is the most common of the nonlysosomal muscle glycogenoses. Exercise intolerance is the characteristic feature of this disease, and it often manifests as early fatigue, myalgia, and stiffness of exercising muscle that is relieved by resting. The EMG is normal in some patients; it shows nonspecific myopathic changes in others. Forearm ischemic exercise testing shows virtually no increase in venous lactate in most patients. Serum CK levels, however, are variably elevated in these patients. Muscle biopsies show subsarcolemmal deposits of glycogen at the periphery of the fibers.

**Mitochondrial Myopathies**

Mitochondrial diseases are heterogeneous and often present a diagnostic challenge. It has been suggested that myopathy is due to mutations in mitochondrial DNA in the skeletal muscle. The clinical course of the pure myopathy varies from rapidly progressive to almost reversible disease, with disease onset occurring from infancy through adulthood. The weakness is facioscapulohumeral and more proximal than distal, with involvement of the orbicular and extraocular muscles. Patients often complain of exercise intolerance and fatigue and have recurrent episodes of myoglobinuria. Muscle biopsy plays a critical role in the diagnosis of these conditions, especially the use of special histochemical stains that detect succinate dehydrogenase, Cox staining, and Gomori trichrome staining.

**Endocrine Myopathies**

**Cushing’s Syndrome**

Cushing’s syndrome is an endogenous glucocorticoid excess disease that manifests as muscle weakness and wasting. Chronic corticosteroid treatment results in similar manifestations and significant loss of strength within a few weeks of treatment. Muscle biopsy shows increased vacuolations and glycogen accumulation in type II muscle. The onset of weakness is usually insidious. The weakness is primarily proximal, with more severe involvement in the legs than in the arms. These patients generally show normal serum muscle enzyme levels (CK, AST, and LDH). Muscle wasting can often be reversed if the glucocorticoid levels are returned to the normal range.

**Hyperthyroid and Hypothyroid Myopathy**

Myopathic thyroid disease is characterized primarily by proximal muscle weakness and muscle wasting. When distal weakness occurs, it often follows proximal myopathy. Exercise intolerance, fatigue, and breathlessness are common complaints, and weakness of the respiratory muscles results in respiratory insufficiency and the need for ventilatory support. Patients often have difficulty rising from a sitting position or lifting their arms above their heads. Serum muscle enzymes (CK, AST, and ALT) are often normal or low in hyperthyroidism and elevated in hypothyroidism. EMG findings are variable, with short-duration motor unit potentials and increased polyphasic potentials in proximal muscles; fibrillations and fasciculations are uncommon. Muscle biopsy shows atrophy in fiber types, nerve terminal damage, fatty infiltration, and isolated fiber necrosis, with macrophage and lymphocyte infiltration.

**Infectious Myopathies**

**Human Immunodeficiency Virus Myopathy**

Neuromuscular manifestations are common in human immunodeficiency virus (HIV)–induced myopathy. The clinical features typically include a myopathy of subacute onset that progresses slowly. The myopathy often starts as proximal symmetric muscle weakness with or without muscle wasting, similar to that in IIMs. Histologic features include muscle fiber necrosis, inflammation, and vacuolated muscle fibers, with a significant increase in serum CK levels. EMG shows spontaneous activity, with fibrillation potentials, positive sharp waves, and brief, low-amplitude polyphasic motor unit potentials.

**HTLV-1 Myopathy**

Myositis associated with human T-lymphotropic virus 1 (HTLV-1) has been noted in certain areas of the world (e.g., Japan, Jamaica). In these patients, symptoms of PM and IBM occur either alone or in combination with tropical spastic paraparesis. Typical features include weakness and increases in serum CK levels. Histologic findings include interstitial inflammation, muscle fiber necrosis in PM and endomyosial inflammation, vacuoles, amyloid deposits, and tubulofilaments in IBM.

**Parasitic Myopathies**

Diseases caused by various parasites—protozoa (e.g., toxoplasmosis, trypanosomiasis, sarcocystosis, malaria), cestodes (e.g., cysticercosis, echinococcosis, coenurosis, sparganosis), and nematodes (trichinellosis, taxocariasis, racunculiasis)—can cause myositis. The clinical features include nonspecific complaints such as myalgia and focal swelling, as well as typical features of PM and DM. Each parasitic infection shows typical changes on muscle biopsy (e.g., the presence of tachyzoites and toxoplasma cysts along with perimysial and endomyosial inflammation). A combination of muscle biopsy and serologic findings is useful in making a diagnosis.

**Drug-Induced Myopathies**

Drugs can induce myopathic changes either by acting directly on the muscle or by indirectly influencing various factors required for muscle cell survival and growth.
Zidovudine Myopathy

Nucleoside analogues such as zidovudine are used to treat HIV because they act as false substrates for the viral reverse transcriptase. These drugs also cause myalgias, proximal muscle weakness, and fatigue and are sometimes associated with increased levels of serum CK. EMG shows typical myopathic changes. Histologically, muscle fibers show ragged red fibers; atrophic fibers show marked sarcoplasmic changes, with rod-body formation. Pronounced abnormalities in mitochondria, myofilaments, and tubules are also noted by electron microscopy. It has been suggested that these drugs also inhibit mitochondrial DNA polymerase, thus producing the mitochondrial abnormalities. Discontinuation of therapy improves muscle strength and function. In these patients, zidovudine-induced mitochondrial myopathy may coexist along with HIV-induced T cell–mediated inflammatory myopathy.\(^\text{188}\)

Statin Myopathy

Statins are lipid-lowering drugs (e.g., lovastatin, simvastatin) that are known to cause necrotizing myopathy. These HMG-CoA reductase inhibitors generally suppress specific cholesterol synthesis and lower plasma concentrations of low-density lipoprotein. The clinical features of this condition include myalgia, cramps, and acute and subacute painful proximal myopathy, with histologic features varying from mild, discrete, and unspecific to muscle fiber necrosis, mononuclear cell infiltration, and myophagocytosis and regeneration. A mild increase in serum CK levels is also noted. Other agents that are known to cause necrotizing myopathy include fibrin acid derivatives (clofibrate, gemfibrozil); nicotinic acid; organophosphate poisoning; and \(\varepsilon\)-aminocaproic acid.

Other Drugs

d-Penicillamine is known to induce clinical features reminiscent of DM; recovery usually occurs after withdrawal of the drug. Agents such as IFN-\(\alpha\) that are used to treat viral hepatitis and certain malignant tumors are also known to induce clinical features that resemble PM. Amphiphilic drugs such as chloroquine, hydroxychloroquine, and amiodyarone are also known to induce cytoplasmic vacuoles, necrosis, and longitudinal branching of muscle fibers.\(^\text{189}\)

Drugs that affect microtubules such as colchicines and vincristine also induce myopathic changes, with the appearance of characteristic autophagic vacuoles in muscle fibers.

Management and Outcome

The recommended treatment for patients with PM or DM is based on a combination of pharmacologic therapy and physical exercise. The optimal pharmacologic treatment in PM and DM is unclear. Few controlled trials have been undertaken, so recommendations are based on clinical observations from case series. With these limitations in mind, a suggested outline of treatment for patients with PM or DM is depicted in Figure 85-12. In addition to providing treatment, it is important to give patients adequate information about their disease and its treatment. This educational component is best provided by a rheumatology team and patient support groups.

Pharmacologic Treatment

The initial pharmacologic treatment in PM and DM is high-dose glucocorticoids: 0.75 to 1 (up to 2) mg/kg body weight per day for 4 to 12 weeks. Most experts recommend that glucocorticoid treatment be combined with another immunosuppressive drug to reduce the side effects of the glucocorticoids and to boost the immunosuppressive effect. The most frequently used immunosuppressive agents are azathioprine and methotrexate. In the extension phase of one of the few double-blind, placebo-controlled trials that have been reported, the combination of azathioprine and glucocorticoids, as compared with prednisone alone, was associated with better functional ability and a lower requirement for prednisone after 1 and 3 years.\(^\text{190,191}\) The recommended azathioprine dosage is 2 mg/kg per day. The dosing regimen for methotrexate is similar to that for rheumatoid arthritis—up to 25 mg weekly, although there have been reports of higher doses. Pulmonary involvement related to myositis does not seem to be a contraindication for methotrexate.

The combination of methotrexate and azathioprine proved to be successful in a few patients with refractory myositis in a prospective, randomized, open-label crossover study comparing two aggressive approaches. There are also newer reports that mycophenolate mofetil might be effective. In patients with interstitial lung disease, cyclophosphamide could be of value. There are also a few reports that cyclosporine A or tacrolimus can be beneficial in these cases.\(^\text{192,193}\) In treatment-resistant DM, a high dose of IVIG was found to have a beneficial effect on muscle strength when compared with placebo; however, the therapeutic effect was temporary, and repeated infusions were required.\(^\text{194}\) In patients with severe, rapidly progressive disease that might be life threatening, high-dose pulses of intravenous methylprednisolone have been reported to be beneficial. Pharmacologic treatment including tapering of the corticosteroid dose should be guided by clinical outcome measures. As discussed earlier, the most appropriate outcome measures are muscle endurance and muscle strength. Side effects of glucocorticoids in these high doses are frequent. Prophylaxis against osteoporosis is recommended with vitamin D and calcium and, when clinically indicated, bisphosphonates. Steroid myopathy is another possible consequence of glucocorticoid treatment that is particularly problematic in patients with inflammatory myopathies. There is no specific test to verify steroid myopathy, but in the absence of active clinical disease, steroid myopathy could contribute to muscle weakness. If steroid myopathy is suspected, tapering of the glucocorticoid dose with careful evaluation of the clinical response is recommended. Glucocorticoids may also cause hypokalemia, and if this is not corrected, it may be associated with muscle weakness and incorrectly interpreted as myositis activity.

The depletion of B cells has recently emerged as a new strategy in autoimmune diseases. One approach is to use rituximab (monoclonal antibody against CD20). There are
a few case series suggesting a beneficial effect of depleting B cells in patients with DM or PM. A large international multicenter trial of B cell depletion is currently under way. Use of anti-TNF treatments produced mixed results. For example, treatment with infliximab at 3 mg/kg resulted in improved muscle function in patients with juvenile DM. However, interpretation of the data from this trial was complicated by the fact that all patients were concomitantly being treated with a tapering dose of prednisone during the trial. A second, larger trial examining the use of infliximab in patients with PM, DM, or IBM found that only 30% of patients experienced an improvement in disease symptoms, while the remaining participants suffered from adverse side effects or worsening of the disease. IBM is usually nonresponsive to glucocorticoids. There are occasional case reports of “stabilization” for a period of months, but this condition probably reflects the natural history of the disease. Prolonged administration of glucocorticoids to IBM patients may actually lead to worsening of clinical aspects of the disease, despite the improvement in CK levels and reduced T cell numbers in biopsies. Prednisone treatment also increases the number of amyloid-containing fibers. A few small studies have shown some beneficial effect of methotrexate, anti–T lymphocyte globulin, or mycophenolate. Anabolic steroids and oxandrolone may have some beneficial effects on muscle strength, although this benefit needs to be confirmed in larger studies. Most experts consider their use justified in combination with glucocorticoids for a limited period in patients who have inflammatory infiltrates on muscle biopsy, or in combination with a more aggressive immunosuppressive treatment (e.g., methotrexate, azathioprine) in patients with another connective tissue disease.

Nonpharmacologic Treatment

With immunosuppressive treatment, approximately 75% of patients improve, but few recover normal muscle function, even in the absence of muscle inflammation. In previous
decades, patients with myositis were cautioned against exercise due to concerns over possible muscle damage and inflammation, but recent publications have demonstrated that combining exercise and immunosuppressive therapy is a safe approach and has clear beneficial effects on muscle function.\textsuperscript{150} In a pair of small studies of immunosuppressed myositis patients, it was demonstrated that personalized sub-maximal exercise programs improved muscle function without causing any increase in inflammation. This was true for patients who had individually tailored exercise regimens\textsuperscript{201} and self-administered home exercise routines.\textsuperscript{202} Examinations of patient biopsies both before and after exercise indicated that the patients showed improved muscle strength due to an increase in the proportion of slow twitch (type I) muscle fibers and reduction of inflammation and fibrosis.\textsuperscript{202,203} Combining exercise and immunosuppressive therapy is a safe approach and has clear beneficial effects on muscle function.\textsuperscript{\textsuperscript{9}} The exercise regimen should be individualized and supervised by a physiotherapist to avoid the overuse of muscles. Physical exercise is now recommended as combination therapy with immunosuppressive treatment.

**Assessing Disease Activity and Outcome**

The most important variable to measure in myositis patients is muscle performance or physical function. However, it is equally important to evaluate whether impaired muscle function reflects disease activity or irreversible muscle damage.

**Muscle Examination**

**Manual Muscle Test.** There are several tools to measure muscle performance, but the most often used method in clinical practice and clinical trials is the manual muscle test with the MRC scale (see earlier). The drawback is that these tools measure muscle strength but not muscle endurance, which is often a major problem in PM or DM. Further, they have not been validated in adults with inflammatory myopathies. Previously, the number of muscle groups tested using the manual muscle test varied and different scales were used (5 or 10 grade). Recently, a consensus was reached to assess eight muscle groups on the dominant side using a 0- to 10-point scale, where 0 is no muscle contraction, 5 is ability to hold the test position without any added pressure, and 10 is ability to hold the test position against strong pressure. The points between these scores are based on gradual increased resistance against the examiner’s pressure. The eight muscle groups tested are neck flexors, shoulder abduction (deltoid middle), biceps brachii, wrist extensors, knee extensors (quadriceps), dorsiflexion of ankle, gluteus maximus, and gluteus medius. The score achieved varies between 0 and 80.

**Functional Index in Myositis.** The Functional Index in Myositis and its revised form, the Functional Index in Myositis-2, were developed as outcome measures for patients with PM or DM. This test measures the number of repetitions that can be performed in defined muscle groups.\textsuperscript{157,162} It is a more sensitive method of measuring impaired muscle function in patients with PM or DM.\textsuperscript{157} The drawback is that it takes a longer time to perform than the manual muscle test, and it may be difficult to use in everyday clinical practice. Preferably, the Functional Index in Myositis-2 is administered by a physiotherapist and can be combined with the manual muscle test.

**Extramuscular Involvement**

In some patients, extramuscular symptoms predominate among the clinical features. These symptoms may require other assessment tools such as those used to evaluate interstitial lung disease. For monitoring the effects of treatment of interstitial lung disease, high-resolution CT and pulmonary function tests are recommended.

**Disease Activity and Damage**

It is also important to distinguish whether symptoms are caused by active inflammatory disease or are a consequence of organ damage. IMACS, an international collaboration, made a consensus recommendation that outcome measures for patients with myositis include tools that measure disease activity, damage, and quality of life. The IMACS network developed one outcome measure to assess myositis disease activity and one to measure organ damage: the myositis disease activity assessment tool and the myositis damage index, respectively.\textsuperscript{156} The disease activity outcome measure is a core set that consists of the six variables listed in Table 85-9. The damage index is recorded by the physician on the basis of the patient’s history and covers several organ systems that can be affected in patients with inflammatory myopathies. To assess the impact on general health, the generic Short Form 36, a self-administered health-related quality of life questionnaire, is recommended. These outcome measures have been developed for clinical trials and research but can also be useful in clinical practice. More detailed information about these outcome measures can be found on the IMACS website: https://dir-apps.niehs.nih.gov/imacs.

IMACS has also reached a consensus on what constitutes improvement. Improvement is based on the disease activity core set and is defined as greater than 20% improvement in three of the six variables of the core set, with two or fewer of the variables (except the manual muscle test) worsening by less than 25%. However, this definition of improvement must be validated in longitudinal studies.

**Table 85-9 Disease Activity Measure: Core Set**

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physician’s overall assessment of disease activity on a visual analogue scale (VAS)</td>
<td>The physician’s overall assessment of disease activity on a visual analogue scale (VAS)</td>
</tr>
<tr>
<td>Patient’s or parent’s overall assessment of disease activity (VAS)</td>
<td>The patient’s or parent’s overall assessment of disease activity (VAS)</td>
</tr>
<tr>
<td>Functional assessment (health assessment questionnaire)</td>
<td>The functional assessment (health assessment questionnaire)</td>
</tr>
<tr>
<td>Muscle strength testing (manual muscle test)</td>
<td>The muscle strength testing (manual muscle test)</td>
</tr>
<tr>
<td>Serum levels of at least 2 of 4 muscle enzymes (CK, LDH, AST, ALT)</td>
<td>The serum levels of at least 2 of 4 muscle enzymes (CK, LDH, AST, ALT)</td>
</tr>
<tr>
<td>Extramuscular score (myositis disease activity assessment VAS)</td>
<td>The extramuscular score (myositis disease activity assessment VAS)</td>
</tr>
<tr>
<td>[MYOACT] or myositis intention to treat activity index (MITAX), in which disease activity in 7 organ systems (general symptoms, skin, joints, gastrointestinal tract, pulmonary, heart, and muscles) is scored</td>
<td>The [MYOACT] or myositis intention to treat activity index (MITAX), in which disease activity in 7 organ systems (general symptoms, skin, joints, gastrointestinal tract, pulmonary, heart, and muscles) is scored</td>
</tr>
</tbody>
</table>

ALT, alanine transaminase; AST, aspartate transaminase; CK, creatine kinase; LDH, lactate dehydrogenase.


Full references for this chapter can be found on www.expertconsult.com.
References


Overlap Syndromes

ROBERT BENNETT

The clustering of symptoms and signs into readily recognizable groupings has an important historic precedence in the classification of disease. With the progress of knowledge, such groupings may become more precisely defined in terms of distinctive pathology, specific laboratory findings, and genetic associations. According to current nosology, there are six autoimmune connective tissue diseases (AICTDs):

1. Systemic lupus erythematosus (SLE)
2. Scleroderma (ScI)
3. Polymyositis (PM)
4. Dermatomyositis (DM)
5. Rheumatoid arthritis (RA)
6. Sjögren’s syndrome

All six classic AICTDs are descriptive syndromes without a “gold standard” for diagnosis. The diagnosis of a well-differentiated AICTD is usually readily apparent without recourse to extensive investigations. However, in the early stages, there are often common features such as Raynaud’s phenomenon, arthralgias, myalgias, esophageal dysfunction, and positive tests for antinuclear antibodies (ANA). In such cases the diagnosis is not always so obvious; this is often referred to as undifferentiated connective tissue disease (UCTD).1 About 35% of such patients have clinical overlap syndromes, whereas most differentiate into a clinical picture consistent with the traditional description of an AICTD. In some instances one AICTD evolves into another AICTD over time.

The propensity for differentiation into a classic AICTD or the maintenance of an overlap state is often associated with distinctive serologic profiles and major histocompatibility (MHC) linkages. Although most rheumatologists generally feel more comfortable thinking in terms of the classic AICTD paradigms, a case can be advanced for using serologic profiles and human leukocyte antigen (HLA) typing to better understand the clinical features and prognoses. In this respect a careful analysis of the overlap syndromes and their serologic associations has provided insights for understanding the clinical heterogeneity of the AICTDs.2 Researchers have reported numerous clinical correlations of autoantibodies (Table 86.1).

EPIDEMIOLOGY

The reported prevalence of AICTDs is variable, depending on methodology, nature of referral bias, and ethnicity.3 It is generally accepted that Sjögren’s syndrome has the highest prevalence (0.5 to 3.6%), with SLE being much lower at about 15 to 50 per 100,000. Scleroderma, polymyositis, and dermatomyositis are relatively rare AICTDs, occurring in fewer than 10 per 100,000. Experts are increasingly realizing that overlap syndromes of scleroderma and myositis are more common than the “pure” forms of the disease.4 There are no epidemiology studies of overlap syndromes, apart from Japan, where the reported prevalence of mixed connective tissue disease (MCTD) was 2.7 per 100,000. The syndrome of MCTD usually occurs as an isolated finding, but there are reports of a familial occurrence. Unlike SLE, precipitation by sun exposure has not been described in patients with MCTD. Likewise, drug exposure has not been related to the onset of MCTD, although a transient appearance of anti-RNP antibodies has been seen at the initiation...
of procainamide therapy.\(^5\) Vinyl chloride and silica are the only environmental agents that have been associated with MCTD so far.

### AUTOIMMUNITY IN OVERLAP SYNDROMES

Compelling evidence indicates that autoimmunity is often antigen driven by components of subcellular particles, in particular spliceosomes, nucleosomes, and proteasomes.\(^6\)

#### Autoimmunity to Spliceosomal Components

Certain components of the spliceosome are common targets of autoimmunity in the AICTDs.\(^7\) Furthermore, it appears that post-translational modifications of these molecules, as occurs during apoptosis, are often associated with increased immunogenicity.\(^8\) Spliceosomes are complex nuclear particles made up of some 300 distinct proteins and 5 RNAs, which are involved in the processing of pre-messenger RNA (pre-mRNA) into mature “spliced RNA.”\(^9\) Two major spliceosomal subunits are antigenic targets in autoimmunity: (1) small nuclear ribonucleoprotein protein particles (snRNPs) and (2) heterogeneous nuclear RNP particles (hnRNPs).

The snRNPs contain small RNA species ranging in size from 80 to 350 nucleotides that are complexed with proteins.\(^6\) These RNAs contain a high content of uridine and are therefore called U-RNAs; 5 different U-RNAs were defined on the basis of immunoprecipitation (U1, U2, U4, U5, and U6). Autoantibodies to these complexes are mainly directed to the protein components. Anti-Sm antibodies precipitate five proteins with molecular weights of 28,000 (B′B), 16,000 (D), 13,000 (E), 12,000 (F), and 11,000 (G); five of these polypeptides are common to the U1, U2, U4, U5, and U6 RNAs. Anti-RNP antibodies precipitate three proteins with molecular weights of 68,000 (70K), 33,000 (A′), and 22,000 (C); these polypeptides are uniquely associated with U1 RNA (Figure 86-1). The clinical correlates considered to be distinctive of MCTD are associated with the 70-kD specificity with an immunodominant epitope embracing amino acid residue 125 flanked by important conformational residues at positions 119-126 (see Figure 86-1). On the other hand, SLE is associated with anti-Sm antibodies.

The hnRNPs are among the most abundant proteins in the eukaryotic cell nucleus. They contain pre-mRNA associated with 30 small proteins that are all structurally related and have molecular weights of 33 to 43 kD. Nine hnRNP core proteins have been designated A1, A2, B1a, B1b, Blc, B2, C1, C2, and C3. An antibody termed anti-RA33, which

### Table 86-1 Correlations of Autoantibodies with Clinical Features

<table>
<thead>
<tr>
<th>Autoantigen</th>
<th>Clinical Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatoid factor</td>
<td>RA, erosive arthritis, cryoglobulinemia</td>
</tr>
<tr>
<td>Anticyclic citrullinated protein</td>
<td>RA</td>
</tr>
<tr>
<td>Nucleosome</td>
<td>SLE, ScI, MCTD</td>
</tr>
<tr>
<td>Proteasome</td>
<td>SLE, PM/DM, Sjögren’s syndrome, multiple sclerosis</td>
</tr>
<tr>
<td>Sm snRNP</td>
<td>SLE</td>
</tr>
<tr>
<td>Histones H1, H2A, H2B, H3, H4</td>
<td>SLE, UCTD, RA, PBC, generalized morphea</td>
</tr>
<tr>
<td>Ribosomal P</td>
<td>SLE psychosis</td>
</tr>
<tr>
<td>dsDNA</td>
<td>SLE, glomerulonephritis, vasculitis</td>
</tr>
<tr>
<td>ACL/β1-glycoprotein</td>
<td>SLE, thrombosis, thrombocytopenia, miscarriages</td>
</tr>
<tr>
<td>β2-glycoprotein–independent ACL</td>
<td>MCTD (not associated with APL syndrome)</td>
</tr>
<tr>
<td>68-kD peptide of U1-RNP</td>
<td>MCTD, Raynaud’s, pulmonary hypertension</td>
</tr>
<tr>
<td>U1 snRNP</td>
<td>MCTD, SLE, PM</td>
</tr>
<tr>
<td>hnRNP-A2 (also called RA-33)</td>
<td>MCTD, RA, erosive arthritis in SLE and ScI</td>
</tr>
<tr>
<td>Ro/La</td>
<td>Sjögren’s, SLE, congenital heart block, photosensitivity, PBC</td>
</tr>
<tr>
<td>Fodrin</td>
<td>Sjögren’s, glaucoma, moyamoya disease</td>
</tr>
<tr>
<td>Platelet-derived growth factor</td>
<td>Diffuse and limited Scl</td>
</tr>
<tr>
<td>Topoisomerase I (Scl-70)</td>
<td>Diffuse Scl with prominent organ involvement</td>
</tr>
<tr>
<td>Centromere</td>
<td>Limited Scl, CREST, Raynaud’s, pulmonary hypertension, PBC</td>
</tr>
<tr>
<td>Th/To</td>
<td>Limited Scl</td>
</tr>
<tr>
<td>U3-snRNP</td>
<td>Limited Scl</td>
</tr>
<tr>
<td>hnRNP-I</td>
<td>Scl (early diffuse and limited)</td>
</tr>
<tr>
<td>RNA polymerases I and III</td>
<td>Scl (diffuse with renovascular hypertension)</td>
</tr>
<tr>
<td>Fibrillarin</td>
<td>Severe generalized Scl</td>
</tr>
<tr>
<td>Ku</td>
<td>Myositis overlap, primary pulmonary hypertension, Graves’ disease</td>
</tr>
<tr>
<td>U5 snRNP</td>
<td>Myositis overlap</td>
</tr>
<tr>
<td>PM/Scl</td>
<td>Myositis overlap with arthritis, skin lesions, mechanic’s hands</td>
</tr>
<tr>
<td>U1-RNP</td>
<td>Myositis overlap with arthritis and interstitial lung disease</td>
</tr>
<tr>
<td>Proteinase-3</td>
<td>Dermatomyositis</td>
</tr>
<tr>
<td>Myeloperoxidase</td>
<td>Granulomatosis with polyangiitis (formerly Wegener’s granulomatosis), pulmonary capillaritis</td>
</tr>
<tr>
<td>Endothelial cell</td>
<td>Churg-Straus, pauci-immune glomerulonephritis</td>
</tr>
<tr>
<td>α-Enolase</td>
<td>Pulmonary hypertension, severe digital gangrene</td>
</tr>
<tr>
<td>Angiotensin-converting enzyme 2</td>
<td>Behçet’s, RA, MCTD, Scl, Takayasu’s</td>
</tr>
</tbody>
</table>

ACL, anticoagulant; AICTDs, autoimmune connective tissue diseases; APL, antiphospholipid syndrome; CREST, syndrome of calcinosis, Raynaud’s phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasias; DM, dermatomyositis; hn, heterogeneous nuclear; MCTD, mixed connective tissue disease; PBC, primary biliary cirrhosis; PM, polymyositis; RA, rheumatoid arthritis; RNP, ribonucleoprotein particle; ScI, scleroderma; SLE, systemic lupus erythematosus; sn, small nuclear; UCTD, undifferentiated connective tissue disease.
 targets the 33-kD hnRNP-A2, is particularly interesting because it is found in about one-third of sera from patient with RA, SLE, and MCTD. It also has associations with patient subsets of erosive arthritis in SLE, scleroderma, and MCTD and predicts the eventual development of RA in patients with early polyarthritis. Importantly, this association with anti-RA33 is not seen in scleroderma (sine erosions), PM, or overlaps of PM/Scl or PM/DM. The antigenic epitopes of hnRNP-A2 contain two RNA binding regions at the N-terminal end and a glycin-rich C-terminal region. Certain disease subsets target these two RNA binding regions differently. For instance, RA and SLE sera preferentially react with the complete second RNA binding domain, whereas MCTD sera target an epitope that spans both RNA binding domains.

### Autoimmunity to Nucleosomal Components

Nucleosomes are the compact building blocks of chromatin and consist of an octamer of two copies of histones H2A, H2B, H3, and H4, wrapped around approximately 146 base pairs of DNA (Figure 86-2). During apoptosis endonucleases cleave chromatin with the liberation of nucleosomal particles. Following the release into the cytoplasm, nucleosomes migrate to the surface of the dying cell and thus become accessible to B cell receptors. The development of autoimmunity has been linked to defective phagocytosis of apoptotically released constituents. Nucleosomal antibodies are directed to antigenic determinants on the intact nucleosome rather than its individual components, DNA and histones. In a study of 496 patients with 13 different AICTDs and 100 patients with hepatitis C, antinucleosome antibodies were found in the sera of patients with SLE (71.7%), Scl (45.9%), and MCTD (45.0%).

### Autoimmunity to Proteasomal Components

The 26S proteasome is a large subcellular particle involved in the degradation of proteins that have been tagged with ubiquitin, resulting in the generation of peptides for presentation by the MHC class I molecules (Figure 86-3). There is good evidence that it is the target of an autoimmune response in AICTD. Antibodies to proteasomal subunits have been reported in patients with autoimmune myositis, systemic lupus erythematosus, and primary Sjögren's syndrome. Circulating 20S proteasomes (c20S) subunits appear to have an association with disease activity in MCTD and SLE.

### Generation of Autoimmunity

The antibody response to just one component of an intracellular structure such as a spliceosome will result in the uptake of the entire particle by antigen-presenting cells

---

**Figure 86-1** The spliceosome is made up of five small nuclear RNAs (snRNAs) complexed with proteins to form a small nuclear ribonucleoprotein particle (snRNP). This subcellular structure is responsible for splicing introns from pre-mRNA to form mRNA via a 5’-splice recognition site. Antibodies to various spliceosomal constituents are a common feature of autoimmune rheumatic disorders with a tendency to associate with different clinical profiles (see Table 86-1). The U1 small nuclear RNP (U1-snRNP) particle of the spliceosome is composed of U1-RNA, RNP proteins (70 kD, A, and C), and common Sm proteins (B’B, D, E, F, and G). The structure of U1-RNA consists of single-stranded RNA and double-stranded RNA called stem loops (SL) 1, 2, 3, and 4. An electron density map of the functional core of U1-snRNP at 5.5 Å resolution has enabled the spatial visualization of the RNA and placement of the seven Sm proteins, U1-C, and U1-70K into the map. A striking feature is the amino (N)-terminal polypeptide of U1-70K, which extends over a distance of 180 Å from its RNA binding domain, wraps around the core domain consisting of the seven Sm proteins, and finally contacts U1-C, which is crucial for 5’-splice-site recognition. (Modified from Newman J: Structural studies of the spliceosome, Curr Opin Struct Biol 20:82–89, 2010; and Pomeranz AD: Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution, Nature 458:457–480, 2009.)
disorders is characterized by differential degrees of epitope spreading. The widest range of antibodies, to both snRNP and hnRNP, is seen in SLE; a more restricted antisn spliceosomal antibody repertoire to snRNP and hnRNP is seen in MCTD; and in RA the antisn spliceosomal antibody repertoire is restricted to hnRNP.22 In general the autoimmune rheumatic diseases are characterized by the production of autoantibodies that recognize evolutionarily conserved molecules. The mechanism whereby these “hidden” intracellular molecules become autoantigens is an area of ongoing research. The two main theories are apoptotic modification23 and molecular mimicry.24 Proteins modified during apoptosis can be presented to the immune system in ways that bypass tolerance to self-proteins.6 Although rheumatic disease autoantigens are not unified by common structure or function, they have the common feature of becoming clustered and concentrated in the surface blebs of apoptotic cells. A population of smaller blebs contains fragmented endoplasmic reticulum and ribosomes, as well as the ribonucleoprotein Ro. Larger blebs (apoptotic bodies) contain nucleosomal DNA, Ro, La, and the small nuclear ribonucleoproteins.25 During the process of apoptosis several enzyme systems are upregulated with resulting post-translational modifications of the cleaved proteins. These modifications, which include citrullination, phosphorylation, dephosphorylation, transglutamination, and conjugation to ubiquitin, render the molecules more antigenic. For instance, the U1-70K protein is cleaved by the enzyme caspase-3, converting it into a C-terminally truncated fragment, which contains a major B cell epitope that is preferentially recognized by autoimmune sera.26

![Figure 86-2](image1.png)

**Figure 86-2** The nucleosome is the fundamental repeating unit of chromatin. The central part of the nucleosome is composed of a tetramer composed of two molecules of histones H3 and H4, flanked by two dimers of histones H2A and H2B. This central core is surrounded by two super-helical turns, consisting of 146 base pairs of histone-free DNA. Histone H1 is located at the point where DNA enters and exits the nucleosome. Antibodies to the nucleosome arise early in the evolution of systemic lupus erythematosus—before anti-DNA and antihistone antibodies. Thus the nucleosome is thought to be an important early autoantigen in the development of epitope spreading. Nucleosome antibodies are also found in scleroderma and mixed connective tissue disease. (From Amoura Z, Koutouzov S, Piette C, et al: The role of nucleosomes in lupus, Curr Opin Rheumatol 12:369–373, 2000.)

![Figure 86-3](image2.png)

**Figure 86-3** Most proteins in the cytosol and nucleus are degraded via the proteasome-ubiquitin pathway. The 26S proteasome is a huge complex of ~2.5 mega-daltons, made up of approximately 35 different subunits. It contains a proteolytic core complex, the 20S proteasome, and one or two 19S regulatory complexes that associate with the termini of the barrel-shaped 20S core. The function of proteasomes is twofold: (1) to degrade intracellular proteins that have been tagged with ubiquitin and (2) to generate antigenic peptides for presentation by the major histocompatibility complex class I molecules. Antibodies to proteasomal subunits have been reported in several autoimmune diseases (especially systemic lupus erythematosus and polymyositis/dermatomyositis), and elevated levels of proteasomes have been correlated with disease activity.
The initial stimulus for a first antibody response may be a non–self-protein possessing a peptide region that mimics a self-epitope—so-called molecular mimicry. Environmental stressors such as infection, toxins, drugs, and ultraviolet light may, under some circumstances, induce accelerated apoptosis. A critical limitation to molecular mimicry is the necessity for the antigenic sequence to undergo TCR recognition. Helper T lymphocytes (CD4+) usually recognize peptides of 12 to 16 amino acids in the context of HLA class II molecules. However, in some instances smaller peptides may be recognized. They can be more immunostimulatory than the parent ligand. Thus antigen recognition by T cells is highly degenerate and expands the potential for molecular mimicry. The universe of molecules containing a pentapeptide, for example, is much greater than for 12 amino residue peptide. Once an immune response to one component of an immunogenic molecular complex has been elicited, other proteins/epitopes of the complex may become antigenic by the same process of epitope spreading.15

**UNDIFFERENTIATED CONNECTIVE TISSUE DISEASE**

**KEY POINTS**

- Nearly all patients with a UCTD have Raynaud’s phenomenon in combination with an unexplained synovitis.
- Nail-fold capillary microscopy is useful in evaluating the potential pathology.
- Antibody profiles are useful in predicting the eventual clinical features: U1-RNP antibodies predict the differentiation into MCTD. DNA antibodies predict the differentiation into SLE. Nucleolar antibodies predict the differentiation into systemic sclerosis (SSc).
- Synthetase and PM/Scl antibodies predict the differentiation into a myositis overlap syndrome.

Rheumatologists frequently see patients who present with a weakly positive ANA and nonspecific symptoms such as arthralgias, fatigue, and cold sensitivity. The critical question in such patients is “will they develop a connective tissue disease?” or “do they have fibromyalgia?”

The answer to this question is not always straightforward because fibromyalgia is not a diagnosis of exclusion; it is a common comorbidity with the well-defined CTDs and is often associated with cold-induced vasospasm.29 An algorithm for diagnosing UCTDs is given in Figure 86-4. In the early stages of a CTD, there may be just one or two suspicious clinical and laboratory features, but a definitive diagnosis cannot always be made. In such cases a working diagnosis of undifferentiated connective tissue disease (UCTD) may be appropriate.29 Most patients with this UCTD have Raynaud’s phenomenon with or without an unexplained polyarthritis and a positive ANA with usually just a single autoantibody specificity, often anti-Ro and anti-RNP.30 A 5-year follow-up study of 665 patients with UCTD reported that only 34% developed a well-defined CTD (RA—13.1%, Sjögren’s—6.8%, SLE—4.2%, MCTD—4%, Scl—2.8%, systemic vasculitis—3.3%, and PM/DM—0.5%).31 Certain combinations of features are predictive for the development of an established CTD: Polyarthritis plus U1-RNP antibodies predict MCTD, sicca symptoms plus anti-SSA/SSB antibodies predict Sjögren’s syndrome, Raynaud’s phenomenon plus a nucleolar ANA pattern predict Scl, polyarthritis plus high levels of rheumatoid factor (RF) predict RA, and fever/serositis plus a homogenous ANA pattern or anti-dsDNA antibodies predict progression into SLE (Table 86-2). The identification of a pathologic nail-fold capillary pattern can provide some early indication that the UCTD may progress to systemic sclerosis (SSc) or MCTD (Figure 86-5).32 Low levels of vitamin D are also reported to be a risk factor for the development of UCTDs and should be evaluated and corrected in all such patients.33

**SCLERODERMA OVERLAPS**

**KEY POINTS**

- Scleroderma overlap syndromes include scleroderma variants such as calcinosis, Raynaud’s phenomenon, esophageal involvement, sclerodactyly, and telangiectasia (CREST), myositis associated with sclerodactyly, and MCTD.
- Raynaud’s phenomenon is often the first clinical feature of SSc overlaps and must be distinguished from primary cold Raynaud’s (i.e., cold-induced vasospasm).
- The finding of thickened and dilated capillaries on nail-fold microscopy and pathologic autoantibodies (e.g., Scl-70, anticentromere, PM/Scl, U1-RNP) are important clues about the development of an overlap syndrome.
- CREST has a common overlap with primary biliary cirrhosis.
- Pulmonary fibrosis and pulmonary hypertension are the main causes of morbidity/mortality.
- Scleroderma-like disorders (e.g., eosinophilic fascitis, scleromyxedema, nephrogenic fibrosis, scleredema) need to be considered in the differential diagnosis of scleroderma overlaps.

**Table 86-2** Disorders Associated with Increased Fibrosis

<table>
<thead>
<tr>
<th>Localized</th>
<th>Systemic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphea</td>
<td>Scleroderma</td>
</tr>
<tr>
<td>Scleromyxedema</td>
<td>Metastatic carcinoid</td>
</tr>
<tr>
<td>Eosinophilic fasciitis</td>
<td>Nephrogenic systemic fibrosis</td>
</tr>
<tr>
<td>Peyronie’s disease</td>
<td>Pathergy</td>
</tr>
<tr>
<td>Dupuytren’s contracture</td>
<td>Pachydermoperiostitis</td>
</tr>
<tr>
<td>Pachydermoperiostitis</td>
<td>Idiopathic pulmonary fibrosis</td>
</tr>
<tr>
<td>Sclerosing cholangitis</td>
<td>Cryptogenic fibrosis</td>
</tr>
<tr>
<td>Primary biliary cirrhosis</td>
<td></td>
</tr>
</tbody>
</table>

| PM/DM | 0.5% |

---

**Scleroderma Overlap Syndromes**

- Scleroderma overlap syndromes include scleroderma variants such as calcinosis, Raynaud’s phenomenon, esophageal involvement, sclerodactyly, and telangiectasia (CREST), myositis associated with sclerodactyly, and MCTD.
- Raynaud’s phenomenon is often the first clinical feature of SSc overlaps and must be distinguished from primary cold Raynaud’s (i.e., cold-induced vasospasm).
- The finding of thickened and dilated capillaries on nail-fold microscopy and pathologic autoantibodies (e.g., Scl-70, anticentromere, PM/Scl, U1-RNP) are important clues about the development of an overlap syndrome.
- CREST has a common overlap with primary biliary cirrhosis.
- Pulmonary fibrosis and pulmonary hypertension are the main causes of morbidity/mortality.
- Scleroderma-like disorders (e.g., eosinophilic fasciitis, scleromyxedema, nephrogenic fibrosis, scleredema) need to be considered in the differential diagnosis of scleroderma overlaps.
Several fibrotic conditions may mimic scleroderma (see Table 86-2). Scleroderma itself has a widespread heterogeneity of disease expression ranging from a diffuse cutaneous disease, with a poor prognosis, to a limited cutaneous involvement, with generally a good prognosis. Furthermore, some patients with Scl have a prominent overlap with other connective tissue diseases. In many cases, these overlaps occur in patients who do not have prominent skin involvement (sine scleroderma) or with the limited form of the disease—CREST. Approximately 90% of patients with Scl have a positive ANA. Scleroderma-related antibodies include topoisomerase I (Scl-70), anticentromere (ACA),

hnPNP-I, RA33, p23, p25, RNA polymerase I (RNAP-1), RNA polymerase III, U1-RNP, PM/Scl, fibrillarin, histone, Ku, endothelial cell, and Th/To (see Table 86-1).

A German registry for scleroderma has reported on patterns of organ involvement in two subsets of 1483 SSc patients. Limited distal skin involvement (distal to the knee and elbows) was seen in 46% (the leSSc group), and 33% had progressive widespread scleroderma (rapid involvement of trunk, face, and extremities—the dcSSc group). An overlap syndrome was seen in 11%, and 8% were undifferentiated. The extent of organ involvement varied between subgroups. For instance, musculoskeletal involvement was seen in 68% of the overlap group compared with 57% of the dcSSc group. Pulmonary fibrosis (56%) and pulmonary hypertension (19%) were most common in the dcSSc group, but pulmonary hypertension was seen in 15% of the dcSSc group.

Specific antibody profiles tend to be associated with distinctive patterns of morbidity and mortality. Patients possessing anticentromere, anti-U3 snRNP, and anti-Th/To antibodies tend to have the limited form of Scl, whereas anti-Scl-70, ACA, and anti-RNAP are associated with diffuse skin involvement and systemic disease. Anti-PM/Scl antibodies are associated with a myositis/Scl overlap and a tendency to develop pulmonary interstitial disease. About 60% of patients with scleroderma have obvious synovitis, and 35% are positive for RF. Erosive arthritis in Scl has an association with anti-RA33; the Scl component in such overlap patients is often an incomplete form of CREST. The limited form of scleroderma has a well-documented overlap with primary biliary cirrhosis (PBC). The distinctive antibody association of scleroderma with PBC is antimitochondrial antibodies. Conversely, anticentromere antibodies have been found in 10% to 29% of patients with PBC; approximately half developed some features of the CREST syndrome. Hence a serologic overlap between the two syndromes is more prevalent in the clinical overlap. Low-grade muscle involvement is not uncommon in scleroderma, being described in between 50% and 80% of patients. A European review of 114 scleroderma overlap patients reported a 95% PM/Scl antibody positivity with 80% having an inflammatory myositis. This “scleromyositis” differed from MCTD by coexistent features of dermatomyositis (myalgia, myositis, Gottron sign, heliotrope rash,
calcinosi(s), but no overlap SLE features, as is characteristic of classic MCTD. Many of these patients had a deforming arthritis of the hands. In general they had a chronic benign course, and most were steroid responsive. Scleroderma lupus overlaps are less common. However, Scl patients often have antinuclear antibodies other than ACA and Scl-70.

Nonscleroderma fibrotic disorder may be mistaken for a scleroderma overlap at initial presentation (see Table 86-2); although these disorders may have some systemic involvement, they seldom exhibit overlap features with other AICTDs.

Nephrogenic systemic fibrosis (NSF) is a fibrotic disorder that develops in some patients following exposure to gadolinium-containing contrast agents; most patients have pre-existing renal disease.\(^4^2\) Histologically there is fibroblast proliferation, thickened collagen bundles, and deposits of mucin, similar to those observed in scleromyxedema. The clinical presentation is a rapid progression with confluence of initially focal areas of indurated skin (Figure 86-6). The face is usually spared, but joint contractures may occur at the elbows and knees, and systemic involvement with pulmonary and neurologic symptoms can develop in refractory cases. NSF is usually nonresponsive to corticosteroids and immunosuppressive therapy.

Eosinophilic fasciitis presents with limited scleroderma-like skin changes involving the extremities (see Figure 86-6). The correct diagnosis is suggested by finding a peripheral eosinophilia and a hypergammaglobulinemia.\(^4^3\) The definitive diagnosis is established by a full-thickness skin biopsy that shows a diffuse inflammation of the fascia. Initial treatment is with corticosteroids (prednisone 0.5 to 1 mg/kg) tapering according to the clinical response; some patients need to continue moderate-dose corticosteroids for up to 2 years. Methotrexate and mycophenolate may be used in refractory cases.

Scleromyxedema is characterized by cutaneous mucinosi(s) and is often associated with a gammopathy, usually IgM and light chains.\(^4^4\) The mucinous skin lesions appear as waxy papules on the face, neck, and limbs. If the papules coalesce, it may be mistaken for scleroderma (see Figure 86-6). Systemic involvement may occur with dysphagia, proximal muscle weakness, pulmonary, cardiac, and renal complications. It is difficult to manage; corticosteroids are usually tried initially, in refractory cases some benefit has

been reported for intravenous immunoglobulin and thalidomide.

Scleroderma is a cutaneous mucinosis that often starts with a febrile episode and resolves spontaneously. More chronic scleroderma has been associated with paraproteinemias including multiple myeloma and diabetes mellitus. The dermis is thickened with increased collagen glycosylation, as in diabetic stiff skin syndrome. The face and neck are commonly involved, and there is relative sparing of the hands and feet (see Figure 86-6). Systemic organ involvement is rare, but a monoclonal gammopathy is sometimes seen. Such cases need to be worked up for lymphoma. Refractory cases have been helped by local radiotherapy.

**MYOSITIS OVERLAPS**

**KEY POINTS**

Myositis overlap syndromes are more common than the classic descriptions of PM or DM.

- Amino-acyl tRNA synthetase antibodies (ARS) are associated with myositis, arthritis, and interstitial lung disease.

- Arthritis and interstitial lung disease may antedate the appearance of myositis in patients with ARS.

- Antibodies to synthetas, signal recognition particle (SRP), and nucleoporin tend to be associated with corticosteroid unresponsiveness.

- Antibodies to U1-RNP, PM/Scl, or Ku are associated with corticosteroid responsiveness.

- Antibodies to 155-kD and 140-kD proteins have been associated with an increased risk of myositis-associated malignancy.

Polymyositis (PM), dermatomyositis (DM), and inclusion body myositis (IBM) are the classic idiopathic inflammatory myopathies (IIM), yet the same clinical picture and investigational findings may be found in patients with SLE, Scl, MCTD, and Sjögren’s syndrome. Such overlaps, especially with scleroderma, have been reported as being more common in the classic description of polymyositis. When clinical overlaps emerge, they are most commonly associated with specific autoantibodies, namely anti-PM-Scl, anti-Ku, U1-RNP, Jo-1, SRP, and ARS. The arthropathy associated with polymyositis is characterized by deforming subluxations (particularly of the distal interphalangeals and thumbs) with only minor erosive changes. Another myositis overlap syndrome is seen in patients with amino-acyl tRNA synthetase antibodies (ARS). This is a family of enzymes that catalyze the transfer of a specific amino acid to its cognate transfer RNA—the commonest association is with anti-Jo-1 (histidine-tRNA synthetase). The clinical syndromes associated with the various antisynthetase antibodies are similar, with remissions and exacerbations characterized by inflammatory myositis, fever, Raynaud’s syndrome, and skin problems (mechanic’s hands). The arthritis of ARS may initially mimic RA with an inflammatory arthritis and nodules; erosions, however, do not occur. Interstitial lung disease may be a presenting clinical feature of patients with ARS antibodies, with myopathy occurring much later. The association of myositis in patients with anti-U1-RNP antibodies is usually seen in the context of MCTD. Antibodies to the signal recognition particle (SRP) have been reported in 4% of patients with Scl/PM overlap; these patients usually have a severe, rapidly progressive myositis with prominent muscle fiber necrosis without much inflammatory cell infiltration.

A 2006 clinical and longitudinal study of 100 consecutive French Canadian patients with idiopathic IIM concluded that the original Bohan and Peter classification of inflammatory myopathies should be abandoned because 60% of patients with IIM were found to have an overlap syndrome. In this study an overlap syndrome was based on the presence of an inflammatory myopathy as per the Bohan and Peter classification, plus at least one clinical overlap feature (Table 86-3), or one of the following autoantibodies: synthetases, centromere, topo I, RNA-polymerases I or III, Th, U1-RNP, U2-RNP, U3-RNP, U5-RNP, PM/Scl, Ku, SRP and nucleoporins (see Table 86-3). The distinction between classic PM/DM and an overlap syndrome was reported to be of prognostic/therapeutic significance because classic PM nearly always pursued a chronic course with 50% of patients being initially unresponsive to corticosteroid therapy. Pure dermatomyositis was almost always chronic, but most had an initial response to corticosteroids. On the other hand, myositis overlap syndromes (usually with scleroderma features) were almost always responsive to corticosteroids (≈90% response rate). When overlap patients were divided according to antibody subsets, antisynthetase, SRP, and nucleoporin autoantibodies were markers for treatment-resistant myositis, whereas autoantibodies to U1-RNP, PM/Scl, or Ku were markers for corticosteroid responsiveness. Patients with autoimmune myositis, especially dermatomyositis, are at risk of developing cancer, and it has been problematic as to how far and how often one should pursue a malignancy workup. It is now apparent that the finding of an antibody against 155-kD and 140-kD protein specificities (anti-155/140 antibody) signifies a significant risk for the co-occurrence of a malignancy and points to the need for a thorough cancer workup.

**MIXED CONNECTIVE TISSUE DISEASE**

**KEY POINTS**

The clinical overlap features of MCTD (i.e., Scl, SLE, and IIM) seldom occur concurrently but develop sequentially over the course of months or years.

- Raynaud’s phenomenon is seen in nearly all patients with MCTD; if Raynaud’s syndrome is absent, the diagnosis should be reconsidered.

- About 25% of MCTD patients develop renal involvement—usually membranous glomerulonephritis. Proliferative glomerulonephritis is uncommon in MCTD.

- Serious CNS involvement is rare in MCTD; the commonest findings are trigeminal neuropathy and sensorineural hearing loss.

- Pulmonary hypertension is the commonest cause of death in MCTD patients and should be screened for on an ongoing basis.
MCTD was described by Sharp and colleagues in a 1971 paper reporting an overlap of SLE, Scl, and PM. This was the first overlap syndrome defined in terms of a specific antibody—namely antibodies to a ribonuclease-sensitive extractable nuclear antigen (ENA). Over the past 38 years, many studies have explored the clinical correlates of this antibody system (now called U1-RNP).

Serologic Features

The basic premise of the MCTD concept is that the presence of high-titer anti-U1-RNP antibodies modifies the expression of an AICTD in ways that are relevant to prognosis and treatment. The first clue to diagnosing MCTD is usually a positive ANA with a high-titer speckled pattern. The titer is often greater than 1:1000 and sometimes greater than 1:10,000. This finding should prompt the measurement of antibodies to U1-RNP, Sm, Ro, and La. It is also pertinent to note whether the serum contains antibodies to dsDNA and histones because patients destined to follow a course most consistent with MCTD have sera with predominant U1-RNP reactivity. Antibodies to dsDNA, Sm, and Ro are occasionally seen as a transient phenomenon in patients with MCTD. But when they are found consistently, as the predominant antibody system, the clinical picture is usually more consistent with classic SLE. Antibodies to the 70-kD antigen, especially in its apoptotic form, are most closely associated with the clinical correlates of MCTD.

Clinical Features

Diagnosis

MCTD is an overlap syndrome that embraces features of SLE, Scl, and PM. These overlap features seldom occur concurrently; it usually takes several years before enough overlapping features have appeared to be confident that MCTD is the most appropriate diagnosis. The commonest clinical associations with U1-RNP antibodies in the early phase of the disease are hand edema, arthritis, Raynaud’s phenomenon, inflammatory muscle disease, and sclerodactyly. No American College of Rheumatology (ACR) criteria are available for the diagnosis of MCTD, but a comparative study reported that two criteria sets, those of Alarcon-Segovia and Kahn, had the best sensitivity and specificity (62.5% and 86.2%, respectively) (Table 86-4). The sensitivity could be improved to 81.3% if the term “myalgia” was substituted for “myositis.” Some patients initially diagnosed as MCTD will evolve into a clinical picture most consistent with SLE or RA; in one long-term follow-up, more than half of the subjects continued to satisfy criteria for MCTD. A comparison of the clinical and serologic features of MCTD with SLE, RA, Scl, and PM/DM is given in Table 86-5.

Early Symptoms

In the early stages most patients destined to develop MCTD cannot be differentiated from the other classic AICTDs.

Table 86-3  Suggested Classification for Inflammatory Myopathies

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>Pure polymyositis</td>
</tr>
<tr>
<td>DM</td>
<td>Pure dermatomyositis</td>
</tr>
<tr>
<td>OM</td>
<td>Overlap myositis: myositis with at least 1 clinical overlap feature and/or an overlap autoantibody</td>
</tr>
<tr>
<td>CAM</td>
<td>Cancer-associated myositis: with clinical paraneoplastic features and without an overlap autoantibody or anti-Mi-2</td>
</tr>
</tbody>
</table>

Bohan and Peter’s Definition of Myositis

1. Symmetric proximal muscle weakness.
2. Elevation of serum skeletal muscle enzymes.
3. Electromyographic triad of short, small, polyphasic motor unit potentials; fibrillations, positive sharp waves, and insertional irritability; and bizarre, high-frequency repetitive discharges.
4. Muscle biopsy abnormalities of degeneration, regeneration, necrosis, phagocytosis, and an interstitial mononuclear infiltrate.
5. Typical skin rash of DM including the heliotrope rash, Gottron sign, and Gottron papules.

Definite myositis: 4 criteria (without the rash) for PM, 3 or 4 criteria (plus the rash) for DM.
Probable myositis: 3 criteria (without the rash) for PM, 2 criteria (plus the rash) for DM.
Possible myositis: 2 criteria (without the rash) for PM, 1 criterion (plus the rash) for DM.

Definition of Clinical Overlap Features

Inflammatory myopathy plus at least 1 or more of the following clinical findings: polyarthritis, Raynaud’s phenomenon, sclerodactyly, scleroderma proximal to metacarpophalangeal joints, typical SSc-type calcinosis in the fingers, lower esophageal or small-bowel hypomotility, DLCO lower than 70% of the normal predicted value, interstitial lung disease on chest radiogram or computed tomography scan, discoid lupus, anti-native DNA antibodies plus hypocomplementemia, 4 or more of 11 American College of Rheumatology criteria for systemic lupus erythematosus, antiphospholipid syndrome.

Definition of Overlap Autoantibodies

Antisynthetases (Jo-1, PL-7, PL-12, OJ, EJ, KS); scleroderma-associated autoantibodies (scleroderma-specific antibodies: centromeres, topoisomerase I, RNA polymerases I or III, Th; and antibodies associated with scleroderma overlap: U1-RNP, U2-RNP, U3-RNP, U5-RNP, Pm-Scl, Ku, and other autoantibodies (signal recognition particle, nucleoporins).

Definition of Clinical Paraneoplastic Features

Cancer within 3 yr of myositis diagnosis, plus absence of multiple clinical overlap features; plus, if cancer was cured, myositis was cured as well.

**Table 86-4 Diagnostic Criteria for Mixed Connective Tissue Disease**

<table>
<thead>
<tr>
<th>Alarcón-Segovia Criteria</th>
<th>Kahn Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Serologic criteria</strong></td>
<td>Anti-RNP at hemagglutination titer of ≥1:1600</td>
</tr>
<tr>
<td><strong>Clinical criteria</strong></td>
<td>1. Swollen hands</td>
</tr>
<tr>
<td></td>
<td>2. Synovitis</td>
</tr>
<tr>
<td></td>
<td>3. Myositis (biologically proven)</td>
</tr>
<tr>
<td></td>
<td>4. Raynaud’s phenomenon</td>
</tr>
<tr>
<td></td>
<td>5. Acrosclerosis</td>
</tr>
<tr>
<td><strong>MCTD present if:</strong></td>
<td>Serologic criterion accompanied by 3 or more clinical criteria, one of which must include synovitis or myositis</td>
</tr>
</tbody>
</table>

ANA, antinuclear antibody; MCTD, mixed connective tissue disease; RNP, ribonucleoprotein particle.


The assumption that a diagnosis of MCTD implies a simultaneous presence of features usually seen in SLE, Scl, and PM is erroneous. It is unusual to see such an overlap during the early course of MCTD, but with the progress of time the overlapping features usually occur sequentially. Early in the course of the disease most patients complain of easy fatigability, poorly defined myalgias, arthralgias, and Raynaud’s phenomenon; at this point in time a diagnosis of RA, SLE, or undifferentiated connective tissue disease (UCTD) seems most appropriate. If such a patient is found to have swollen hands or puffy fingers (Figure 86-7) in association with a high-titer speckled ANA, he or she should be carefully followed for the evolution of overlap features (see Table 86-3). A high titer of anti-RNP antibodies in a patient with UCTD is a powerful predictor for a later evolution into MCTD. Less commonly there is an acute onset of MCTD, which gives little clue to the subsequent course; such presentations have included polymyositis, acute arthritis, aseptic meningitis, digital gangrene, high fever, acute abdomen, and trigeminal neuropathy.

**Fever**

Fever may be a prominent feature of MCTD in the absence of an obvious cause. Fever of unknown origin has been the initial presentation of MCTD; after careful evaluation, fever in MCTD can usually be traced to a coexistent myositis, aseptic meningitis, serositis, lymphadenopathy, or intercurrent infection.

**Joints**

Joint pain and stiffness is an early symptom in nearly all patients who develop the MCTD syndrome. Over the past 2 decades it has become increasingly apparent that joint involvement in MCTD is more common and more severe than in classic SLE. About 60% of patients eventually develop an obvious arthritis, often with deformities commonly seen in RA such as ulnar deviation, swan neck, and boutonnière changes. Radiographs usually show a characteristic absence of severe erosive changes; they often

**Table 86-5 Differential Features of the Classic Autoimmune Connective Tissue Diseases**

<table>
<thead>
<tr>
<th>Clinical Feature</th>
<th>SLE</th>
<th>RA</th>
<th>Scl</th>
<th>PM</th>
<th>MCTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleurisy/pericarditis</td>
<td>+++</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+++</td>
</tr>
<tr>
<td>Erosive joint disease</td>
<td>±</td>
<td>+++</td>
<td>±</td>
<td>+</td>
<td>±</td>
</tr>
<tr>
<td>Raynaud’s phenomenon</td>
<td>++</td>
<td>−</td>
<td>+++</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>Inflammatory myositis</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Sclerodactyly</td>
<td>±</td>
<td>−</td>
<td>+++</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Nonacral skin thickening</td>
<td>−</td>
<td>−</td>
<td>+++</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Interstitial pulmonary fibrosis</td>
<td>+</td>
<td>+</td>
<td>+++</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Pulmonary hypertension</td>
<td>++</td>
<td>±</td>
<td>+</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>Butterfly rash</td>
<td>+++</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Oral ulcers</td>
<td>+++</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Seizures/psychosis</td>
<td>+++</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Trigeminal neuropathy</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+++</td>
</tr>
<tr>
<td>Peripheral neuropathy</td>
<td>++</td>
<td>+</td>
<td>±</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Transverse myelopathy</td>
<td>+++</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Aseptic meningitis</td>
<td>+++</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+++</td>
</tr>
<tr>
<td>Diffuse proliferative glomerulonephritis</td>
<td>+++</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Membranous glomerulonephritis</td>
<td>+++</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Renovascular hypertension</td>
<td>+</td>
<td>−</td>
<td>+++</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Inflammatory vasculitis</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Noninflammatory vasculopathy</td>
<td>−</td>
<td>−</td>
<td>+++</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Esophageal dysmotility</td>
<td>+</td>
<td>±</td>
<td>+++</td>
<td>+</td>
<td>+++</td>
</tr>
</tbody>
</table>

MCTD, mixed connective tissue disease; PM, polymyositis; RA, rheumatoid arthritis; Scl, scleroderma; SLE, systemic lupus erythematosus.
SCLERODERMA, INFLAMMATORY MYOPATHIES, AND OVERLAP SYNDROMES

develop a flexor tenosynovitis, bone edema, and pericapsular inflammation reminiscent of a seronegative spondyloarthropathy (Figure 86-8). A positive RF is found in 50% to 70% of patients; indeed, patients may be diagnosed as having RA and fulfill ACR criteria for RA.

Skin and Mucous Membranes

Most patients with MCTD develop mucocutaneous changes sometime during the course of the syndrome. Raynaud’s phenomenon is the commonest problem and one of the earliest manifestations of MCTD. It may be accompanied by puffy, swollen digits and sometimes total hand edema. In some patients, skin changes commonly associated with classic SLE are prominent findings, particularly malar rash and discoid plaques. Other problems have included buccal ulceration, sicca complex, orogenital ulceration, livedo vasculitis, subcutaneous nodules, and nasal septal perforation.

Muscle

Myalgia is a common symptom in patients with the MCTD syndrome. In most patients there is no demonstrable weakness, EMG abnormalities, or muscle enzyme changes. It is often unclear whether the symptom represents a low-grade myositis, physical deconditioning, or an associated...
CHAPTER 86 | OVERLAP SYNDROMES

fibromyalgia syndrome. The inflammatory myopathy associated with MCTD is similar histologically to IIM, with features of both the vascular involvement of DM and the cell-mediated changes of PM (Figure 86-9). In most patients myositis occurs as an acute flare against a background of general disease activity. Such patients usually respond well to a short course of high-dose corticosteroid therapy. Another scenario is that of a low-grade inflammatory myopathy, which is often insidious in its onset; these patients often have a poor therapeutic response to corticosteroids. Some patients with PM associated with MCTD develop an impressive fever; other patients may give a history of febrile myalgias that were diagnosed as “flu.”

**Heart**

All three layers of the heart may be involved in MCTD. An abnormal electrocardiogram (ECG) is noted in about 20% of patients. The most common ECG changes are right ventricular hypertrophy, right atrial enlargement, and interventricular conduction defects. Pericarditis is the commonest clinical manifestation of cardiac involvement, reported in 10% to 30% of patients. Pericardial tamponade is rare. Involvement of the myocardium is increasingly recognized. In some patients myocardial involvement is secondary to pulmonary hypertension (PAH); this occurs in some 20% of patients and is often asymptomatic in its early stages. The early detection of pulmonary hypertension is increasingly important because there are now more effective therapeutic options. PAH is probably underdiagnosed in its early stages; in a community rheumatology practice setting an elevation of the estimated right ventricular systolic pressure (ERVSP), consistent with the diagnosis of PAH, was found in 13% of previously undiagnosed subjects. This diagnosis should be suspected in patients with increasing exertional dyspnea. Two-dimensional echocardiography with Doppler flow studies is the most useful screening test, with a definitive diagnosis requiring cardiac catheterization showing a mean resting pulmonary artery pressure greater than 25 mm Hg at rest. The development of pulmonary hypertension has been correlated with a nail-fold capillary pattern similar to that seen in Scl, antiendothelial cell antibodies, anticardiolipin antibodies, and anti-U1-RNP antibodies. Both left and right ventricular dysfunction appears to be a common finding that is not always associated with PAH; regular echocardiographic evaluations are recommended for all MCTD patients, especially those with PAH. Elevated levels of anti-U1-RNP antibodies, antiendothelial cell antibodies, serum thrombomodulin, and Willebrand factor are prognostic clues to the development of PAH.

![Figure 86-9](attachment:image.png)

*Figure 86-9* Muscle biopsy from the biceps brachii of a mixed connective tissue disease patient (H&E stain, ×300). A, Moderate variation of fiber size and degenerated fibers (→) with mononuclear cell infiltration. B and C show perivascular inflammatory infiltration and thickening of vessel walls. (Modified from Vianna M, Borges MT, Barba EF, et al: Myositis in mixed connective tissue disease: a unique syndrome characterized by immuno-histopathologic elements of both polymyositis and dermatomyositis, Arq Neuro-Psiquiatr 62:923–934, 2004.)
Lung

Lung involvement occurs in up to 75% of patients. Early symptoms that should prompt further investigations are dry cough, dyspnea, and pleuritic chest pain. Interstitial lung disease (ILD) occurs in up to 50% of subjects. High-resolution computed tomography (HRCT) is the most sensitive test to determine the presence of ILD (Figure 86-10). The commonest HRCT findings are septal thickening and ground-glass opacities. Untreated ILD is usually progressive with the development of severe pulmonary fibrosis in 25% of subjects after 4 years of follow-up. Pulmonary hypertension (PAH) is prognostically the most severe form of pulmonary involvement in MCTD. Unlike Scl, where pulmonary hypertension is often secondary to an interstitial pulmonary fibrosis, PAH in MCTD is usually caused by a bland intimal proliferation and medial hypertrophy of pulmonary arterioles (Figure 86-11).

Gastrointestinal

Gastrointestinal involvement is a major feature of the overlap with scleroderma, occurring in about 60% to 80% of patients. The commonest abdominal problem in MCTD is disordered motility in the upper gastrointestinal tract. There have been case reports of hemoperitoneum, hemobilia, duodenal bleeding, megacolon, pancreatitis, ascites, protein-losing enteropathy, primary biliary cirrhosis, portal hypertension, pneumatosis intestinalis, and autoimmune hepatitis. Abdominal pain in MCTD may result from

Kidney

In the initial description of MCTD, renal involvement was considered to be rare. After some 4 decades of observations, it is now evident that renal involvement occurs in about 25% of patients. However, high titers of anti-U1 RNP antibodies are relatively protective against the development of diffuse proliferative glomerulonephritis, irrespective of whether they occur in a setting of classic SLE or MCTD. When patients with MCTD do develop renal changes, it usually takes the form of a membranous glomerulonephritis. This is often asymptomatic but may sometimes cause an overt nephrotic syndrome. The development of diffuse proliferative glomerulonephritis or parenchymal interstitial disease has been rarely recorded in MCTD. There is increasing recognition that MCTD patients are at risk of developing a renovascular hypertensive crisis similar to the scleroderma kidney.

Figure 86-10  Computed tomography scans of a patient with mixed connective tissue disease and pulmonary hypertension (A, upper zones; B, lower zones). There are bilateral pleural effusions and enlarged bilateral mediastinal lymph nodes in the right paratracheal region and left perivascular areas. The pulmonary artery has a diameter greater than that of the ascending aorta—consistent with the diagnosis of pulmonary hypertension. Both hilar pulmonary arteries are also enlarged. A fairly large pericardial effusion is present. The lung windows show evidence of a diffuse abnormality with linear opacities and some areas of ground-glass attenuation in the upper zones. At the lung bases there are more confluent opacities, both reticular and ground glass, and some air-space consolidation. No honeycombing is identified, and there is no distortion of the lung architecture. (From Saito Y, Terada M, Ishida T, et al: Pulmonary involvement in mixed connective tissue disease: comparison with other collagen vascular diseases using high resolution CT, J Comput Assist Tomogr 26:349–357, 2002.)

Figure 86-11  Intimal hyperplasia and smooth muscle hypertrophy without accompanying inflammation are the characteristic features of the vasculopathy of mixed connective tissue disease. When it occurs in the lung, as shown here, it may give rise to severe pulmonary hypertension. (Note absence of pulmonary fibrosis.) The plexiform lesion (arrow) is a characteristic pathologic finding in this disease process. (From Bull TM, Fagan KA, Badesch DB: Pulmonary vascular manifestations of mixed connective tissue disease, Rheum Dis Clin North Am 31:451–464, 2005.)
bowel hypomotility, serositis, mesenteric vasculitis, colonic perforation, and pancreatitis. Malabsorption syndrome can occur secondarily to small bowel dilation with bacterial overgrowth. Liver involvement in the form of chronic active hepatitis and Budd-Chiari syndrome has been described. Pseudodiverticula, identical to those seen in SCC, may be seen along the antimesenteric border of the colon.

**Nervous System**

In keeping with Sharp’s original description, CNS involvement has not been a conspicuous feature of MCTD. The commonest problem is a trigeminal neuropathy.\(^{81}\) In a review of 81 cases of trigeminal neuropathy seen in a neurologic clinic, the most frequently associated CTDs were undifferentiated connective tissue disease (47%), mixed connective tissue disease (26%), and scleroderma (19%). A sensorineural hearing loss has been reported in nearly 50% of MCTD patients.\(^{82}\) In contrast to CNS involvement in classic SLE, frank psychosis and convulsions have rarely been reported in MCTD.\(^{81}\) Headaches are a relatively common symptom; in the majority of patients they are vascular in origin with many of the components of classic migraine. In a subset of these patients, signs of meningeal irritation develop and examination of the cerebrospinal fluid (CSF) reveals the changes of aseptic meningitis.\(^{83}\) Aseptic meningitis in MCTD has also been described as a hypersensitivity reaction to nonsteroidal anti-inflammatory drugs, in particular sulindac and ibuprofen. There are isolated reports of transverse myelitis, cauda equina syndrome, cerebral hemorrhage, retinal vasculitis, optic neuropathy, progressive multifocal leukoencephalopathy, cold-induced brain ischemia, myasthenia gravis, polyradiculopathy, demyelinating disorder, and peripheral neuropathy. Elevated CSF levels of anti-U1-RNP antibodies, with a predominance of anti-70-kD antibodies, have been reported in both SLE and MCTD patients with diffuse central neuropsychiatric involvement.\(^{84}\) Many patients with AICTDs have changes on brain magnetic resonance imaging that are referred to as unspecific bright objects (UBOs). In many instances UBOs occur in the absence of neurologic symptoms. However, there is a modest correlation between the density and positioning of UBOs; in MCTD these lesions tend to cluster at the corticomedullary junction and periventricular region.\(^{85}\)

**Blood Vessels**

Raynaud’s phenomenon is an early feature of nearly all patients who are eventually diagnosed as having MCTD.\(^{56}\) A bland intimal proliferation and medial hypertrophy affecting medium- and small-sized vessels is the characteristic vascular lesion of MCTD\(^{80}\) (see Figure 86-11) and is the characteristic pathology in pulmonary hypertension and renovascular crises. Both nail-fold capillary microscopy\(^{87}\) and color Doppler\(^{88}\) are useful in distinguishing benign primary Raynaud’s from secondary involvement in MCTD and other AICTDs (see Figure 86-5). Fingernail capillaroscopy is abnormal in most MCTD patients with the same pattern of capillary dilation and dropout that has been reported in Scl.\(^{89}\) An angiographic study reported a high prevalence of medium-size vessel occlusions\(^{85}\) (Figure 86-12). Endothelial cell and anticardiolipin antibodies have been reported to be associated with endothelial dysfunction and the development of atherosclerosis in MCTD.\(^{91}\)

![Figure 86-12](image-url) A, Digital angiogram showing multiple arterial occlusions with collateral formation. B, Digital angiogram showing ulnar artery occlusions. (From Peller JS, Gabor GT, Porter JM, Bennett RM: Angiographic findings in mixed connective tissue disease: correlation with fingernail capillary photomicroscopy and digital photoplethysmography findings, Arthritis Rheum 28:768, 1985. Reprinted with permission of the American College of Rheumatology.)
Blood
Hematologic abnormalities are a common finding in MCTD. Anemia is found in 75% of patients, and the usual profile is most consistent with the anemia of chronic inflammation. A positive Coombs test is seen in about 60% of patients, but an overt hemolytic anemia is uncommon. In SLE, a leukopenia affecting mainly the lymphocyte series is seen in about 75% of patients and tends to correlate with disease activity. Less common associations have been thrombocytopenia, thrombotic thrombocytopenia purpura, and red cell aplasia. Hypocomplementemia has been described in several studies; it is not as prevalent as in classic SLE and has not been correlated with any particular clinical situation. Positive tests for RF have been found in about 50% of patients. The presence of RF is associated with more severe degrees of arthritis, especially if anti-A2/RA33 are also present. Anticardiolipin antibodies or lupus anticoagulants, or both, have also been reported. Unlike the anticardiolipin antibodies found in SLE, they are β2-glycoprotein independent and tend to be associated with thrombocytopenia rather than thrombotic events.

Pregnancy
Reports of maternal and fetal morbidity in MCTD are quite diverse. In a comparison study of patients with MCTD and SLE, the fertility rates in both diseases were unaltered, whereas the parity and fetal wastage was increased in both. Some studies have reported an exacerbation of MCTD during pregnancy and postpartum flares, whereas others have not. Antiphospholipid antibodies have been linked to spontaneous abortion in MCTD. A single case of neonatal “lupus” has been reported, suggesting a pathogenic role for the transplacental passage of anti-U1-RNP antibodies.

Juvenile Mixed Connective Tissue Disease
MCTD may first become apparent in childhood. The average age of onset in one report was 10.7 years. Polymyositis and Raynaud’s phenomena are the most common presenting features. There tends to be a progression of organ involvement with 20% involvement at 5 years and 48% at 10 years. Significant myocarditis, glomerulonephritis, thrombocytopenia, seizures, hemolytic uremic syndrome, an acute coronary syndrome, and aseptic meningitis have been described in isolated cases.

MANAGEMENT OF CONNECTIVE TISSUE DISEASE OVERLAPS
The rational management of overlap CTDs is confounded by the absence of controlled trials. Recommendations for management are based on conventional treatments for SLE, PM/DM, RA, and Scl. General guidelines for treating specific features of the overlap CTDs are given in Table 86-6. Nearly all patients with CTDs experience Raynaud’s phenomenon. Apart from advice on minimizing cold exposure, most patients should be tried on calcium channel blockers (e.g., nifedipine). The use of topical nitrates, endothelin antagonists (e.g., bosentan), phosphodiesterase-5 inhibitors (e.g., tadalafil), and prostaglandin analogues (e.g., iloprost) should be considered in severe refractory cases. Pulmonary hypertension is the main cause of death in MCTD, and patients should be evaluated at regular intervals for the development of this complication because early intervention is the key to effective management. Recent advances in the treatment of pulmonary hypertension have led to reduced morbidity and mortality. Overall effective management requires anticoagulation and vasodilator therapy such as calcium channel blockers or prostacyclin analogues. Long-term treatment with intravenous epoprostenol or prostacyclin improves exercise capacity, hemodynamics, and survival in many patients, as does therapy with inhaled iloprost. Evidence indicates that some patients respond to a regimen of intravenous cyclophosphamide and corticosteroids. Bosentan, an oral endothelin-1 antagonist, has been reported to improve dyspnea and slow PAH progression in MCTD.

The management of overlap syndromes has not been the subject of controlled trials. Therefore management is based on an analysis of the clinical features and the application of management strategies used in the usual treatment of presenting features in terms of inflammatory arthritis, Raynaud’s, inflammatory muscle disease, serositis, interstitial lung disease, pulmonary hypertension, and the gastrointestinal features of scleroderma. By definition, the clinical features of an overlap syndrome will be quite diverse and often change over time. Thus a constant reappraisal of management strategies is necessary at each patient visit.

Many of the problems causing morbidity in overlap syndromes tend to be intermittent and responsive to corticosteroids (e.g., aseptic meningitis, myositis, pleurisy, pericarditis, and myocarditis). On the other hand, nephrotic syndrome, Raynaud’s phenomenon, deforming arthropy, acrosclerosis, and peripheral neuropathies are usually steroid resistant. Many of the scleroderma-like gastrointestinal problems can be managed according to the usual practice in scleroderma such as management of renal crisis with angiotensin-converting enzyme inhibitors, Raynaud’s phenomenon with calcium channel blockers, and gastrointestinal reflux disease with proton pump inhibitors. Fibrotic lung disease is notoriously resistant to corticosteroids and immunosuppressives; there is some evidence that a new class of drugs, the tyrosine kinase inhibitors (e.g., imatinib), may be effective in some patients.

In patients with steroid-resistant thrombocytopenia, refractory myositis, or hemolytic anemia, it is worth considering the use of intravenous gammaglobulin or danazol.

Successful autologous peripheral blood stem cell transplantation has been reported in a patient with refractory myositis and MCTD. Over the long term, concern usually mounts over the total corticosteroid burden and the possibility of inducing an iatrogenic steroid myopathy, nosocomial infection, aseptic necrosis of bone, or accelerated osteoporosis. Routine evaluation of bone mineral density is warranted to detect early presymptomatic osteoporosis and initiation of therapy with antiresorptive agents. Unless contraindicated, all patients should take supplementary calcium and vitamin D. In patients requiring long-term corticosteroids it would seem reasonable to use antimalarials or methotrexate in an attempt to minimize the cumulative
Table 86-6 Guidelines for Managing Overlap Syndromes

<table>
<thead>
<tr>
<th>Problems</th>
<th>Treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue, arthralgia, myalgia</td>
<td>NSAIDs, antimalarials, low-dose prednisone (&lt;10 mg/day); trial use of modafinil</td>
</tr>
<tr>
<td>Arthritis</td>
<td>NSAIDs, antimalarials, methotrexate. Consider TNF inhibition*</td>
</tr>
<tr>
<td>Raynaud’s phenomenon</td>
<td>Keep warm, avoid finger trauma, avoid β-blockers, stop smoking; dihydropyridine calcium channel blocker (e.g., nifedipine); α-sympatholytic (e.g., prazosin); consider endothelin receptor antagonist (e.g., bosentan) in recalcitrant cases</td>
</tr>
<tr>
<td>Acute-onset digital gangrene</td>
<td>Local chemical sympathectomy (infiltration of lidocaine at base of involved digit), anticoagulation, topical nitrates; consider hospitalization for intra-arterial prostacyclin; start endothelin receptor antagonist therapy</td>
</tr>
<tr>
<td>Pleurisy</td>
<td>NSAID or short course of prednisone (~20 mg/day)</td>
</tr>
<tr>
<td>Pericarditis</td>
<td>NSAID or short course of prednisone (~20 mg/day); tamponade will require percutaneous or surgical drainage</td>
</tr>
<tr>
<td>Aseptic meningitis</td>
<td>Discontinue NSAIDs and give short course of high-dose prednisone, about 60 mg/day</td>
</tr>
<tr>
<td>Myositis</td>
<td>Acute onset, severe: prednisone 60-100 mg/day</td>
</tr>
<tr>
<td>Chronic, low grade: prednisone, 10-30 mg/day</td>
<td>Consider methotrexate and/or IVIG in recalcitrant cases</td>
</tr>
<tr>
<td>Membranous glomerulonephropathy</td>
<td>Progressive proteinuria: trial of ACE inhibitor; trial of low-dose aspirin combined with dipyridamole</td>
</tr>
<tr>
<td>Nephrotic syndrome</td>
<td>Severe: trial of prednisone 15-60 mg/day plus monthly pulse cyclophosphamide or daily chlorambucil</td>
</tr>
<tr>
<td>Scleroderma-like renal crisis</td>
<td>ACE inhibitor</td>
</tr>
<tr>
<td>Myocarditis</td>
<td>Trial of steroids and cyclophosphamide; avoid digoxin*</td>
</tr>
<tr>
<td>Incomplete heart block</td>
<td>Avoid chloroquine/</td>
</tr>
<tr>
<td>Asymptomatic pulmonary hypertension</td>
<td>Trial of steroids and cyclophosphamide, low-dose aspirin and ACE inhibitors; consider endothelin receptor antagonist (oral bosentan)</td>
</tr>
<tr>
<td>Symptomatic pulmonary hypertension</td>
<td>Intravenous prostacyclin, ACE inhibitors, anticoagulation, endothelin receptor antagonist (oral bosentan); trial of sildenafil; heart-lung transplantation</td>
</tr>
<tr>
<td>Vascular headache</td>
<td>Trial of propranolol and/or alternate-day aspirin, 350 mg</td>
</tr>
<tr>
<td>Autoimmune anemia/thrombocytopenia</td>
<td>Symptomatic use of a triptan (e.g., sumatriptin, eletriptan)</td>
</tr>
<tr>
<td>Thrombotic thrombocytopenic purpura</td>
<td>Immediate infusion of fresh-frozen plasma; may require plasma exchange and transfusion of platelet-depleted RBCs; consider splenectomy in recalcitrant cases</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>Mild: no treatment required</td>
</tr>
<tr>
<td>Intestinal dysmotility</td>
<td>With reflux: proton pump inhibitor; consider Nissen fundoplication</td>
</tr>
<tr>
<td>Intestinal dysmotility</td>
<td>Severe: calcium channel antagonist, alone or in combination with an anticholinergic agent</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>Prokinetic agents such as metoclopramide and erythromycin</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>Ca/Vit D supplements, estrogen replacement or raloxifene; bisphosphonates; nasal calcitonin; carboxyl-truncated PTH analogues such as hPTH-(1-34).</td>
</tr>
<tr>
<td>Heartburn/Dyspepsia</td>
<td>High-dose steroids (~prednisone 80 mg/day) with taper dependent on clinical course. Consider danazol, IVIG, and immunosuppression in recalcitrant cases</td>
</tr>
<tr>
<td>Trigeminal neuropathy</td>
<td>Immediate infusion of fresh-frozen plasma; may require plasma exchange and transfusion of platelet-depleted RBCs; consider splenectomy in recalcitrant cases</td>
</tr>
<tr>
<td>Trigeminal neuropathy</td>
<td>Immediate infusion of fresh-frozen plasma; may require plasma exchange and transfusion of platelet-depleted RBCs; consider splenectomy in recalcitrant cases</td>
</tr>
<tr>
<td>Trigeminal neuropathy</td>
<td>No effective therapy for numbness; trial of an antiepileptic (e.g., gabapentin) or tricyclic antidepressant (e.g., nortryptiline) for pain</td>
</tr>
</tbody>
</table>

*Has been associated with flares in MCTD and SLE.
#Sulindac and ibuprofen have been associated with a hypersensitivity aseptic meningitis.
•Remain alert for steroid myopathy, aseptic necrosis of bone, and accelerated osteoporosis.
•Cardiotoxic at high doses.
#Predisposes to ventricular arrhythmias.
#Predisposes to complete heart block.
•Cannot be used if esophagus is more than mildly involved.
#ACE, angiotensin-converting enzyme; IVIG, intravenous immunoglobulin; NSAID, nonsteroidal anti-inflammatory drug; PTH, parathyroid hormone; RBC, red blood cell; TNF, tumor necrosis factor.

Steroid burden. Antimalarials should be used with caution in overlap patients with a fascicular or bundle branch block due to the risk of causing a complete heart block or an idiosyncratic hepatitis. Digitalis is relatively contraindicated in patients with myocarditis due to the risk of inducing ventricular arrhythmias. As in SLE, the tumor necrosis factor inhibitor etanercept has been reported to exacerbate MCTD. Rituximab has been beneficial in some patients with severe refractory antisynthetase syndrome. Patients with severe hand deformities may be helped by soft tissue release operations and selected joint fusions.

The management of pregnancy presents several special problems in patients with overlap CTDs. Doria and colleagues have provided the following general advice:

1. Patients should be correctly informed about the risk of becoming pregnant.
2. Pregnancies should be planned when the disease is in remission because it increases the probability of successful maternal and fetal outcome.

3. Patients should be regularly monitored during gestation and postpartum by a multidisciplinary team including a rheumatologist, an obstetrician, and a neonatologist.

4. In the case of disease relapse an adequate treatment, even aggressive if necessary, should be recommended because active disease can be more detrimental for a fetus than drugs.

There is often a tendency to assume that all patients with overlap CTDs should be on long-term corticosteroids; this mistake is compounded by the assumption that all medical problems in these patients are related to their underlying overlap CTDs. For instance, apparent flares of discomfort and pain in overlap CTDs may be due to myofascial pain syndrome or fibromyalgia and are thus unresponsive to corticosteroids. Likewise, the feeling of malaise and easy fatigability may be related to a reactive depression or the fact that the patient has become deconditioned. Premature atherosclerosis is now well recognized as a cause of increased morbidity and mortality in AICTDs, and all patients with overlap syndromes need ongoing evaluation for risk factors and appropriate advice and therapy for hypertension and hyperlipidemia. The management of patients with overlap CTDs requires continuing reassessment of an ever-changing pattern of clinical problems and a constant alertness to the iatrogenic disease. As with any disease of unknown etiology, effective management of patients with the overlap CTDs presents a constant and ever-evolving challenge.

OUTCOME

The prognosis for overlap syndromes is often better than the classic AICTDs. For instance, Troynov reported on the follow-up of 100 patients with idiopathic inflammatory myopathy. It was found that the long-term course after treatment with prednisone, with a dose/duration that initially resulted in good symptomatic improvement, was strikingly different; all PM patients (100%) and most DM patients (92%) progressed to chronic myositis, whereas only 58% of overlap patients developed persistent muscle disease. The tendency for overlap patients to develop chronic disease was more common in those with antisynthetase and nucleoparin antibodies and less with antibodies to U1-RNP, PM/Scl, or Ku. Patients with three or four U1 snRNP antibodies (i.e., anti-70 kD, anti-A, anti-C, and anti-U1 snRNA) tended to have minimal renal disease compared with patients with just one or two reactivities. Antibodies to 155-kD and 140-kD proteins in myositis are a risk factor for the development of cancer. There is unequivocal evidence that patients with high-titer U1-RNP antibodies have a low prevalence of serious renal disease and life-threatening neurologic problems; in this sense MCTD can be favorably compared with classic SLE. However, not all patients with MCTD have a favorable prognosis and death may occur from progressive pulmonary hypertension and its cardiac sequelae. A 38-year follow-up of 47 MCTD patients at the University of Missouri reported a favorable course in 62% and continuing active disease in 38%. Eleven (23%) patients had a fatal outcome related to pulmonary hypertension in nine patients and two deaths unrelated to MCTD. It is evident that the course of overlap syndromes is unpredictable; many patients do follow a relatively benign course, but it is major organ involvement that ultimately dictates the morbidity and mortality of the disease.

References


The references for this chapter can also be found on www.expertconsult.com.
First Modern Case: “Periarteritis Nodosa”

In 1866 Kussmaul and Maier reported the case of a 27-year-old tailor who died during a month-long hospital stay. On presentation, the patient was strong enough to climb two flights of stairs to the clinic but “afterward felt so weak that he immediately had to go to bed.” He complained of numbness on the volar aspect of his thumb and the two neighboring fingers on the right hand. Over the ensuing days, “the general weakness increased so rapidly that he was unable to leave the bed, [and] the feeling of numbness also appeared in the left hand.” Muscle paralysis progressed quickly: “Before our eyes, a young man developed a general paralysis of the voluntary muscles … [He] had to be fed by attendants, and within a few weeks was robbed of the use of most of his muscles.”

The patient’s weakness, caused by vasculitic neuropathy (mononeuritis multiplex), was accompanied by tachycardia, abdominal pains, and the appearance of cutaneous nodules over his trunk. His death was described as follows: “He was scarcely able to speak, lay with persistent severe abdominal and muscle pains, opisthotonically stretched, whimpering, and begged the doctors not to leave him … Death occurred … at 2 o’clock in the morning.” At autopsy, grossly visible nodules were present along the patient’s medium-sized arteries. Kussmaul and Maier suggested the name “periarteritis nodosa” for this disease because of the apparent localization of inflammation to the perivascular sheaths and outer layers of the arterial walls, leading to nodular thickening of the vessels. The name was later revised to polyarteritis nodosa (PAN), to reflect the widespread arterial involvement of this disease and the fact that inflammation in PAN extends through the entire thickness of the vessel wall.

Polyarteritis Nodosa as a Reference Point

In addition to its status as the first form of vasculitis recognized, several features of PAN make it a logical reference for classifying many subsequently recognized forms of vasculitis. Because of the importance of that first report in the understanding and classification of vasculitis, the case is described in detail here.
Table 87-1  Classification Scheme of Vasculitides According to Size of Predominant Blood Vessels Involved

<table>
<thead>
<tr>
<th>Primary Vasculitides</th>
<th>Predominantly Large Vessel Vasculitides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takayasu’s arteritis</td>
<td></td>
</tr>
<tr>
<td>Giant cell arteritis (temporal arteritis)</td>
<td></td>
</tr>
<tr>
<td>Cogan’s syndrome</td>
<td></td>
</tr>
<tr>
<td>Behçet’s disease*</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predominantly Medium Vessel Vasculitides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyarteritis nodosa</td>
</tr>
<tr>
<td>Cutaneous polyarteritis nodosa</td>
</tr>
<tr>
<td>Buerger’s disease</td>
</tr>
<tr>
<td>Kawasaki disease</td>
</tr>
<tr>
<td>Primary angitis of the central nervous system</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predominantly Small Vessel Vasculitides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immune complex mediated</td>
</tr>
<tr>
<td>Goodpasture’s disease (anti-glomerular basement membrane disease)*</td>
</tr>
<tr>
<td>Cutaneous leukocytoclasttic angiitis (“hypersensitivity vasculitis”)</td>
</tr>
<tr>
<td>Henoch-Schönlein purpura</td>
</tr>
<tr>
<td>Hypocomplementemic urticarial vasculitis</td>
</tr>
<tr>
<td>Essential cryoglobulinemia†</td>
</tr>
<tr>
<td>Erythema elevatum diutinum ANCA-associated disorders§</td>
</tr>
<tr>
<td>Granulomatosis with polyangiitis (formerly Wegener’s granulomatosis)</td>
</tr>
<tr>
<td>Microscopic polyangiitis†</td>
</tr>
<tr>
<td>Churg-Strauss syndrome‡</td>
</tr>
<tr>
<td>Renal-limited vasculitis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Secondary Forms of Vasculitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miscellaneous Small Vessel Vasculitides</td>
</tr>
<tr>
<td>Connective tissue disorders (rheumatoid vasculitis, lupus erythematosus, Sjögren’s syndrome, inflammatory myopathies)</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>Paraneoplastic</td>
</tr>
<tr>
<td>Infection</td>
</tr>
<tr>
<td>Drug-induced vasculitis: ANCA-associated, other</td>
</tr>
</tbody>
</table>

*May involve small, medium, and large blood vessels.
†Immune complexes formed in situ, in contrast to other forms of immune complex–mediated vasculitis.
‡Frequent overlap of small and medium blood vessel involvement.
§Not all forms of these disorders are always associated with ANCA.
ANCA, antineutrophil cytoplasmic antibody.

Classification by Vessel Size

Because the etiologies of most forms of vasculitis are unknown, the most valid basis for classifying the vasculitides is the size of the predominant blood vessels involved. Under such classification schemes, the vasculitides are categorized initially by whether the vessels affected are large, medium, or small (see Table 87-1 and Figure 87-1). “LARGE” generally denotes the aorta and its major branches (and the corresponding vessels in the venous circulation in some forms of vasculitis, e.g., Behçet’s disease). “Medium” refers to vessels that are smaller than the major aortic branches yet still large enough to contain four elements: (1) an intima, (2) a continuous internal elastic lamina, (3) a muscular media, and (4) an adventitia. In clinical terms, medium vessel vasculitis (see Table 87-1) is generally macrovascular (i.e., involves vessels large enough to be observed in gross pathologic specimens or visualized by angiography). “Small vessel” vasculitis, which incorporates all vessels below macroscopic disease, includes capillaries, postcapillary venules, and arterioles. Such vessels all are typically less than 500 µm in outer diameter. Because glomeruli may be viewed simply as differentiated capillaries, forms of vasculitis that affect glomerulonephritis are considered to be small vessel vasculitides. Table 87-2 presents the typical clinical manifestations associated with small, medium, and large vessel vasculitides.

All discussions of vasculitis classification schemes involving vessel size must acknowledge the frequent occurrence of overlap. Although PAN primarily involves medium-sized arteries, palpable purpura—a manifestation of small vessel disease—is observed in some cases. Despite the possibility of vessel size overlap within individual cases, the categorization of a patient’s vasculitis as primarily large, medium, or small vessel in nature remains enormously useful in focusing the differential diagnosis and initiating plans for treatment.

Additional Considerations in Classification

Many other considerations are important in the classification of vasculitis (Table 87-3): (1) the patient’s demographic profile (see Epidemiology section), (2) the disease’s tropism for particular organs, (3) the presence or absence of granulomatous inflammation, (4) the participation of immune complexes in disease pathophysiology, (5) the finding of characteristic autoantibodies in the patients’ serum (e.g., ANCA, anti-GBM antibodies, or rheumatoid factor), and (6) the detection of certain infections known to cause specific forms of vasculitis.

The organ tropisms of these disorders are illustrated by the following examples. Granulomatosis with polyangiitis (formerly Wegener’s granulomatosis) and herein abbreviated GPA, classically involves the kidneys, upper airways, and lungs. In contrast, Henoch-Schönlein purpura often affects the kidneys, but never the nose or sinuses and almost never the lungs. In contrast to both of these forms of

The fact of vessel size overlap in vasculitis syndromes is acknowledged and discussed subsequently.
vasculitis, Cogan’s syndrome is defined by the simultaneous occurrence of ocular inflammation (most often interstitial keratitis) and sensorineural hearing loss (and, in 10% of cases, a large vessel vasculitis). The histopathologic findings in these three disorders are equally distinctive, ranging from granulomatous inflammation of small to medium vessels (GPA), to IgA deposition in small vessels (Henoch-Schönlein purpura), to large vessel vasculitis centered on the adventitia (Cogan’s syndrome).

The granulomatous features of some forms of vasculitis resemble chronic infections (e.g., infections caused by fungi or mycobacteria) or the inflammation induced by the presence of a foreign body. Granulomatous inflammation is more likely to be found in some organs (e.g., the lung) than in others (e.g., the kidney or skin). Some patients without evidence of granulomatous inflammation at early points in their courses later exhibit such features as their diseases unfold. Patients initially diagnosed with cutaneous leukocytoclastic angiitis or microscopic polyangiitis may be reclassified as having GPA if disease manifestations appear in new organs and granulomatous inflammation is found on biopsy specimens. Table 87-4 presents forms of vasculitis commonly associated with granulomatous inflammation.

Immune complexes are essential to the pathophysiology of some small and medium vessel vasculitides. Immune complex–mediated tissue injury does not produce a single clinical syndrome, but rather applies to many forms of vasculitis and overlaps with injuries caused by other immune mechanisms. Anti-GBM disease (Goodpasture’s disease) is a unique form of immune complex disease in which the immune complexes form in situ rather than in the circulation.6 Complexes of IgA1 are found in Henoch-Schönlein purpura. Immune complexes comprising IgG, IgM, complement components, and the hepatitis C virion characterize most cases of mixed cryoglobulinemia. HBV surface antigen/antibody complexes are present in the circulation and involved tissues of patients with HBV-associated PAN. Rheumatoid factor and complement proteins are found within organs involved by rheumatoid vasculitis.

In contrast, other small and medium vessel vasculitides such as GPA, microscopic polyangiitis, and Churg-Strauss syndrome are disorders associated with “pauci-immune”

Table 87-2 Typical Clinical Manifestations of Large, Medium, and Small Vessel Involvement by Vasculitis

<table>
<thead>
<tr>
<th>Large</th>
<th>Medium</th>
<th>Small</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limb claudication</td>
<td>Cutaneous nodules</td>
<td>Purpura</td>
</tr>
<tr>
<td>Asymmetric blood pressures</td>
<td>Ulcers</td>
<td>Vesiculobullous lesions</td>
</tr>
<tr>
<td>Absence of pulses</td>
<td>Livedo reticularis</td>
<td>Urticaria</td>
</tr>
<tr>
<td>Bruits</td>
<td>Digital gangrene</td>
<td>Glomerulonephritis</td>
</tr>
<tr>
<td>Aortic dilation</td>
<td>Mononeuritis multiplex</td>
<td>Alveolar hemorrhage</td>
</tr>
<tr>
<td>Renovascular hypertension</td>
<td>Microaneurysms</td>
<td>Cutaneous extravascular necrotizing granulomas</td>
</tr>
<tr>
<td></td>
<td>Renovascular hypertension</td>
<td>Splinter hemorrhages</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Constitutional symptoms: fever, weight loss, malaise, arthralgias/arthritis (common to vasculitides of all vessel sizes)
inflammation. “Pauci-immune” refers not to a lack of immunologic involvement in these disorders, but rather to the absence of significant immunoreactant deposition (immunoglobulin or complement) within diseased tissues. Many (but not all) patients with pauci-immune forms of vasculitis have ANCA in their serum. Three decades before the description of ANCA, Godman and Churg observed pathologic links between these three entities, noting that the disorders “group themselves into a compass, [ranging from] necrotizing and granulomatous processes with angiitis, through mixed forms, to vasculitis without granuloma.”

ANCAs (see Chapter 89) are directed against antigens that reside within the primary granules of neutrophils and monocytes. Two types of ANCA seem to be relevant to vasculitis: (1) ANCA directed against proteinase-3 (PR3), a serine protease found within the primary granules of neutrophils and monocytes; and (2) ANCA directed against myeloperoxidase and another serine protease found within the same granules. Although rigorous serologic assays for these antibodies are helpful in diagnosis, evidence for a primary etiologic role of these antibodies in human forms of pauci-immune vasculitis is still lacking. In contrast, anti-GBM antibodies have been proven to play a major role in the pathogenesis of Goodpasture’s disease. In RA, systemic rheumatoid vasculitis occurs only in patients who are rheumatoid factor positive. Rheumatoid factor is believed to play an essential role in the immune complex nature of that disease complication. Finally, although the causes of most forms of vasculitis are unknown, several infections have been linked definitively with specific forms of these diseases (e.g., HBV with some cases of PAN, hepatitis C with type II mixed cryoglobulinemia).

**Table 87-3** Considerations in the Classifications of Systemic Vasculitis

<table>
<thead>
<tr>
<th>Size of predominant blood vessels affected</th>
<th>Epidemiologic features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Gender</td>
</tr>
<tr>
<td>Ethnic background</td>
<td>Pattern of organ involvement</td>
</tr>
<tr>
<td>Pathologic features</td>
<td>Granulomatous inflammation</td>
</tr>
<tr>
<td>Immune complex deposition versus pauci-immune histopathology</td>
<td>Linear staining along glomerular basement membrane</td>
</tr>
<tr>
<td>Presence of ANCA, anti-GBM antibodies, or rheumatoid factor in serum</td>
<td>Demonstration of a specific associated infection (hepatitis B or hepatitis C)</td>
</tr>
</tbody>
</table>

ANCA, antineutrophil cytoplasmic antibody; GBM, glomerular basement membrane.

**Table 87-4** Forms of Vasculitis Associated with Granulomatous Inflammation

| Giant cell arteritis                     |
| Takayasu’s arteritis                    |
| Cogan’s syndrome                        |
| Granulomatosis with polyangiitis        |
| Churg-Strauss syndrome                  |
| Primary angiitis of the central nervous system* |
| Buerger’s disease†                      |
| Rheumatoid vasculitis                   |

*Sometimes granulomatous.
†Giant cells occur within inflammatory thrombi (and are diagnostic of Buerger’s disease) but do not occur within the blood vessel wall.

**Historical Attempts at Classification and Nomenclature**

For decades after this initial description of vasculitis, most forms of systemic inflammatory vascular disease were termed *periarteritis nodosa.* In the 1900s, two major factors led to the recognition of new forms of vasculitis: (1) the use of microscopy in the evaluation of pathologic specimens became routine, and (2) horse serum and sulfonamides came to be employed in the treatment of many medical conditions. These new therapies frequently induced small vessel vasculitides on the basis of serum sickness or “hypersensitivity” phenomena, which were observed readily through the microscope. In some cases, the histopathologic findings of hypersensitivity reactions (e.g., serum sickness) were confused with periarteritis nodosa. The gradual recognition that these syndromes represented departures from PAN spurred interest in the first classification scheme for necrotizing angiitis.

In 1952 Zeek identified five major categories of necrotizing angiitis. The Zeek classification included (1) hypersensitivity angiitis, (2) allergic granulomatous angiitis (Churg-Strauss syndrome), (3) rheumatic arteritis (vasculitis associated with fulminant rheumatic fever), (4) periarteritis nodosa, and (5) temporal arteritis. This classification scheme omitted several forms of systemic vasculitides that were known but not yet described in the English medical literature (e.g., GPA, Takayasu’s arteritis).

**Sources of Confusion in Classification**

Two forms of vasculitis, now termed *microscopic polyangiitis* and *cutaneous leukocytoclastic angiitis,* have been consistent sources of confusion in vasculitis nosology. In the current understanding of systemic vasculitides, these two conditions are separate entities. Microscopic polyangiitis affects capillaries, veins, and arteries (in contrast to PAN) and is recognized to be a disorder associated with ANCA in approximately 70% of cases. In 1923 Wohlwill observed unequivocal evidence of small vessel involvement in cases of vasculitis that he still considered part of the spectrum of “periarteritis nodosa.” Davson and colleagues remarked on two forms of “periarteritis nodosa” with differential effects on the kidney—one with a predilection for medium-sized, muscular arteries and the other with a predilection for small vessels including glomerulonephritis. Davson and colleagues termed this latter form *microscopic periarteritis nodosa.* Swayed by human and animal models of hypersensitivity that showed small vessel disease involving the kidneys, lungs, and other organs, Zeek chose to group microscopic periarteritis nodosa under the heading of *hypersensitivity vasculitis.*

Over the next several decades, hypersensitivity vasculitis came to refer to an immune complex–mediated small vessel of the skin that spared internal organs and often followed drug exposures. The Chapel Hill Consensus Conference (CHCC) (see later) recommended eliminating the term *hypersensitivity* altogether because evidence for
hypersensitivity is lacking in many cases. Participants in the CHCC preferred the term cutaneous leukocytoclastic angiitis because of the disorder's typical confinement to the skin and the usual predominant cell type—the neutrophil. Although cutaneous leukocytoclastic angiitis can mimic the skin features of microscopic polyangiitis, cutaneous leukocytoclastic angiitis does not involve the kidneys, lungs, peripheral nerves, and other internal organs and is not associated with ANCA.

In 1990 the American College of Rheumatology (ACR) performed a study designed to establish criteria for the classification of vasculitis, through the identification of features that distinguished one form of vasculitis from others.18,19 An important caveat: This study was not designed to establish criteria for diagnosis, but rather to facilitate research by permitting the inclusion of similar types of patients in studies. Patients with giant cell arteritis, Takayasu’s arteritis, PAN, GPA, Churg-Strauss syndrome, Henoch-Schönlein purpura, and hypersensitivity vasculitis were included in this study.16,20-25 The findings of the ACR study remain useful for the purposes of the study’s original intention—the insurance of uniform inclusion criteria for patients in research studies. The passage of time and the development of new insights have shown the need for updates, however. First, because the study was performed before the days of reliable and widely available assays for ANCA, ANCA positivity was not considered as a possible classification criterion. Second, the ACR classification criteria study did not include microscopic polyangiitis as a separate disease, but rather lumped such patients under the heading of PAN.22 Third, the study did not define classification criteria for such rarer forms of vasculitis as Cogan’s syndrome and Behçet’s syndrome. As noted subsequently, Behçet’s syndrome is rare in North America but not in countries bordering the Old Silk Route. Classification criteria for this disease have been defined.26

In 1994 the CHCC reviewed the nomenclature of systemic vasculitides. Formal diagnostic criteria were not attempted, but definitions were created for 10 forms of vasculitis (in addition to the 7 forms of vasculitis included in the ACR study, microscopic polyangiitis, Kawasaki disease, and “essential” cryoglobulinemic vasculitis were defined). The CHCC emphasized the important role that ANCA plays in the diagnosis of several forms of vasculitis and carefully distinguished microscopic polyangiitis from classic PAN. The conference defined classic PAN as necrotizing inflammation of medium-sized or small arteries without glomerulonephritis.17 Microscopic polyangiitis was defined as a necrotizing vasculitis with few or no immune deposits that (1) affects small blood vessels (capillaries, venules, or arterioles), (2) often includes glomerulonephritis and pulmonary capillaritis, and (3) is often associated with either myeloperoxidase ANCA’s or PR-3 ANCA’s.

The classification of vasculitis continues to evolve. In the years since the CHCC, it has become clear that hepatitis C plays a major role in 90% of cases that formerly were termed essential mixed cryoglobulinemia. Some cases of this syndrome are not associated with hepatitis C infections, however, and probably have some other infectious etiology. Similarly, with the availability of the HBV vaccine, increasingly fewer cases of PAN are associated with this infection. Most PAN cases have no known cause. As emphasized by Churg,27 cases of PAN currently termed idiopathic do not represent a single entity but almost certainly include several different disorders. Some of these, as indicated by low serum complement levels and measurable immune complexes in the blood, are mediated by immune complex deposition. Others seem to be independent of this mechanism. Finally, although cutaneous leukocytoclastic angiitis may be preferable to hypersensitivity vasculitis in describing small vessel vasculitis confined to the skin, the term fails to acknowledge the few cases in which lymphocytic infiltrates predominate and the nonspecific term “lymphocytic vasculitis” is used.

**Epidemiology**

Accurate definition of the epidemiology of vasculitis confronts several challenges, as follows: (1) the uncommon nature of many forms of vasculitis; (2) the frequent difficulties in making the correct diagnosis of vasculitis (and in distinguishing one form of vasculitis from another); (3) the fact that the etiologies of most types of vasculitis remain unknown; and (4) historical uncertainty with regard to the classification of these conditions. Nevertheless, in recent years, the epidemiology of some forms of vasculitis has been defined with reasonable precision. Table 87-5 presents the major epidemiologic features of several forms of systemic vasculitis.

**Geography**

The epidemiologic features of systemic vasculitis vary tremendously by geography. This variation may reflect

<table>
<thead>
<tr>
<th>Disease</th>
<th>Incidence</th>
<th>Age/Gender/Ethnic Predispositions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giant cell arteritis</td>
<td>United States: 240/1 million</td>
<td>Elsewhere: 240/1 million (Scandinavia)</td>
</tr>
<tr>
<td>Takayasu’s arteritis</td>
<td>3/1 million</td>
<td>300-300/1 million (India)</td>
</tr>
<tr>
<td>Behçet’s disease</td>
<td>3/1 million</td>
<td>3000/1 million (Turkey)</td>
</tr>
<tr>
<td>Polyarteritis nodosa</td>
<td>7/1 million</td>
<td>7/1 million (Spain)</td>
</tr>
<tr>
<td>Kawasaki disease</td>
<td>100/1 million*</td>
<td>900/1 million (Japan)</td>
</tr>
<tr>
<td>Granulomatosis with polyangiitis</td>
<td>4/1 million</td>
<td>8.5/1 million (United Kingdom)</td>
</tr>
<tr>
<td>Henoch-Schönlein purpura</td>
<td>In children: 135-180/1 million; in adults: 13/1 million</td>
<td>Only 10% of cases occur in adults</td>
</tr>
</tbody>
</table>

*Among children younger than 5 years of age.

The genetics of these disorders are polygenic and complex. The rarity of familial cases in vasculitis indicates that the disease has higher incidences in patients of Asian ancestry. Arteritis and GPA occur with an overwhelming predominance among patients lacking HLA-DRB1*04 shared epitope alleles, suggesting that HLA-C and HLA-DR genes influence the risk of extra-articular manifestations, which may be associated with the genotypes 0401/0404.

For some forms of vasculitis, there are striking variations in tendencies to affect specific ethnic groups. Giant cell arteritis and GPA occur with an overwhelming predominance in whites. Takayasu's arteritis and Kawasaki disease have higher incidences in patients of Asian ancestry.

### Genes

Although genetic risk factors are undoubtedly important in the susceptibility to some forms of vasculitis, familial cases are rare (with the exception of giant cell arteritis; see later). The rarity of familial cases in vasculitis indicates that the genetics of these disorders are polygenic and complex. The strongest link between any single gene and vasculitis is the association of HLA-B51 with Behçet's disease. In Behçet's disease, 80% of Asian patients have the HLA-B51 gene. The prevalence of HLA-B51 is significantly higher among patients with Behçet's disease in Japan than among non-disease controls (55% vs. <15%). Among the sporadic cases of Behçet's disease involving white patients in the United States, however, HLA-B51 occurs in less than 15% of cases. In addition to increasing the risk of disease susceptibility in some patients, HLA-B51 increases disease severity. Patients with this gene are more likely to have posterior uveitis, central nervous system involvement, or other severe manifestations.

Reports of familial aggregation in giant cell arteritis are common. Genetic studies have indicated roles for HLA class II alleles such as HLA-DRB1*0401 and HLA-DRB1*0101, albeit the specific associations have varied from study to study.

The greatest progress in understanding the relationship of genetics to systemic vasculitis has come in the area of rheumatoid vasculitis. The contribution of the "shared epitope" found on class II human leukocyte antigen (HLA) molecules (DR4) to the development of rheumatoid arthritis (RA) has been appreciated for 2 decades. Possession of the shared epitope is now known to substantially increase the risk of extra-articular manifestations in RA including vasculitis, at least in Northern European populations. A gene dosage effect for extra-articular RA with severe organ manifestations has been noted; patients with two copies of shared epitope alleles have a substantially higher risk of extra-articular disease manifestations, many of which are mediated by vasculitis. One study reported an association between rheumatoid vasculitis and 0401/0404.

In a case-control study of patients with severe extra-articular RA compared with RA patients without extra-articular disease manifestations, the presence of two HLA-DRB1*04 alleles encoding the shared epitope was associated with extra-articular RA (odds ratio [OR], 1.79; 95% confidence interval [CI], 1.04 to 3.08) and rheumatoid vasculitis (OR, 2.44; 95% CI, 1.22 to 4.89). In a meta-analysis of HLA-DRB1 genotyping studies of patients with rheumatoid vasculitis, rheumatoid vasculitis was found to be associated with the genotypes 0401/0401, 0401/0404, and 0401/0101. An association between rheumatoid vasculitis and class I HLA molecules has also been reported. An analysis of 159 patients with severe extra-articular RA (46 of whom had vasculitis) and 178 RA patients without extra-articular disease reported a strong association between the HLA-C3 allele and vasculitis. Among vasculitis patients, the allele frequency of HLA-C3 was 0.411 compared with 0.199 in RA patients without extra-articular disease (P < .001). The odds ratio for vasculitis in patients with HLA-C3 was 4.15 (95% CI, 2.14 to 8.08).

The association between HLA-C3 and vasculitis was not due to linkage disequilibrium with HLA-DRB1, suggesting that these two genetic risk factors operate through different pathways. HLA-C3 was a stronger predictor of vasculitis in patients lacking HLA-DRB1*04 shared epitope alleles, suggesting that HLA-C and HLA-DR genes influence the RA
disease process through different pathways. Linkage disequilibrium with other genes in the major histocompatibility complex (MHC) could not be excluded in this study, however.

In GPA, the allele frequency of a functional polymorphism, 620W, in the intracellular tyrosine phosphatase gene PTPN22 was found to be increased significantly among ANCA-positive patients compared with healthy controls. Analyses of families with multiple autoimmune disorders have identified this allele as a risk factor for type 1 diabetes, seropositive RA, systemic lupus erythematosus, and autoimmune thyroid disease. In GPA, the allelic association was particularly strong among patients with generalized disease (i.e., vasculitis involving the kidney, lung, eye, and peripheral nervous system).

Study of the relationships between genes and systemic vasculitis is in its infancy. Substantial progress can be anticipated in this area in the future.

Environment

Several environmental and occupational exposures have been linked to the development of vasculitis. Medications and certain infections have well-known associations with vasculitides. As examples, antibiotics such as the penicillins and cephalosporins are common causes of hypersensitivity vasculitis (see Chapter 91), and virtually any medication can trigger this syndrome. Hepatitis B and C and have well-recognized links to polyarteritis nodosa and mixed cryoglobulinemia, respectively. Moreover, both medications and infections can serve as triggers for Henoch-Schönlein purpura.

The strongest environmental exposure, now linked convincingly to Buerger's disease and to rheumatoid vasculitis, is cigarette smoking. Buerger's disease does not occur in the absence of cigarette smoking. The relationship between smoking and Buerger's disease is usually one of primary exposure (usually heavy), but cases related to second-hand smoke have been alleged.

In a case-control study of patients with recent-onset RA, the interactions among smoking, shared epitope genes, and antibodies to citrullinated proteins were studied. A dose-dependent relationship between smoking and the occurrence of anti-cyclic citrullinated peptide antibodies was found. The presence of shared epitope genes was a risk factor for RA only among patients who were positive for anticyclic citrullinated peptide. A major gene-environment interaction between smoking and HLA-DR SE genes was evident in the large subgroup of patients who possessed anticyclic citrullinated peptide antibodies: The combination of smoking history and double copies of shared epitope alleles increased the risk of RA 21-fold. Smoking may trigger RA-specific immune reactions to citrullinated proteins in the context of shared epitope genes.

Associations have been reported, but not confirmed, between exposure to inhaled silica dust and some types of pauci-immune vasculitis. Precise definitions of the relationships between exposures and vasculitis are complicated by difficulties in obtaining reliable measurements of the levels of such exposures, the likelihood of recall bias among patients who are diagnosed with vasculitis, and the choice of appropriate control groups.

Finally, estimates of disease prevalence in vasculitis may be subject to revision because of changing disease definitions. In the ACR classification criteria study, manifestations of small and medium vessel involvement were included in the criteria for PAN. Four years later, the CHCC defined PAN as a form of arterial inflammation limited to medium-sized vessels, sparing capillaries, arterioles, and venules. Under this definition, classic PAN is believed to be a rare condition. Applying the CHCC definition retrospectively, not a single case of classic PAN was reported over a 6-year period in the region of the Norwich Health Authority (United Kingdom), an area that included a population of more than 400,000.

The epidemiologic differences among individual types of vasculitis raise compelling questions about the etiologies of these diseases. Ultimately, better insights into the pathogenesis of these conditions should explain these epidemiologic differences and facilitate the development of more refined classification schemes.

References


The references for this chapter can also be found on www.expertconsult.com.
Giant Cell Arteritis, Polymyalgia Rheumatica, and Takayasu’s Arteritis

DAVID B. HELLMANN

**KEY POINTS**

- Giant cell arteritis affects adults older than 50 years.
- The most common manifestations of giant cell arteritis are constitutional symptoms, headache, jaw claudication, and visual symptoms; almost all untreated patients have an elevated erythrocyte sedimentation rate.
- The diagnosis of giant cell arteritis is usually confirmed by temporal artery biopsy.
- Early treatment of giant cell arteritis can prevent blindness.
- Polymyalgia rheumatica can occur by itself or with giant cell arteritis.
- Polymyalgia rheumatica responds to prednisone 10 to 20 mg/day, whereas giant cell arteritis requires an initial dose of prednisone of approximately 60 mg/day.
- Takayasu’s arteritis most frequently affects the aorta and its major branches in young women.

Giant cell arteritis (GCA) and polymyalgia rheumatica (PMR) are discussed together because they affect similar epidemiologic subsets of patients and often occur together in the same individual. Although GCA is a disease of older people and Takayasu’s arteritis is a disease of younger people, their shared predilection for causing vasculitis of large arteries and their nearly identical histopathologic changes prompt their inclusion in the same chapter.

**GIANT CELL ARTERITIS AND POLYMYALGIA RHEUMATICA**

**American College of Rheumatology Criteria**

Classification criteria for the diagnosis of GCA have been proposed by the American College of Rheumatology (ACR) (Table 88-1). Three classification schemes have been proposed for PMR (Table 88-2). The provisional criteria for 2012 were developed in part to incorporate information from ultrasound evaluation.

**Definitions**

**Giant Cell Arteritis**

GCA is the most common form of systemic vasculitis in adults. The disease affects primarily the extracranial branches of the carotid artery in patients older than 50 years. The most feared complication of GCA is irreversible loss of vision. Because the cause of GCA is unknown, various names—including temporal arteritis, cranial arteritis, and granulomatous arteritis—have been used to highlight different salient features. All the names for this disease have both merits and shortcomings. The designation temporal arteritis or cranial arteritis, for example, conveys how frequently the temporal arteries or other cranial arteries are involved but fails to capture GCA’s more widespread nature. Although granulomatous arteritis and GCA pay homage to an important pathologic finding, this focus is undeserved because giant cells are absent in about half the cases and may be present in other forms of vasculitis. With no perfect name available, this chapter bows to convention and refers to this disease as GCA.

**Polymyalgia Rheumatica**

Polymyalgia rheumatica, a term suggested by Barber, is a syndrome characterized by aching in the proximal portions of the extremities and torso. Because there are no specific diagnostic tests or pathologic findings, PMR is defined by its clinical features. The features included in most definitions of PMR are as follows (1) aching and morning stiffness lasting half an hour or longer in the shoulder, hip girdle, neck, or some combination; (2) duration of these symptoms for 1 month or longer; (3) age older than 50 years; and (4) laboratory evidence of systemic inflammation such as an elevated erythrocyte sedimentation rate (ESR). Some definitions also include a rapid response to small doses of glucocorticoids such as prednisone 10 mg/day. The presence of another specific disease other than GCA such as rheumatoid arthritis (RA), chronic infection, polymyositis, or malignancy excludes the diagnosis of PMR.

**Epidemiology**

The incidence of GCA varies widely in different populations, from less than 0.1 per 100,000 to 77 per 100,000 persons aged 50 years and older. The greatest risk factor for developing GCA is aging; the disease almost never occurs before age 50, and its incidence rises steadily thereafter. The average age at diagnosis has risen over the past half century, from 74.7 years in the 1950s to 79.2 years now. Nationality, geography, and race are also important, with the highest incidence figures found in Scandinavians and in Americans of Scandinavian descent. The lowest incidence of GCA is reported in Japanese, northern Indians, and African-Americans. In western Europe, GCA is more common in the northern latitudes than the southern ones. The incidence of GCA has been increasing over the past...
Gender and health status also influence the development of GCA. Women are affected about twice as often as men.\textsuperscript{10} Having diabetes reduces the risk of developing GCA by 50% in women.\textsuperscript{2} Although patients with GCA have an increased risk of developing thoracic aortic aneurysms, they do not have overall higher mortality rates.\textsuperscript{10} PMR is two to three times more common than GCA.\textsuperscript{5,8,9,19} In Olmsted County, Minnesota, 245 cases of PMR were diagnosed during the 22-year period from 1970 through 1991, providing an average annual incidence rate of 52.5 cases per 100,000 persons aged 50 years or older.\textsuperscript{19} The prevalence of PMR (active plus remitted cases) was

### Table 88-1 American College of Rheumatology Classification Criteria for Giant Cell Arteritis

<table>
<thead>
<tr>
<th>Criterion*</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at disease onset ≥ 50 yr/New headache</td>
<td>Development of symptoms or findings beginning at age 50 or older/New onset or new type of localized pain in the head</td>
</tr>
<tr>
<td>Temporal artery abnormality</td>
<td>Temporal artery tenderness to palpation or decreased pulsation, unrelated to arteriosclerosis of cervical arteries</td>
</tr>
<tr>
<td>Elevated ESR/Abnormal artery biopsy</td>
<td>ESR ≥ 50 mm/hr by the Westergren method/Biopsy specimen with artery showing vasculitis characterized by a predominance of mononuclear cell infiltration or granulomatous inflammation, usually with multinucleated giant cells</td>
</tr>
</tbody>
</table>

*For purposes of classification, a patient with vasculitis is said to have giant cell (temporal) arteritis if at least three of these five criteria are present. The presence of any three or more criteria yields a sensitivity of 93.5% and a specificity of 91.2%.

ESR, erythrocyte sedimentation rate.


### Table 88-2 Diagnostic* and Classification Criteria for Polymyalgia Rheumatica

#### Diagnostic Criteria of Chuang and Colleagues\textsuperscript{2} (1982)

- Age 50 yr or older
- Bilateral aching and stiffness for 1 mo or more and involving two of the following areas: neck or torso, shoulders or proximal regions of the arms, and hips or proximal aspects of the thighs
- ESR > 40 mm/hr
- Exclusion of all other diagnoses except giant cell arteritis

#### Diagnostic Criteria of Healey\textsuperscript{1} (1984)

- Pain persisting for at least 1 mo and involving two of the following areas: neck, shoulders, and pelvic girdle
- Morning stiffness lasting > 1 hr
- Rapid response to prednisone (≤20 mg/day)
- Absence of other diseases capable of causing the musculoskeletal symptoms
- Age older than 50 yr
- ESR > 40 mm/hr

#### Classification Criteria of Dasgupta and Colleagues\textsuperscript{14} (2012)

- Age 50 years or older, bilateral shoulder aching and abnormal C-reactive protein and/or ESR*[^14]

<table>
<thead>
<tr>
<th>Points without US (0-6)</th>
<th>Points with US\textsuperscript{*} (0-8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning stiffness duration &gt;45 min</td>
<td>2</td>
</tr>
<tr>
<td>Hip pain or limited range of motion</td>
<td>1</td>
</tr>
<tr>
<td>Absence of rheumatoid factor or anticitrullinated protein antibody</td>
<td>2</td>
</tr>
<tr>
<td>Absence of other joint involvement</td>
<td>1</td>
</tr>
<tr>
<td>At least one shoulder with subdeltoid bursitis and/or biceps tenosynovitis and/or glenohumeral synovitis (either biceps tenosynovitis and/or trochanteric bursitis and/or glenohumeral synovitis) and/or severe shoulder bursitis</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Both shoulders with subdeltoid bursitis, biceps tenosynovitis or glenohumeral synovitis</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

*For each set of criteria, all the findings must be present for polymyalgia rheumatica to be diagnosed.

\[^{14}\]A score of 4 or more is categorized as polymyalgia rheumatica in the algorithm without US, and a score of 5 or more is categorized as polymyalgia rheumatica in the algorithm with US.

\[^{*}\]Optional ultrasound criteria.

ESR, erythrocyte sedimentation rate; US, ultrasound.

approximately 600 per 100,000 persons aged 50 years and older. PMR is associated with the same HLA-DR4 genes as GCA.

**Cause, Pathology, and Pathogenesis**

The causes of GCA and PMR are unknown. Because pathologic studies have provided important clues to the pathogenesis, they are discussed first.

In GCA, inflammation is found most often in medium-size muscular arteries that originate from the arch of the aorta. The inflammation tends to affect the arteries in a segmental fashion (possibly leading to “skip lesions” within arteries), but long portions of arteries may be involved. In patients who died during the active phase of GCA, the greatest frequency of severe involvement was noted in the superficial temporal arteries, vertebral arteries, and ophthalmic and posterior ciliary arteries. The internal carotid, external carotid, and central retinal arteries were affected somewhat less frequently. In other postmortem studies, lesions were commonly found in the proximal and distal aorta and internal and external carotid, subclavian, brachial, and abdominal arteries. Because GCA affects vessels with an internal elastic lamina and vasa vasorum, and because intracranial arteries lose these structures after penetrating the dura, it is not surprising that GCA rarely involves intracranial arteries. In some patients with GCA, follow-up biopsy or autopsy surveys showed the persistence of mild chronic inflammation, even though symptoms had resolved.

Early in the disease, collections of lymphocytes are confined to the region of the internal or external elastic lamina or adventitia. The inflammation may be limited to the vasa vasorum in some cases. Intimal thickening with prominent cellular infiltration is a hallmark of more advanced cases. In heavily involved areas, all layers are affected (Figure 88-1). Transmural inflammation of portions of the arterial wall (including the elastic laminae) and granulomas containing multinucleated histiocytic and foreign body giant cells, histiocytes, lymphocytes (which are predominantly CD4+ T cells), and some plasma cells and fibroblasts are found. Eosinophils may be seen, but polymorphonuclear leukocytes are rare. Thrombosis may develop at sites of active inflammation; later, these areas may recanalize. The inflammatory process is usually most marked in the inner portion of the media adjacent to the internal elastic lamina. Fragmentation and disintegration of elastic fibers occur, closely associated with an accumulation of giant cells (Figure 88-2). However, giant cells are seen in only about half of routinely examined specimens; therefore they are not required to make the diagnosis if other features are compatible. In contrast to some other forms of systemic vasculitis (e.g., polyarteritis nodosa, microscopic polyangiitis, granulomatosis with polyangiitis [GPA] [formerly Wegener’s granulomatosis]), fibrinoid necrosis is rarely if ever observed in GCA.

Immunohistochemical studies demonstrate inflammatory changes that are specific for each layer of the affected artery. Dendritic cells, which can present antigen and activate T cells, are found in the adventitia, along with two separate lineages of CD4+ T cells: (1) Th1 cells that secrete interferon-γ (IFN-γ) and interleukin-2 (IL-2) and (2) Th17 cells that secrete interleukin-17 (IL-17) family members including IL-17A. The adventitial T cells show evidence of clonal expansion. The adventitia is also infiltrated with macrophages that secrete IL-1, IL-6, and transforming growth factor-β (TGF-β). The media is populated mostly by macrophages that, in contrast to those in other layers, produce matrix metalloproteinases and oxygen free radicals. Closer to the intima, the macrophages secrete nitric oxide and unite to form syncytia—the giant cells—which produce platelet-derived growth factor (PDGF) and substances that stimulate intimal proliferation.

Although microscopic examination of arteries in PMR is usually normal, immunohistochemical studies of apparently uninvolved temporal arteries reveal upregulation of the same macrophage-related inflammatory cytokines found in GCA. The T cell cytokine IFN-γ is abundantly expressed in GCA and is absent in arteries from patients having only PMR. Pathologically, relatively little else has been found in the arteries of patients with PMR. Granulomatous myocarditis and hepatitis have been noted. Muscle biopsy specimens may be normal or show nonspecific type II muscle atrophy. However, a number of reports have shown the presence of lymphocytic synovitis in the knees, sternoclavicular joints, and shoulders and evidence of a similar
constituents of microorganisms or some self-antigens (e.g., oxidized lipids) might also be TLR ligands that activate the arterial dendritic cells.

Once its TLRs have been engaged, dendritic cells differentiate from the resting to the active state and release cytokines such as IL-6 and IL-18, which recruit, activate, and retain CD4\(^+\) T cells in the blood vessel. The crucial role of dendritic cells in activating T cells and maintaining vasculitis has been demonstrated in the GCA-SCID mouse model: Depleting the dendritic cells markedly reduces the T cell infiltrate and suppresses the vasculitis. As noted, two separate populations of CD4\(^+\) T cells become activated and expand: IFN-\(\gamma\) producing Th1 cells and IL-17 producing Th17 cells. IFN-\(\gamma\) causes macrophages to migrate, differentiate, and form granulomas,\(^{34,40}\) Production of matrix metalloproteinases and lipid peroxidation agents by macrophages in the media results in the destruction of elastic laminae. The vessel attempts to counter the tissue destruction by elaborating a variety of growth factors including PDGF, vascular endothelial growth factor (VEGF), and TGF-\(\beta\), which prompt smooth muscle cells in the media to revert to a contractile phenotype to a secretory one and migrate to the intima. Proliferation of the intimal smooth muscle cells results in occlusion of the lumen.\(^{22,32}\)

PMR appears to result from a similar but less intense adaptive immune response in blood vessels (as evidenced by the in situ production of many inflammatory cytokines). According to this model, both PMR and GCA begin with the activation of dendritic cells at the adventitia-media border.\(^{32,41}\) However, the distinguishing feature of PMR is the absence of T cells producing IFN-\(\gamma\). Without IFN-\(\gamma\) to

reaction in sacroiliac joints.\(^{37,38}\) Synovitis (mostly subclinical) was shown in bone scans demonstrating an increased uptake of technetium pertechnetate in the joints of 2 of 25 patients with PMR.\(^{37}\) More sensitive studies using magnetic resonance imaging (MRI) and ultrasonography have convincingly demonstrated that in PMR, the principal foci of inflammation are the bursae surrounding the shoulder more than the glenohumeral joint itself.\(^{3}\) Sera from patients with GCA, PMR, or both demonstrate evidence of systemic inflammation, with increased levels of circulating immune complexes during active disease\(^{39}\) and elevated levels of IL-6 and IL-1.\(^{7}\)

These observations, together with the results from experiments in which temporal arteries from patients with GCA have been implanted into mice with severe combined immunodeficiency (SCID), suggest that GCA results from an adaptive immune response in the wall of an artery.\(^{22,33}\) In this model (Figure 88-3) the key initiating event is activation of dendritic cells located in the adventitia, the only arterial layer normally penetrated by the vasa vasorum. In large and medium-size arteries, immunohistochemical studies reveal that the dendritic cells in temporal arteries have a specific phenotype and express fascin and CD11c.\(^{22,32}\) Dendritic cells express Toll-like receptors (TLRs) that normally serve as sentinels, monitoring for any immunologic breach of the adventitia.\(^{40}\) In GCA, activation of TLRs on dendritic cells appears to be the initial triggering event.\(^{32,40,41}\) Of the many different types of TLRs, TLR-2 and TLR-4 might be the most important in GCA. Although the exact trigger of the adaptive immune response in GCA is not known, experiments in the GCA-SCID mouse model have shown that blood-borne lipopolysaccharide serves as an effective TLR ligand in temporal arteries. Other
stimulate the recruitment and differentiation of macrophages, the level of arterial inflammation in PMR remains subclinical. Thus the development of GCA appears to require both vascular dendritic cell activation and a disease-inducing repertoire of T cells. The constitutional symptoms of PMR and GCA are attributed to the high levels of inflammatory cytokines (e.g., IL-1, IL-6) found in the sera. Whether these serum cytokine elevations result from blood vessel inflammation alone or from some other source of inflammation is not yet clear. The attractiveness of this model is increased by its ability to explain why subsets of clinical features occur together.

**Clinical Features**

The mean age at onset of GCA and PMR is approximately 79 years, with a range of about 50 to 90 years of age. Women are affected about twice as often as men. Although the onset of the disease is usually insidious, typically evolving over weeks or months, in one-third of cases, the disease begins so abruptly that some patients recall the very day they became ill.

**Giant Cell Arteritis**

**Classic Manifestations.** The most common manifestations of GCA are constitutional symptoms, headache, visual symptoms, jaw claudication, and PMR (Table 88-3). Almost all patients experience one or more constitutional symptoms including fatigue, weight loss, malaise, and fever. Besides constitutional symptoms, headache is the most common symptom in GCA, being present in nearly three-quarters of patients. The pain is typically described as boring in quality, of moderate severity, and most commonly appreciated in the temporal area. However, the description of the headache varies enormously. It can be mild to so severe that the patient seeks immediate relief by presenting to the emergency department. The pain may localize to any part of the skull including the occiput (owing to involvement of the occipital artery). The most consistent characteristic is that the patient experiences the headache as something new and unusual. In untreated patients, the headache may subside over weeks, even though the disease activity continues. Often, the headache of GCA is not associated with any particular findings on physical examination. Abnormalities of the temporal artery—including enlargement, nodular swelling, tenderness, or loss of pulse—develop in only about half of patients (Figure 88-4). Some patients note tenderness of the scalp, which can be aggravated by brushing or combing the hair.

Visual symptoms are common in GCA, especially loss of vision and diplopia. Vision loss can be unilateral or (less commonly) bilateral, transient or permanent, and partial or complete. Vision loss lasting more than a few hours usually does not reverse. Loss of vision often reflects an anterior ischemic optic neuropathy caused by occlusive arteritis of the posterior ciliary artery, the chief blood supply to the head of the optic nerve. The posterior ciliary artery is a branch of the ophthalmic artery (which derives, in turn, from the internal carotid artery). Less frequently, vision loss in GCA stems from a retinal artery occlusion. Regardless of the site of the culprit lesion, vision loss in GCA is usually profound, with more than 80% of patients unable to see hand waving. GCA patients who present with fever or other systemic symptoms are less likely to develop vision loss. One possible explanation of this protective effect of fever and other systemic manifestations is that patients with prominent systemic inflammation demonstrate more extensive angiogenesis in temporal artery biopsies. The angiogenesis associated with increased inflammation may result in the development of collateral circulation that reduces the chance of ischemic events.

The early funduscopic appearance in the setting of blindness caused by anterior ischemic optic neuropathy is that of ischemic optic neuritis: slight pallor and edema of the optic

### Table 88-3 Symptoms of Giant Cell Arteritis

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headache</td>
<td>76</td>
</tr>
<tr>
<td>Weight loss</td>
<td>43</td>
</tr>
<tr>
<td>Fever</td>
<td>42</td>
</tr>
<tr>
<td>Fatigue</td>
<td>39</td>
</tr>
<tr>
<td>Any visual symptom</td>
<td>37</td>
</tr>
<tr>
<td>Anorexia</td>
<td>35</td>
</tr>
<tr>
<td>Jaw claudication</td>
<td>34</td>
</tr>
<tr>
<td>Polymyalgia rheumatica</td>
<td>34</td>
</tr>
<tr>
<td>Arthralgia</td>
<td>30</td>
</tr>
<tr>
<td>Unilateral visual loss</td>
<td>24</td>
</tr>
<tr>
<td>Bilateral visual loss</td>
<td>15</td>
</tr>
<tr>
<td>Vertigo</td>
<td>11</td>
</tr>
<tr>
<td>Diplopia</td>
<td>9</td>
</tr>
</tbody>
</table>

Modified from Smetana GW, Shmerling RH: Does this patient have temporal arteritis? JAMA 287:92, 2002. Data from a review of 2475 patients reported in the literature.
disc, with scattered cotton-wool patches and small hemorrhages (Figure 88-5). Later, optic atrophy occurs. Rarely, blindness may be the initial symptom; however, it tends to follow other symptoms by several weeks or even months. Ophthalmoscopic examination in patients without eye involvement is generally normal. In most reports, the incidence of blindness is 20% or less. In a series of 245 patients from the modern era, 34 (14%) had some permanent loss of vision. In 32 of these patients, the deficit developed before glucocorticoid therapy was begun; in the other 2, vision loss occurred after therapy was started. Vision loss progressed in 3 of the 32 after therapy was initiated, and it improved in 5. At 5 years’ follow-up, among the patients who had visual deficits caused by GCA at the time glucocorticoids were started, the risk of additional loss of vision was 13% over the follow-up period. If no loss had occurred at the beginning of glucocorticoid therapy, there was only a 1% risk of new loss of vision over the subsequent 5 years.

Another potential ocular complication of GCA is ophthalmoplegia. Diplopia usually results from ocular motor nerve palsy caused by ischemia and usually resolves after therapy is started. Oculomotor nerve involvement in GCA usually spares the pupil. Rarely, arterial lesions cause infarction of the occipital cortex and vision loss.

Intermittent claudication may occur in the muscles of mastication (jaw claudication), the extremities, and occasionally the muscles of the tongue or those involved in swallowing. In the jaw muscles, the discomfort is noted, especially when chewing meat, and may involve the muscles on one side of the mandible more than those on the other. In some instances, facial artery involvement results in spasm of the jaw muscles. More marked vascular narrowing may lead to gangrene of the scalp or tongue.

Atypical Manifestations. Approximately 40% of patients present with disease manifestations that are considered atypical (Table 88-4). In these patients, headache, jaw claudication, visual symptoms, and PMR do not occur or are less prominent.

Fever occurs in up to 40% of patients with GCA, but it is usually low grade and overshadowed by other classic symptoms. However, 15% of GCA patients may present with fever of unknown origin (FUO) in which the temperature spikes are high, dominating the clinical picture. Although GCA causes only 2% of all cases of FUO, it is responsible for 16% of all such fevers in individuals older than 65 years. Approximately two-thirds of patients experience shaking chills and drenching sweats, features often attributed to infection or malignancy. The median temperature is 39.1°C, and the maximum 39.8°C. The white blood cell count in GCA-induced FUO is usually normal or nearly so (at least before the initiation of prednisone).

Neurologic problems occur in approximately 30% of patients. These are diverse, but most common are neuropa thyies and transient ischemic attacks or strokes. Hemiparesis or brain stem events are due to narrowing or occlusion of the carotid or verteobasilar artery. GCA preferentially involves the posterior circulation; the 3:2 ratio of anterior to posterior strokes and transient ischemic attacks seen in the normal population reaches nearly 1:1 in patients with GCA. Delirium, reversible dementia, and myelopathy have also been reported. However, the assignment of an exact cause to ischemic central nervous system events is often challenging, given the older population in which GCA occurs. The neuropathies of GCA include mononeuropathies and peripheral polyneuropathies and may affect the upper or lower extremities. Presumably, they are secondary to the involvement of nutrient arteries, but little pathologic documentation is available. Among the vasculitides, GCA has a nearly unique propensity for involving the C5 nerve root, resulting in loss of shoulder abduction. Mono neuropathies affecting the hands and feet, so typical of polyarteritis and other forms of vasculitis, develop less often in GCA.

### Table 88-4  Atypical Manifestations of Giant Cell Arteritis

<table>
<thead>
<tr>
<th>Manifestation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever of unknown origin</td>
</tr>
<tr>
<td>Respiratory symptoms (especially cough)</td>
</tr>
<tr>
<td>Otolaryngeal manifestations</td>
</tr>
<tr>
<td>Glossitis</td>
</tr>
<tr>
<td>Lingual infarction</td>
</tr>
<tr>
<td>Throat pain</td>
</tr>
<tr>
<td>Hearing loss</td>
</tr>
<tr>
<td>Large artery disease</td>
</tr>
<tr>
<td>Aortic aneurysm</td>
</tr>
<tr>
<td>Aortic dissection</td>
</tr>
<tr>
<td>Limb claudication</td>
</tr>
<tr>
<td>Raynaud’s phenomenon</td>
</tr>
<tr>
<td>Neurologic manifestations</td>
</tr>
<tr>
<td>Peripheral neuropathy</td>
</tr>
<tr>
<td>Transient ischemic attack, stroke</td>
</tr>
<tr>
<td>Dementia</td>
</tr>
<tr>
<td>Delirium</td>
</tr>
<tr>
<td>Myocardial infarction</td>
</tr>
<tr>
<td>Tumor-like lesions</td>
</tr>
<tr>
<td>Breast mass</td>
</tr>
<tr>
<td>Ovarian and uterine mass</td>
</tr>
<tr>
<td>Syndrome of inappropriate antidiuretic hormone secretion</td>
</tr>
<tr>
<td>Microangiopathic hemolytic anemia</td>
</tr>
</tbody>
</table>
Prominent respiratory tract symptoms occur in about 10% of patients. These include cough with or without sputum, sore throat, and hoarseness. When these symptoms are severe or an initial manifestation of GCA, they may direct the attention of the examining physician away from the underlying arteritis. Vasculitis may induce these symptoms by causing ischemia or hyperirritability of the affected tissues. Otolaryngeal manifestations of GCA include throat pain, dental pain, tongue pain, glossitis, and ulceration or infarction of the tongue.

Clinical evidence of large artery involvement occurs in 10% to 15% of cases at presentation and in up to 27% eventually. Positron emission tomography (PET) studies using fluorodeoxyglucose (FDG) revealed that subclinical involvement of large arteries occurs in the vast majority of GCA patients. One PET study, for example, showed that 88% of 35 patients had increased FDG uptake in large arteries, with subclavian involvement in 7% and aortic involvement in 54%.

Generally, clinically evident disease can be divided into early (within a year of diagnosis) and late (years after diagnosis) stages. Usually, early disease consists chiefly of large artery stenosis resulting in upper extremity claudication; bruits over the carotid, subclavian, axillary, and brachial arteries; absent or decreased pulses in the neck or arms; and Raynaud’s phenomenon. Angiographic features that suggest GCA are smooth-walled arterial stenoses or occlusions alternating with areas of normal or increased caliber in the absence of irregular plaques and ulcerations, located especially in the carotid, subclavian, axillary, and brachial arteries. Late disease most frequently involves thoracic aortic aneurysm. The tendency for aneurysm to develop late was confirmed in one series of 41 patients in which the average time between diagnosis of GCA and recognition of this complication was 7 years. Thoracic aortic aneurysm is 17 times more likely to develop in patients with GCA than in persons without this disease. To place this risk in context, thoracic aortic aneurysms are twice as likely to complicate GCA as lung cancer is to result from smoking. Abdominal aortic aneurysm is also 2.4 times more common in patients with GCA. In aggregate, nearly one out of five patients (18%) with GCA develops an aortic aneurysm or dissection. Patients with large artery disease often do not have headache or other classic manifestations of GCA, and less than 50% have an abnormal temporal artery biopsy. Computed tomography (CT) angiography and magnetic resonance angiography (MRA) are the imaging modalities most commonly used to detect large artery disease in GCA.

In women, GCA can uncommonly present as a breast or ovarian mass. The mass lesions in these tissues result from granulomatous inflammation in and around the arteries. Angina pectoris, congestive heart failure, and myocardial infarction secondary to coronary arteritis occur rarely.

Clinical Subsets. Studies suggest that GCA is not just one disease but rather a number of clinical subsets that are explained by the differential expression of inflammatory cytokines. Ischemic events including blindness, stroke, and large artery disease occur more commonly in patients who express high levels of IFN-γ and low levels of IL-6.

In contrast, patients who produce high amounts of IL-6 are more likely to have strong inflammatory features (such as fever and constitutional symptoms) and are less likely to develop vision loss or other ischemic events.

Polymyalgia Rheumatica

As in GCA, PMR patients are characteristically in good health before their disease begins. Systemic manifestations such as malaise, low-grade fever, and weight loss are present in more than half the patients and may be the initial symptoms. High, spiking fevers are uncommon in PMR in the absence of GCA. Arthralgias and myalgias may develop abruptly or evolve insidiously over weeks or months. Malaise, fatigue, and depression, along with aching and stiffness, may be present for months before the diagnosis is made. In most patients, the shoulder girdle is the first to become symptomatic; in the remainder, the hip or neck is involved at the onset. The discomfort may begin in one shoulder or hip but usually becomes bilateral within weeks. Symptoms center on the proximal limb, axial musculature, and tendinous attachments. Morning stiffness resembling that of RA and “gelling” after inactivity are usually prominent. If the symptoms are severe, aching is more persistent. Although movement of the joints accentuates the pain, it is often felt in the proximal extremities rather than in the joints. Distal joint pain and swelling occur in some cases including diffuse distal extremity swelling with pitting edema. Pain at night is common, and movement during sleep may awaken the patient. Muscle strength is generally unimpaired, although pain with movement makes the interpretation of strength-testing maneuvers difficult. Pain with movement also makes it difficult for patients to get out of bed or the bathtub. In the later stages of the syndrome, muscle atrophy may develop, and contracture of the shoulder capsule may result in limitation of passive and active motion.

As noted, the presence of bursal inflammation and synovitis in PMR has been described by many authors and is undoubtedly the cause of many of the findings in this condition. A careful examination may reveal transient synovitis of the knees, wrists, and sternoclavicular joints.
The shoulders and hips are covered by heavy muscles, and minimal effusions of slight synovitis are not palpable on physical examination. Synovitis has been documented by biopsies, synovial analysis, joint scintiscans, ultrasonography, and MRI.\textsuperscript{16,38}

**Relationship between Polymyalgia Rheumatica and Giant Cell Arteritis**

Evidence that PMR and GCA are related and should be considered different manifestations of a common disease process is abundant.\textsuperscript{10,22} The associations with age, ethnicity, geographic region, and HLA class II alleles are the same in both disorders. Moreover, both disorders involve overproduction of many of the same inflammatory cytokines. Between 30% and 50% of patients with GCA develop PMR. Approximately 10% to 15% of patients who appear to have only PMR have positive temporal artery biopsies. In the absence of symptoms of GCA (e.g., headache, jaw claudication, visual symptoms, high fever), PMR by itself does not appear to cause vision loss and responds to low doses of glucocorticoids (see later).\textsuperscript{10}

**Laboratory Studies**

Except for the findings on arterial biopsy, laboratory results in PMR and GCA are similar (Table 88-5). A mild to moderate normochromic anemia is usually present in both diseases during their active phases. Leukocyte and differential counts are generally normal. A markedly elevated ESR and C-reactive protein (CRP) level are characteristic of both. An ESR higher than 100 mm/hr (Westergren method) is common, but untreated biopsy-proven cases of GCA may be associated with normal or nearly normal levels. In a study of 167 GCA patients, 10.8% presented with an ESR of less than 50 mm/hr and 3.6% had a rate of less than 30 mm/hr. Rare individuals appear to be unable to develop an elevated ESR during any inflammatory process including active GCA. The ESR is also likely to be relatively low or normal in patients who have been receiving corticosteroids for another condition.\textsuperscript{61} Thus a normal ESR does not exclude GCA, especially in a patient with otherwise classic symptoms and findings. Platelet counts are often increased.

Nonspecific changes in plasma proteins are often present and include a decrease in the concentration of albumin and an increase in $\alpha_2$-globulins, fibrinogen, and other acute-phase reactant proteins. Slight increases in $\gamma$-globulins and complement may be present. Results of tests for antinuclear antibodies and rheumatoid factor are generally negative.

Liver function test results are mildly abnormal in approximately one-third of patients with GCA and in a slightly smaller fraction of those with PMR. An increased alkaline phosphatase level is the most common abnormality, but increases in aspartate transaminase and prolonged prothrombin time may also be found. Liver biopsy specimens are generally normal; granulomatous hepatitis has been observed.\textsuperscript{10} Renal function and urinalysis are usually normal. Red blood cell casts are found in some instances, but their presence does not correlate with clinical large artery involvement.\textsuperscript{23}

Levels of serum creatine kinase and other enzymes reflecting muscle damage are normal. Electromyograms are usually normal, and muscle biopsy shows normal histologic features or only the mild atrophy characteristic of disease.

Synovial fluid analyses reported in GCA or PMR show evidence of mild inflammation including increased synovial fluid leukocyte counts, with a mean of 2900 cells/mm\textsuperscript{3} but a range from 300 to 20,000 cells/mm\textsuperscript{3}, with 40% to 50% being polymorphonuclear leukocytes. Synovial fluid complement levels are usually normal. In some instances, synovial biopsy has shown lymphocytic synovitis.\textsuperscript{32,38} Serum IL-6 levels are elevated in patients with PMR and GCA and appear to closely parallel the inflammatory activity.\textsuperscript{62} Levels of factor VIII or von Willebrand factor are elevated in patients with GCA and PMR.

**Differential Diagnosis**

The diagnosis of GCA should be considered in any patient older than 50 years who experiences loss of vision, diplopia, new form of headache, jaw claudication, PMR, FUO, unexplained constitutional symptoms, anemia, and a high ESR. GCA can cause so many forms of cranial discomfort (e.g., headache, scalp tenderness, jaw claudication, pain of the throat, gums, and tongue) that the disease should also be considered in any patient older than 50 who develops new, unexplained “above-the-neck” pain. The protean manifestations of GCA mean that it should also be considered in the differential diagnosis of an older patient presenting with dry cough, stroke, arm claudication, or acute C5 radiculopathy accompanied by other classic symptoms or findings of GCA.

Only a few individual symptoms or findings substantially increase or decrease the likelihood of a patient having this disease\textsuperscript{42} (Table 88-6). For example, jaw claudication, diplopia, abnormal temporal artery signs, scalp tenderness, and ESR greater than 50 mm/hr increase the likelihood that a patient has GCA.\textsuperscript{42} In one series of 373 patients, the presence of either jaw claudication or diplopia increased the likelihood of a positive biopsy by more than threefold; the presence of both jaw claudication and double vision had a 100% positive predictive value for a diagnostic temporal artery biopsy.\textsuperscript{63} Conversely, the absence of headache or temporal artery abnormalities on physical examination, the presence of synovitis, and a normal ESR reduce the likelihood of GCA.

---

**Table 88-5  Physical Findings and Laboratory Abnormalities in Giant Cell Arteritis**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any temporal artery abnormality</td>
<td>65</td>
</tr>
<tr>
<td>Prominent or enlarged temporal artery</td>
<td>47</td>
</tr>
<tr>
<td>Absent temporal artery pulse</td>
<td>45</td>
</tr>
<tr>
<td>Scalp tenderness</td>
<td>31</td>
</tr>
<tr>
<td>Any funduscopic abnormality</td>
<td>31</td>
</tr>
<tr>
<td>Abnormal ESR</td>
<td>96</td>
</tr>
<tr>
<td>ESR &gt;50 mm/hr</td>
<td>83</td>
</tr>
<tr>
<td>ESR &gt;100 mm/hr</td>
<td>39</td>
</tr>
<tr>
<td>Anemia</td>
<td>44</td>
</tr>
</tbody>
</table>

ESR, erythrocyte sedimentation rate.

Table 88-6  Likelihood Ratios* for Symptoms, Signs, and Laboratory Findings in Giant Cell Arteritis

<table>
<thead>
<tr>
<th>Finding</th>
<th>Positive Likelihood Ratio (95% CI)</th>
<th>Negative Likelihood Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptoms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jaw claudication</td>
<td>4.2(2.8-6.2)</td>
<td>0.72(0.65-0.81)</td>
</tr>
<tr>
<td>Diplopia</td>
<td>3.4(1.3-8.6)</td>
<td>0.95(0.91-0.99)</td>
</tr>
<tr>
<td>Weight loss</td>
<td>1.3(1.1-1.5)</td>
<td>0.89(0.79-1.0)</td>
</tr>
<tr>
<td>Any headache</td>
<td>1.2(1.1-1.4)</td>
<td>0.7(0.57-0.85)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Anorexia</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Arthralgia</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Polymyalgia rheumatica</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Fever</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Visual loss</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Signs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beaded temporal artery</td>
<td>4.6(1.1-18.4)</td>
<td>0.93(0.88-0.99)</td>
</tr>
<tr>
<td>Tender temporal artery</td>
<td>2.6(1.9-3.7)</td>
<td>0.82(0.74-0.92)</td>
</tr>
<tr>
<td>Any temporal artery</td>
<td>2.0(1.4-3.0)</td>
<td>0.83(0.38-0.75)</td>
</tr>
<tr>
<td>abnormality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scalp tenderness</td>
<td>1.6(1.2-2.1)</td>
<td>0.93(0.86-1.0)</td>
</tr>
<tr>
<td>Synovitis</td>
<td>0.41(0.23-0.72)</td>
<td>1.1(1.0-1.2)</td>
</tr>
<tr>
<td>Optic atrophy</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Laboratory Results</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESR abnormal</td>
<td>1.1(1.0-1.2)</td>
<td>0.2(0.08-0.51)</td>
</tr>
<tr>
<td>ESR &gt;50 mm/hr</td>
<td>1.2(1.0-1.4)</td>
<td>0.35(0.18-0.67)</td>
</tr>
<tr>
<td>ESR &gt;100 mm/hr</td>
<td>1.9(1.1-3.3)</td>
<td>0.8(0.68-0.95)</td>
</tr>
<tr>
<td>Anemia</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

*Based on literature review, with the number of patients for each variable ranging from 68 to 2475.

CI, confidence interval; ESR, erythrocyte sedimentation rate; NS, not significant.

A large number of disorders can mimic GCA (Table 88-7). There are many causes of monocular vision loss besides vasculitis including arteriosclerosis-induced thromboembolic disease. Patients with nonarteritic vision loss do not have other GCA-related symptoms, signs, or findings. The funduscopic examination may help by revealing Hollenhorst plaques in cases caused by cholesterol emboli. Anterior ischemic optic neuropathy, the most common cause of vision loss in GCA, can also be caused by arteriosclerosis. Nonarteritic optic neuritis invariably produces a small optic disc and cup-to-disc ratio, whereas GCA-related optic neuritis results in an optic disc of variable size. Thus a normal size or large cup in a patient with anterior ischemic optic neuropathy suggests GCA until proved otherwise.

Constitutional symptoms with anemia and an elevated ESR in an older person may also be produced by occult infections (e.g., tuberculosis, bacterial endocarditis, human immunodeficiency virus [HIVI]) or malignancy (especially lymphoma and multiple myeloma). These diagnoses highlight the value of selective serologic tests, imaging studies, and immunoelectrophoresis in appropriate patients. Systemic amyloidosis can closely mimic GCA, being one of the few disorders other than GCA that causes jaw claudication. The amyloid deposits in the temporal artery may not be detected unless the specimen is stained with Congo red. Polyarthritis in an older patient is much more likely caused by RA than by GCA. In one study of 520 GCA patients, less than 2% developed polyarthritis before GCA was diagnosed.

Criteria for the classification of GCA have been formulated and can help differentiate this arteritis from other forms of vasculitis (see Table 88-1). Takayasu’s arteritis, like GCA, can affect the aorta and the major arterial branches to the head and arms. Takayasu’s arteritis, however, is a disease of young women. Antineutrophil cytoplasmic antibody (ANCA)-associated granulomatous vasculitis can affect the temporal artery and, along with systemic amyloidosis, is an exception to the rule that jaw claudication is pathognomonic for GCA. AGV, however, almost always produces telltale involvement of the respiratory tract or kidneys and is associated with ANCA. Polyarteritis nodosa can also affect the temporal artery and should be considered if the biopsy does not contain giant cells and the patient has other features atypical for GCA such as mesenteric arteritis. Fibroinoid necrosis of the vasa vasorum occurs in polyarteritis but rarely, if ever, in GCA. Primary angiitis of the central nervous system differs from GCA in that it affects the intracranial arteries.

The diagnosis of PMR is clinical and depends on eliciting the symptoms and findings noted earlier. Two sets of criteria for the diagnosis have been proposed (see Table 88-2). Several disorders can mimic PMR (see Table 88-7). Distinguishing early RA from PMR can be difficult, especially in the 15% of patients who are rheumatoid factor negative and in those few RA patients who have not yet developed prominent synovitis of the small joints of the hands and feet. Patients with polymyositis complain much more of weakness than of pain—the opposite symptom pattern reported by patients with PMR. In addition, in polymyositis, levels

Table 88-7  Differential Diagnosis of Giant Cell Arteritis and Polymyalgia Rheumatica

<table>
<thead>
<tr>
<th>Disease Type</th>
<th>Specific Entities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Giant Cell Arteritis</strong></td>
<td></td>
</tr>
<tr>
<td>Occult infections</td>
<td>Tuberculosis, bacterial endocarditis, human immunodeficiency virus</td>
</tr>
<tr>
<td>Malignancy</td>
<td>Lymphoma and multiple myeloma</td>
</tr>
<tr>
<td>Systemic amyloidosis</td>
<td></td>
</tr>
<tr>
<td>Other forms of vasculitis</td>
<td>Takayasu’s arteritis, antineutrophil cytoplasmic antibody–associated granulomatous vasculitis, polyarteritis nodosa, primary angiitis of the central nervous system</td>
</tr>
<tr>
<td>Polymyalgia Rheumatica</td>
<td></td>
</tr>
<tr>
<td>Early rheumatoid arthritis</td>
<td>—</td>
</tr>
<tr>
<td>Polymyositis</td>
<td>—</td>
</tr>
<tr>
<td>Fibromyalgia</td>
<td>—</td>
</tr>
<tr>
<td>Complication of medication</td>
<td>Statins</td>
</tr>
<tr>
<td>Endocrine disorders</td>
<td>Hypothyroidism</td>
</tr>
<tr>
<td>Remitting, seronegative</td>
<td></td>
</tr>
<tr>
<td>synovitis with pitting edema</td>
<td></td>
</tr>
</tbody>
</table>
of muscle enzymes are elevated and electromyograms are abnormal. Although patients with neoplasms may have generalized musculoskeletal aching, there is no association between PMR and malignant neoplasia. Therefore a search for an underlying tumor is not necessary unless some clinical evidence for a tumor is present or the patient has an atypically poor response to low-dose prednisone.

Some patients with chronic infections such as bacterial endocarditis may have findings simulating PMR, and blood cultures should be obtained in patients with fever. Patients with fibromyalgia usually do not have typical morning stiffness and have laboratory test results that are normal or nearly so. Rarely, the stiffness of early Parkinson’s disease can be confused with PMR if the bradykinesia and tremor of Parkinson’s disease are subtle or absent. Lumbar spinal stenosis sometimes causes patients to complain of pain and stiffness in the hip girdle area. The absence of symptoms above the waist helps differentiate it from PMR. Cholesterol-lowering statin drugs can produce myalgia with or without muscle enzyme elevations but rarely mimic PMR. Hypothyroidism in the elderly can mimic many conditions including PMR. A peculiar syndrome of remitting, seronegative synovitis with pitting edema (designated RSPE syndrome) may be difficult to differentiate from PMR, and they may be related disorders. Patients with the RSPE syndrome develop acute symmetric polyarthritis of distal joints with pitting edema of the hands and feet. The RSPE syndrome and PMR both respond to nonsteroidal anti-inflammatory drugs (NSAIDs) and low-dose prednisone.10,64

Diagnostic Evaluation in Giant Cell Arteritis

Temporal artery biopsy is the “gold standard” for diagnosing GCA.10,64 Because GCA does not involve the artery in a continuous fashion, temporal artery biopsy should be directed to the symptomatic side, if evident. Removing a small (1- to 2-cm) section of temporal artery is usually adequate in patients who have palpable abnormalities of the vessel.66 Otherwise, the surgeon should try to excise a 4- to 6-cm sample, and the pathologist should examine multiple sections.60 In skilled hands, temporal artery biopsy is virtually free of morbidity or mortality. Scalp necrosis can rarely complicate active GCA but has not developed as a consequence of temporal artery biopsy.65

Temporal artery biopsies performed at institutions experienced in treating GCA are sensitive and have a high negative predictive value. At the Mayo Clinic, the sensitivity of temporal artery biopsy is approximately 90% to 95%, meaning that only 5% to 10% of patients with negative biopsies will subsequently be proved (by additional biopsy, angiography, or autopsy) to have GCA and require corticosteroid therapy.65 The sensitivity figures noted previously include some patients who underwent bilateral temporal artery biopsy. Estimates of the value of bilateral biopsies vary.64,67,68 Of 234 cases of biopsy-proven GCA, unilateral biopsy was positive in 86% and the second biopsy was positive in 14%.64 Other studies indicate that a second temporal artery biopsy improves the diagnostic yield by only 3% to 5%.68

Management of a patient with a negative unilateral biopsy depends on how strongly the patient’s clinical picture suggests GCA (Figure 88-7). When GCA is still strongly suspected, a second biopsy or an imaging test (see later) should be considered. Opting for a second temporal artery biopsy probably makes most sense in patients who have jaw claudication or diplopia. Patients with chiefly occipital headache may be best diagnosed by biopsy of the occipital artery.69 Patients who have signs of subclavian and axillary disease manifested by arm claudication, unequal arm blood pressures, and supraclavicular or axillary bruits may be diagnosed by angiogram, MRA, or CT scan.57 Typically, patients with extracranial GCA have smooth, tapered stenosis or occlusion of the subclavian, axillary, and proximal brachial arteries. In one series, temporal artery biopsy was positive in only 58% of patients with larger artery involvement.57 MRI and CT are the best-established methods for detecting aortic involvement by GCA.10

Other imaging techniques have been proposed to assist in the diagnosis of GCA. Color duplex ultrasonography showed abnormalities of the temporal artery in 28 of 30 patients with GCA (sensitivity of 93%).70 The most characteristic finding was a dark halo around the lumen of the temporal artery (Figure 88-8). However, the diagnostic value of ultrasonography is undermined by the absence of expertise with this technique at most medical centers and conflicting estimates of its sensitivity and specificity.71-73

One meta-analysis of 2036 patients in 23 studies estimated that the sensitivity and specificity of the halo sign were 69% and 82%, respectively. Another study found that ultrasonography did not improve the diagnostic accuracy of a carefully performed physical examination.74 High-resolution MRI can demonstrate contrast enhancement and mural thickening of superficial cranial arteries in GCA. Small studies have found a sensitivity of 91% and a specificity of 73% compared with temporal artery biopsy.71,72 Best results are achieved with 3-Tesla MRI scanners.7 PET has shown promise in detecting occult involvement of the aorta and great vessels by GCA; however, estimates of sensitivity have varied from 65% to 100% and specificity from 77% to 99%.71,72 PET scanning cannot image small arteries that are near highly metabolic tissues such as the brain and thus

![Figure 88-7 Algorithm for diagnosing giant cell arteritis (GCA).](image-url)
In patients with recent loss of vision. Unfortunately, the loss remains permanent in the vast majority of patients. The occlusive nature of the vasculitis argues against any role of acute thrombotic therapy in the treatment of blindness.

Because all patients with GCA require months of glucocorticoid therapy, measures to prevent osteoporosis should be started early, as outlined in Table 88-8. In addition, because traditional risk factors for atherosclerosis (e.g., smoking, hypertension, diabetes, hypercholesterolemia) might increase the risk of vision loss or stroke in GCA, reducing or eliminating these risk factors is an important part of overall management.

The initial effective dose of prednisone should be continued until all reversible symptoms, signs, and laboratory abnormalities have reverted to normal. This usually takes 2 to 4 weeks. After that, the dose can be gradually reduced by a maximum of 10% of the total dose each week or every 2 weeks. The decision to reduce prednisone should be based on a composite assessment of the patient’s symptoms, signs, and laboratory markers of inflammation. The ESR and serum concentration of CRP are generally the most convenient and helpful laboratory markers of inflammation. The ESR is reliable only if performed promptly after the blood sample is obtained. Serum levels of IL-6 appear to be the most sensitive marker of activity of GCA, but this test is not widely available. CRP may be slightly more sensitive than ESR in detecting flares. At some point during drug tapering, the ESR or CRP may rise above normal again, and further reductions of prednisone should be temporarily interrupted. If, over the next week or so, the patient does not develop signs or symptoms of active GCA, reductions in prednisone (at smaller decrements and at longer intervals) can usually be resumed. Doses of 10 to 20 mg/day or more are often required for several months before further reductions are possible. However, making the prednisone dose a slave to the levels of inflammatory markers without regard to the patient’s overall clinical context risks corticosteroid-related side effects. Gradual reductions allow the identification of the minimal suppressive dose and help avoid exacerbations resulting from too-rapid tapering. Even with a gradual reduction of prednisone, more than 50% of

### Table 88-8 Measures to Prevent Corticosteroid-Induced Osteoporosis in Giant Cell Arteritis or Polymyalgia Rheumatica

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoid or stop smoking</td>
<td>Reduce alcohol consumption if excessive</td>
</tr>
<tr>
<td>Participate in weight-bearing exercise</td>
<td>Supplement diet with calcium (1000 to 1500 mg/day)</td>
</tr>
<tr>
<td>Supplement diet with vitamin D (800 IU/day)</td>
<td>Measure BMD at lumbar spine and hip</td>
</tr>
<tr>
<td>If BMD is normal, repeat BMD annually</td>
<td>If BMD is not normal (i.e., T score below −1), prescribe bisphosphonate.</td>
</tr>
</tbody>
</table>

BMD, bone mineral density.

patients experience flares of disease activity during the first year. These exacerbations can usually be handled by increasing the prednisone 10 mg above the last dose at which the disease was controlled.

Most patients with GCA have a chronic disease that requires prednisone treatment for at least 1 year and often for several years; a minority of patients run a self-limited course of only several months. Glucocorticoids can eventually be reduced and discontinued in some patients. Many patients require low doses of prednisone for several years or more to control musculoskeletal symptoms.

The nearly universal experience of serious side effects associated with daily corticosteroids has prompted the search for alternative steroid-sparing treatments. Unfortunately, to date, none has been convincingly effective. Alternate-day prednisone, for example, is not effective initial therapy for GCA. The combination of weekly low-dose oral methotrexate and prednisone was steroid sparing in one placebo-controlled, double-blind treatment trial but not in another. These conflicting results argue against using methotrexate in combination with prednisone as initial therapy for GCA. Methotrexate may be worth adding to the treatment regimen of a patient who has experienced several exacerbations despite slow tapering of prednisone. Although open-label studies suggested that anti–tumor necrosis factor (TNF) agents might be effective in a granulomatous vasculitic process, neither infliximab nor etanercept is effective in GCA. In fact, in a treatment trial of patients with GPA, etanercept in combination with cyclophosphamide increased the number of solid tumors compared with cyclophosphamide alone. Similarly, cytotoxic drugs, dapsone, antimalarials, and cyclosporine have not been clearly shown to be effective, but they may be considered in patients who cannot achieve an acceptably low dose of prednisone.

No prospective, double-blind trials have tested the potential adjunctive role of aspirin or anticoagulants in the treatment of GCA. However, aspirin is theoretically appealing because, in experimental models of GCA, it inhibits IFN-γ production more effectively than prednisone. In addition, two retrospective studies found that GCA patients taking low-dose aspirin or anticoagulant therapy had a threefold to fivefold lower risk of developing an ischemic event such as vision loss. Together, these studies suggest that it is reasonable to add low-dose aspirin in GCA patients who do not have an excessive risk of gastrointestinal bleeding.

Arm claudication from GCA affecting the subclavian and axillary arteries usually improves or resolves with corticosteroid therapy. Rare GCA patients with severe upper extremity claudication unresponsive to corticosteroid therapy may benefit from balloon angioplasty. In one series of 10 patients, all improved initially after angioplasty but 50% developed symptomatic restenosis over 2 months. In all cases, recurrent stenosis developed in vascular lesions that were greater than 3 cm long.

Thoracic aortic aneurysms is greatly increased in patients with GCA. Although it can be present at the outset, aneurysms are usually noted late in the disease course, an average of 7 years after onset. Some authorities have recommended annual chest radiographs to detect thoracic aortic aneurysms.

No prospective, double-blind trials have tested the potential adjunctive role of aspirin or anticoagulants in the treatment of GCA. However, aspirin is theoretically appealing because, in experimental models of GCA, it inhibits IFN-γ production more effectively than prednisone. In addition, two retrospective studies found that GCA patients taking low-dose aspirin or anticoagulant therapy had a threefold to fivefold lower risk of developing an ischemic event such as vision loss. Together, these studies suggest that it is reasonable to add low-dose aspirin in GCA patients who do not have an excessive risk of gastrointestinal bleeding.

Arm claudication from GCA affecting the subclavian and axillary arteries usually improves or resolves with corticosteroid therapy. Rare GCA patients with severe upper extremity claudication unresponsive to corticosteroid therapy may benefit from balloon angioplasty. In one series of 10 patients, all improved initially after angioplasty but 50% developed symptomatic restenosis over 2 months. In all cases, recurrent stenosis developed in vascular lesions that were greater than 3 cm long.

Thoracic aortic aneurysms is greatly increased in patients with GCA. Although it can be present at the outset, aneurysms are usually noted late in the disease course, an average of 7 years after onset. Some authorities have recommended annual chest radiographs to detect thoracic aortic aneurysms.

"Treatment for Polymyalgia Rheumatica"

Patients with PMR without symptoms or signs or biopsy evidence of GCA are usually treated initially with prednisone 10 to 20 mg/day or equivalent. A systematic review of 30 studies recommended starting treatment with prednisone at 15 mg/day: Lower initial doses were less effective and higher doses were more toxic. Salicylates and NSAIDs have been used but are less appealing; salicylates and NSAIDs adequately control symptoms in only a minority of patients with milder symptoms and add to overall adverse drug reactions when they are used with glucocorticoids.

Prednisone therapy usually results in rapid (often overnight) and dramatic improvement of the musculoskeletal aching and stiffness and a more gradual return of the ESR and CRP level to normal. A minority of patients with isolated PMR fails to respond to prednisone 20 mg/day after 1 week and may require up to 30 mg/day as initial treatment. Studies suggest that these resistant cases are more likely to have ESRs greater than 50 mm/hr and high levels of IL-6. Failure to respond to prednisone 30 mg/day for 1 week should prompt a search for an alternative diagnosis (Figure 88-9). Lower doses of prednisone might not suppress an underlying arteritis if it is present. Thus the patient must be observed carefully even though the aching improves. In patients with PMR, the dose should be reduced gradually as soon as symptoms permit. Pretreatment ESR, CRP, and IL-6 concentrations and initial responses to therapy appear to be helpful in dividing patients into subsets with different treatment requirements. If the laboratory test results become normal while the patient is receiving a smaller dose, the likelihood of an underlying active vasculitis seems to be much less, and the risk of vascular complications is smaller. However, this is not true in all instances because active arteritis has been observed even though the ESR improved.

Once the symptoms, signs, and laboratory abnormalities of PMR have resolved (usually after 2 to 3 weeks of therapy), the daily dose of prednisone can be slowly tapered. Some experts recommend tapering prednisone by 2.5 mg every week until 10 mg/day is reached, at which point the decrement should be reduced by 1 mg each month. Flares are common, necessitating a dose increase to achieve remission before attempting a slower taper. The minority of patients with PMR succeed in tapering off prednisone in less than 1 year. Many require at least 2 years of low-dose prednisone.

Some, but not all, studies suggest that oral methotrexate (10 mg once a week for 48 weeks) can reduce the long-term need for corticosteroids in patients with PMR. It is not yet known whether the small but statistically significant reduction in prednisone use achieved with methotrexate results in a clinically important reduction in prednisone-related side effects.

"Takayasu’s Arteritis"

Takayasu’s arteritis (TA), also known as pulseless disease or occlusive thorboaortopathy, is a form of vasculitis of unknown cause that chiefly affects the aorta and its major branches, most frequently in young women. The disease is named for the Japanese ophthalmologist who in 1908 described a young woman with peculiar retinal arteriovenous
anastomoses caused by retinal ischemia from large vessel vasculitis.\textsuperscript{1}

**American College of Rheumatology Criteria**

The ACR classification criteria for the diagnosis of TA are listed in Table 88-9.

**Epidemiology**

Although TA has been described worldwide, it occurs most commonly in Japan, China, India, and Southeast Asia; the disease is also prevalent in Mexico. Whereas the incidence of TA in Japan is nearly 150 per million per year, it is only 0.2 to 2.6 per million in western Europe and North America. TA affects women eight times more frequently than men. The median age of onset is 25 years; however, approximately 25% of cases begin before age 20, and 10% to 20% present after age 40.\textsuperscript{89} Immunogenetic studies in Japanese patients suggest an association with several HLAs, especially HLA-Bw52, Dw12, DR2, and DQw1. Different HLA associations have been found in Koreans and Indians. No HLA association has been found in North American patients. In Mexican patients, TA has been associated with previous exposure to *Mycobacterium tuberculosis*.

**Cause and Pathogenesis**

The cause of TA is unknown. The nearly identical pathology in TA and GCA has invited speculation that the model of immunopathogenesis of GCA described earlier (see Figure 88-3) applies to TA as well.\textsuperscript{22,40} Indeed, some have argued that TA and GCA represent a spectrum within the same disease. Like GCA, TA is thought to result from an autoimmune process that targets large elastic-containing arteries. Both feature panarteritis involving infiltration of dendritic cells; T cells (including $\alpha\beta$, $\gamma\delta$, and cytotoxic); natural killer cells; and macrophages. In TA, the majority of lymphocytes are perforin-secreting killer lymphocytes such as T cells and natural killer cells. The T cell receptors in TA, as in GCA, are oligoclonal, suggesting that the vasculitis is driven in both diseases by a T cell response to a specific but unknown antigen.\textsuperscript{91} Chronic inflammation of the vessel wall leads to aneurysm formation, stenosis, or thrombosis more frequently in TA than in GCA. Dissection occurs in TA but is rare and less frequent than in syphilitic aortitis.\textsuperscript{92} The late phase of TA, like that of GCA, is characterized by intima proliferation with superimposed atherosclerosis, medial necrosis with scarring, and adventitial fibrosis. As in GCA, the inflammatory involvement in TA can be continuous or segmental, with skip areas of normal vessel interposed between involved areas.\textsuperscript{92} It is possible

<table>
<thead>
<tr>
<th>Table 88-9 American College of Rheumatology Classification Criteria for Takayasu’s Arteritis*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Onset before age 40 yr</strong></td>
</tr>
<tr>
<td>Limb claudication</td>
</tr>
<tr>
<td>Decreased brachial artery pulse</td>
</tr>
<tr>
<td>Unequal arm blood pressure (&gt;10 mm Hg)</td>
</tr>
<tr>
<td>Subclavian or aortic bruit</td>
</tr>
<tr>
<td>Angiographic evidence of narrowing or occlusion of aorta or its primary branches, or large limb arteritis</td>
</tr>
</tbody>
</table>

*The presence of three or more of the six criteria is sensitive (91%) and specific (98%) for the diagnosis of Takayasu’s arteritis.

that the humoral immune system may play some role in the pathogenesis of TA; most TA patients possess antiendothelial cell antibodies that can damage vessels by inducing endothelial inflammatory cytokine production, adhesion molecules, and apoptosis.

The geographic clustering of cases has suggested that genetics and environmental factors participate in the pathogenesis of TA. However, immunogenetic studies (described earlier) have not identified any other universally shared genetic risk factors. The young age of onset and the female predominance in TA and in systemic lupus erythematosus have invited speculation about the influence of female hormones in promoting an autoimmune disease process. In some countries, the apparent association of TA with high rates of exposure to tuberculosis has suggested an infectious cause. An animal model of TA has been produced in mice using a herpesvirus that infects the smooth muscle cells of the media. In that model, the media of large elastic arteries serves as an immunoprivileged site that allows the herpesvirus to propagate a chronic inflammatory response in the aorta and its major branches.

Clinical Features

Symptoms and Signs

Although the presenting manifestations of TA are protean, the vast majority of patients present with symptoms and signs of vascular insufficiency (from stenosis, occlusion, or aneurysm); systemic inflammation; or both.4,89 (Table 88-10). In a North American series of 60 patients followed at the National Institutes of Health, the most common presenting vascular symptoms were claudication (35%), reduced or absent pulse (25%), carotid bruit (20%), hypertension (20%), carotidynia (20%), lightheadedness (20%), and asymmetric arm blood pressures (15%).89 Stroke, aortic regurgitation, and visual abnormalities were present at onset in less than 10% of patients. The extreme manifestations of retinal ischemia noted in Takayasu’s original patient are rarely seen now.89 Permanent loss of vision, the major concern in GCA, rarely develops in TA.

Claudication affects the arms at least twice as frequently as the legs. For many young women, arm claudication first reveals itself as arm pain or fatigue experienced while trying to hold a hair dryer. Overall, bruit is the most common sign, eventually found in 80% of patients. Although bruit over the carotid artery is most frequent, it can also be found in the supraclavicular, infraclavicular, axillary, flank, chest, abdominal, and femoral areas. One-third of patients have multiple bruits.89 Unequal arm blood pressures eventually develop in half of all patients. Headache, which is common in TA, does not correlate with carotid or vertebral disease, which develops in nearly 40% of patients.89,94

Constitutional, musculoskeletal, and other symptoms of systemic inflammation are also common presenting complaints.74,89 About one in five TA patients presents with fever and malaise, which can be accompanied by night sweats and weight loss. A few patients who have minimal or no signs of vascular insufficiency may appear to have FUO for weeks or months before the diagnosis of TA becomes evident. A minority of patients present with myalgia or arthralgia (see Table 88-10). Some patients have striking midthoracic back pain, perhaps as a result of aortic inflammation irritating nociceptive nerve fibers.

Cardiac involvement occurs eventually in nearly one-third of patients (see Table 88-10).90 Aortic regurgitation develops in 20% of patients as a result of aortic root dilation. Aortic regurgitation is important because it frequently progresses and may lead to left ventricular dilation with secondary mitral regurgitation and congestive heart failure. Aortic valve replacement is often required eventually. Angina can develop as a result of coronary artery disease. TA of the coronary arteries most often produces ostial lesions but can also produce either diffuse vasculitis of the coronary arteries or aneurysms.95-97 Myocarditis also occurs in TA and causes potentially reversible congestive heart failure. Pericarditis is rare. TA is, along with Behçet’s disease, one of the few forms of vasculitis that can affect the large pulmonary arteries. Although TA of the pulmonary arteries is rare (<3%), affected patients can present with cough, chest wall pain, dyspnea, or hemothypsis.

Unlike polyarteritis or GPA, TA rarely causes peripheral neuropathies. Cutaneous manifestations develop in less than 10% of patients with TA.89 Erythema nodosum is most common, but purpura, livedo reticularis, and ulceration may rarely occur. As in GCA, a minority of TA patients with active disease have a persistent, dry cough.
Laboratory Findings

At presentation, the ESR is more frequently elevated (80%) than the CRP (=50%). Mild anemia and hypergamma-globulinemia are common. The white blood cell count is usually normal or slightly elevated. The platelet count is elevated in one-third of patients and may exceed 500,000/µL in those with active disease. The serum creatinine and urinalysis are usually normal. Any renal abnormalities are usually secondary to hypertension; unlike ANCA-associated vasculitis, TA rarely causes glomerulonephritis.

Imaging Studies

Vascular abnormalities in TA can be imaged by conventional angiography, MRI, MRA, CT angiography, or ultrasonography (Figures 88-10 to 88-12). Each imaging technique has advantages and disadvantages (Table 88-11). The earliest detectable abnormality in TA is thickening of the vessel wall from inflammation. MRI, ultrasonography, and, to a lesser degree, CT can detect this early vessel wall thickening. Conventional angiography is invasive and provides the least sensitive method for visualizing wall thickness; however, conventional angiography is the "gold standard" for precisely delineating the stenoses, occlusions,
and aneurysms that characterize the latter stages of TA. Also, only conventional angiography allows the direct measurement of central arterial blood pressure, which may be otherwise unobtainable in patients with stenotic lesions affecting all four extremities. Although MRA does not provide the same level of detail as conventional angiography, it comes close. Because MRA is not invasive and does not involve ionizing radiation, it has become the preferred imaging method for assessing the extent of vascular involvement and damage, both initially and during follow-up. The role of MRI in assessing disease activity is not clear because vessel wall edema and contrast enhancement have not correlated well with clinical measures of disease activity. Although PET scanning would be expected to be more sensitive than angiography in the early detection of vascular inflammation that precedes the development of stenoses, its sensitivity may be no greater than MRI. PET scanning has no established role in following disease activity in TA. Indeed, one study of 28 TA patients showed that PET scans did not correlate with clinical, biologic, or MRI assessment of disease activity.99

The most common sites of lesions in TA are the aorta (65%) and the left subclavian arteries (93%) (Table 88-12). The left subclavian artery is affected slightly more frequently than the right. Carotid, renal, and vertebral arteries are also commonly affected. Lesions may be stenotic (93%), occluded (37%), dilated (16%), or aneurysmal (7%). Stenotic lesions are about four times more common than aneurysmal lesions. Stenotic segments often extend a few centimeters and may be followed by areas of dilation (see Figure 88-10). The majority of patients (53%) have vascular lesions above and below the diaphragm. However, the frequency distribution of aortic lesions varies considerably from country to country.99

<table>
<thead>
<tr>
<th>Table 88-11</th>
<th>Comparison of Imaging Techniques in Takayasu’s Arteritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>Advantages</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td>Conventional angiography</td>
<td>“Gold standard” image quality</td>
</tr>
<tr>
<td></td>
<td>Allows CAP measurement</td>
</tr>
<tr>
<td></td>
<td>Allows angioplasty at same time</td>
</tr>
<tr>
<td>Magnetic resonance angiography</td>
<td>Excellent image quality</td>
</tr>
<tr>
<td></td>
<td>Noninvasive</td>
</tr>
<tr>
<td></td>
<td>No ionizing radiation exposure</td>
</tr>
<tr>
<td>Ultrasoundography</td>
<td>Visualizes vascular wall thickness</td>
</tr>
<tr>
<td></td>
<td>Noninvasive</td>
</tr>
<tr>
<td></td>
<td>No ionizing radiation exposure</td>
</tr>
<tr>
<td></td>
<td>Can visualize vessel wall edema</td>
</tr>
<tr>
<td>Computed tomography angiography</td>
<td>Excellent image quality</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Positron emission tomography</td>
<td>Can measure intensity of vascular inflammation</td>
</tr>
</tbody>
</table>

CAP, central arterial blood pressure.

Diagnosis and Diagnostic Tests

The ACR has established classification criteria for the diagnosis of TA (see Table 88-9). In clinical practice, the diagnosis of TA is almost always secured by an imaging procedure (see Table 88-11) that demonstrates the characteristic abnormalities of the aorta and its major branches (Figure 88-13). Rarely, the diagnosis is first suggested when a pathologist finds granulomatous inflammation in a section of aorta or other larger artery that was removed or biopsied during a vascular surgery procedure. Unfortunately, the diagnosis of TA is often delayed; the delay averaged 44 months in one large series. The most frequent impediment to a speedy diagnosis is a physician’s failure to consider TA in the differential diagnosis. Although the rarity of TA helps explain its omission from diagnostic consideration, another reason is that some patients have striking features of inflammation that camouflage or overshadow the somewhat more familiar vascular abnormalities. Indeed, a few patients with TA present chiefly with FUO. Most of these patients have other, albeit subtle, manifestations of TA such as bruits, diminished pulses, unequal arm blood pressures, or aortic regurgitation. In other patients with more striking vascular abnormalities, the physician might be lured into focusing on familiar and dramatic abnormalities such as anemia or thrombocytopenia. Thus instead of ordering an imaging test that would explain the patient’s unequal and low blood pressure in the left arm, the physician mistakenly

<table>
<thead>
<tr>
<th>Table 88-12</th>
<th>Frequency of Blood Vessel Involvement in Takayasu’s Arteritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood Vessel</td>
<td>% Abnormal</td>
</tr>
<tr>
<td>Aorta</td>
<td>65</td>
</tr>
<tr>
<td>Aortic arch or root</td>
<td>35</td>
</tr>
<tr>
<td>Abdominal aorta</td>
<td>47</td>
</tr>
<tr>
<td>Thoracic aorta</td>
<td>17</td>
</tr>
<tr>
<td>Subclavian artery</td>
<td>93</td>
</tr>
<tr>
<td>Common carotid artery</td>
<td>58</td>
</tr>
<tr>
<td>Renal artery</td>
<td>38</td>
</tr>
<tr>
<td>Vertebral artery</td>
<td>35</td>
</tr>
<tr>
<td>Celiac axes</td>
<td>18</td>
</tr>
<tr>
<td>Common iliac artery</td>
<td>17</td>
</tr>
<tr>
<td>Pulmonary artery</td>
<td>5</td>
</tr>
</tbody>
</table>

diverts the patient to a hematologist, gastroenterologist, or oncologist for additional blood tests and procedures that further delay the diagnosis.

Many of these delays can be prevented by remembering that TA should be included in the differential diagnosis of any person younger than 40 years who presents with FUO, aortic regurgitation, hypertension, or unequal or absent pulses. Delays in diagnosis can also be reduced by carefully searching for unequal or absent upper extremity pulses and by listening for bruits not only over the carotid arteries but also above and below the clavicle (for subclavian artery bruits) and over the abdomen and flanks (for renal and other mesenteric artery bruits). Recognizing that anemia and thrombocytosis can be manifestations of active inflammatory disorders such as vasculitis can also help speed the diagnosis of TA.

Once an imaging test demonstrates disease of the aorta or its major branches, the differential diagnosis narrows to a set of disorders that are usually easily differentiated (Table 88-13). Most rheumatic diseases that can affect the aorta are distinguished by their associated features. For example, Cogan’s syndrome typically produces ocular inflammation (especially keratitis) and vestibuloauditory dysfunction. The one rheumatic disease that can be difficult to distinguish from TA is GCA (Table 88-14). Usually, the patient’s age and the distribution of lesions allow their rapid differentiation, but distinguishing TA beginning after age 40 from GCA affecting chiefly the major branches of the aorta can be difficult or even impossible. The similarity of treatment (see later) diminishes the practical importance of solving this diagnostic dilemma.

Infections of the aorta are rare in most countries. Tertiary syphilis can be excluded by a negative fluorescent treponemal antibody test (the rapid plasma reagin test is falsely negative in about one-quarter of patients with late syphilis). Other diseases of the aorta (see Table 88-14) are usually readily separated from TA by the history and physical examination. There has been growing appreciation for a small subset of patients who have noninfectious aortitis that is difficult to categorize. At one center, noninfectious aortitis was found in 8% of patients undergoing resection of the ascending aorta. Although nearly 70% had giant cells, only a minority neatly fit a diagnosis of TA or GCA. A fraction of the cases of thoracic or abdominal aortitis are caused by a recently recognized inflammatory condition known as IgG4-related disease (IgG4-RD) associated with high levels of IgG4.

### Table 88-13 Differential Diagnosis of Takayasu’s Arteritis: Other Diseases That Can Affect the Aorta

<table>
<thead>
<tr>
<th>Disease Type</th>
<th>Specific Entities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatic</td>
<td>Giant cell arteritis, Cogan’s syndrome, relapsing polychondritis, ankylosing spondylitis, rheumatoid arthritis, systemic lupus erythematosus, Buerger’s disease, Behçet’s disease</td>
</tr>
<tr>
<td>Infectious</td>
<td>Syphilis, tuberculosis</td>
</tr>
<tr>
<td>Others</td>
<td>Atherosclerosis, ergotism, radiation-induced damage, retroperitoneal fibrosis, inflammatory bowel disease, sarcoidosis, neurofibromatosis, congenital coarctation, Marfan’s syndrome, Ehlers-Danlos syndrome, IgG4-related systemic disease</td>
</tr>
</tbody>
</table>

### Table 88-14 Comparison of Giant Cell Arteritis and Takayasu’s Arteritis

<table>
<thead>
<tr>
<th>Feature</th>
<th>Giant Cell Arteritis</th>
<th>Takayasu’s Arteritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female-male ratio</td>
<td>2:1</td>
<td>8:1</td>
</tr>
<tr>
<td>Age range (yr)</td>
<td>≥50</td>
<td>&lt;40</td>
</tr>
<tr>
<td>Average age of onset (yr)</td>
<td>72</td>
<td>25</td>
</tr>
<tr>
<td>Visual loss</td>
<td>10%-30%</td>
<td>Rare</td>
</tr>
<tr>
<td>Involvement of aorta or its major branches</td>
<td>25%</td>
<td>100%</td>
</tr>
<tr>
<td>Histopathology</td>
<td>Granulomatous arteritis</td>
<td>Granulomatous arteritis</td>
</tr>
<tr>
<td>Pulmonary artery involvement</td>
<td>No</td>
<td>Occasionally</td>
</tr>
<tr>
<td>Renal hypertension</td>
<td>Rare</td>
<td>Common</td>
</tr>
<tr>
<td>Claudication</td>
<td>Uncommon</td>
<td>Common</td>
</tr>
<tr>
<td>Ethnic groups with highest incidence</td>
<td>Scandinavians</td>
<td>Asians</td>
</tr>
<tr>
<td>Corticosteroid responsive</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bruits present</td>
<td>Minority</td>
<td>Majority</td>
</tr>
<tr>
<td>Surgical intervention needed</td>
<td>Rare</td>
<td>Common</td>
</tr>
</tbody>
</table>

### Treatment

**Medical Therapy**

Corticosteroids are the cornerstone of treatment of active TA. Prednisone, at a dose of 0.5 to 1 mg/kg per day, is indicated for the treatment of active disease. Criteria for active disease include new onset or worsening of two or more of the following: (1) fever or other systemic features (in the absence of other cause); (2) elevated ESR; (3) symptoms or signs of vascular ischemia or inflammation (e.g., claudication, absent pulse, carotidynia); and (4) typical
angiographic lesions. Although about 85% of TA patients present with active disease, about 15% do not. The initial dose of prednisone is continued for 4 to 12 weeks before commencing a gradual taper, as is done when treating GCA (see earlier). Although nearly two-thirds of patients achieve remission, more than half later relapse. Relapses are especially common as the prednisone dose falls below 20 mg per day.

Relapses can be treated by increasing the prednisone dose or adding an immunosuppressive agent. No agent used for TA has been evaluated in a double-blind, placebo-controlled trial. However, open trials have suggested that weekly oral methotrexate (started at 0.3 mg/kg per week, with the initial dose not to exceed 15 mg/wk) is a moderately effective corticosteroid-sparing drug. Methotrexate can be gradually increased to 25 mg/wk. The emphasis is on lowering the corticosteroid dose because methotrexate seldom allows the elimination of prednisone completely; most patients continue to require at least 5 to 10 mg/day of prednisone. Small open trials have provided even more encouraging results about the effectiveness of anti-TNF inhibitors (etanercept and infliximab) in treating patients with refractory TA. These trials emphasize that although these agents can treat TA, they rarely cure it: relapses are likely when the treatment is stopped. Tocilizumab, which blocks the IL-6 receptor, has also been reported effective in individual patients. Caution should be used when interpreting these results because similar open-label studies in GCA and PMR were also positive.

Small studies and series suggest that other corticosteroid-sparing drugs include azathioprine (2 mg/kg per day), mycophenolate mofetil (2000 mg/day), and cyclophosphamide (2 mg/kg per day). The toxicity of cyclophosphamide in young women is so high that it is rarely used in TA.

To prevent osteoporosis, patients on chronic corticosteroids should take calcium and vitamin D and perform weight-bearing exercises (see Table 88-8). In postmenopausal women, a bisphosphonate should be added if the bone mineral density is low (see Table 88-8). In premenopausal women, caution in use of bisphosphonates is advised: Their usefulness is less clear and recent use can harm a fetus. Modifiable risk factors for atherosclerosis—especially hypertension, smoking, inactivity, diabetes, and hyperlipidemia—should be treated maximally.

**Surgical Therapy**

TA is the form of vasculitis most frequently requiring revascularization procedures. Unfortunately, medical therapy rarely reduces or reverses stenotic lesions. Treating stenotic or aneurysmal lesions may require bypass surgery (especially of stenotic cervicobrachial arteries, coronary arteries, or renal arteries); aortic valve replacement (for aortic regurgitation); or percutaneous transluminal angioplasty (especially for stenotic renal arteries causing hypertension).

A review of the experience with vascular interventions in TA supports several general recommendations. First, the mere presence of stenosis does not necessitate intervention. The gut, for example, has such rich collaterals that even critical stenoses of the celiac, superior, or inferior mesenteric arteries usually produce no symptoms and require no surgical intervention. Moreover, many patients with arm claudication will develop collateral circulation and improve substantially over time with medical therapy alone. For upper extremity vascular insufficiency, patiently waiting for a response to medical therapy usually pays higher dividends than undertaking rapid surgical intervention. Second, whenever possible, surgical intervention should be deferred until TA is in remission; procedures done during active disease often produce disappointing results. Third, bypass surgery yields better results than angioplasty. With bypass graft procedures, autologous vessels give better results than synthetic grafts (restenosis rates of 9% vs. 36%). Patients who undergo aortic surgery are liable to develop anastomotic aneurysms; such aneurysms developed in nearly 14% of patients followed for 20 years. Although angioplasty gives good short-term results, long-term results are often disappointing except for very short stenotic segments.

The experience with conventional stents has been mostly disappointing.

**Outcome and Prognosis**

Twenty percent of TA patients have a self-limited disease. The rest have a relapsing-remitting or progressive course requiring chronic corticosteroid and/or immunosuppressive therapy. Nearly two-thirds of patients experience new angiographic lesions. In one study from the National Institutes of Health, 74% of patients experienced some form of morbidity and 47% were permanently disabled. In a Japanese study of 120 patients, survival at 15 years was 83%. Age of onset older than 35 years; development of major complications (i.e., retinopathy, hypertension, aortic regurgitation, and aneurysm); or a progressive course was associated with decreased survival. Congestive heart failure and renal failure are the most common causes of death. Pregnancy appears to be relatively well tolerated in the presence of good medical care and in the absence of abdominal aortic involvement.

**References**

Antineutrophil Cytoplasm Antibody–Associated Vasculitis

CAROLINE O.S. SAVAGE • LORRAINE HARPER

The primary systemic vasculitides involve small and medium-sized vessels and are associated with autoantibodies that target neutrophil cytoplasm antigens (antineutrophil cytoplasm antibodies [ANCA]). Thus they are frequently referred to as ANCA-associated vasculitides (AAVs), although the terminology ANCA disease has also been proposed. The presence of ANCA suggests that AAVs are autoimmune diseases. Individual disease descriptions include granulomatosis with polyangiitis (GPA), formerly known as Wegener's granulomatosis; microscopic polyangiitis (MPA); allergic granulomatosis with polyangiitis (AGPA), formerly known as Churg-Strauss syndrome; and renal-limited pauci-immune necrotizing and crescentic glomerulonephritis (RLV). The name changes have been recommended by the Boards of the ACR, the American Society of Nephrology, and EULAR, which wished a shift from eponyms to disease-descriptive or cause-based nomenclature.

The major autoantigens to which ANCs are directed within neutrophils and monocytes include two enzyme proteins identified in the 1980s as autoantigens, namely, proteinase-3 (PR3), myeloperoxidase (MPO) (Figure 89-1), and, more recently, lysosome-associated membrane protein 2 (LAMP2). Confirmation of the presence of antibodies directed against LAMP2 in AAV is awaited, so this antibody subclass is not routinely tested for in current clinical practice.

The initial descriptions of GPA, MPA, and AGPA occurred in the 1930s, 1940s, and 1950s, respectively. However, the term polyarteritis nodosa (PAN) was often used during this period in a generic manner to cover these conditions, particularly MPA, despite the fact that Kussmaul and Meyer had clearly described PAN in 1866 as a systemic condition in which inflammation of medium-sized muscular arteries occurs, leading to aneurysm formation and ischemic tissue or organ infarction. The AAVs, in contrast, are generally considered disorders affecting both small and medium-sized arteries. In 1990, the ACR published classification criteria for seven types of vasculitis that included GPA and AGPA, but not MPA, thereby helping to begin to unravel the complexities of vasculitis classification. Four criteria were selected for GPA:

- Abnormal urinary sediment (red cell casts or greater than five red blood cells per high-power field)
- Abnormal findings on chest radiograph (nodules, cavities, or fixed infiltrates)
- Oral ulcers or nasal discharge
- Granulomatous inflammation on biopsy

The presence of two or more of these four criteria was associated with a sensitivity of 88.2% and a specificity of 92%. It is worth bearing in mind that these criteria were
based on clinical features of 85 patients with GPA who were compared with 722 control patients with other forms of vasculitis.

Six criteria were selected for AGPA:
1. Asthma
2. Eosinophilia greater than 10% on differential blood cell count
3. Mononeuropathy (including multiplex) or polyneuropathy
4. Nonfixed pulmonary infiltrates on chest radiograph
5. Paranasal sinus abnormality
6. Biopsy containing a blood vessel with extravascular eosinophils

The presence of four or more of these six criteria gave a sensitivity of 85% and a specificity of 99.7%. The criteria for AGPA were based on clinical findings in 20 patients compared with 787 controls with other forms of vasculitis.

The next major step forward was development of definitions for forms of vasculitis by the Chapel Hill Consensus Conference (CHCC) in 1994, which included GPA, MPA, and AGPA in the 10 types of vasculitis considered. These are shown in totality in Table 89-1. Neither the ACR classification criteria nor the CHCC definitions included ANCA as a criterion. Lack of classification criteria for MPA prompted development of an algorithm by international consensus that has been validated in two separate populations, covering GPA, MPA, and AGPA, as well as PAN.

The ACR classification criteria and the CHCC definitions were not designed as diagnostic criteria. Indeed, the positive predictive value of the ACR classification criteria can be as low as 29% in clinical practice. Although attempts have been made to develop diagnostic criteria based on the CHCC, these Sorensen criteria have not been validated. Indeed, believing the ACR criteria for AGPA, PAN, and some other vasculitides, and the CHCC definitions for GPA, MPA, and PAN to no longer be fit for purpose, EULAR convened an expert consensus group to consider re-evaluating definitions, classifications, and diagnostic criteria in systemic vasculitis. Seventeen points to consider were formulated before the development of classification criteria and definitions in the systemic vasculitides that related to biopsy, laboratory testing, diagnostic radiology, nosology, definitions, and the research agenda. Not all of the points were relevant to AAV. Those that were relevant included the following:

- Histology point 1: Although histology is fundamental to the diagnosis of vasculitis and exclusion of its mimics, biopsy of affected organs is not always possible, and yields vary significantly according to conditions and target organs.
- Laboratory testing point 4: ANCA testing plays an important diagnostic role in suspected small vessel vasculitis.
- Diagnostic radiology point 10: Computed tomography (CT) and magnetic resonance imaging (MRI) may be useful in diagnosing ear, nose, and throat (ENT) involvement associated with GPA/AGPA.
- Nosology points 12 to 14: The nomenclature in use for distinguishing between “disease definitions,” “classification,” and “diagnostic” criteria is confusing and should be clarified wherever possible. Nosology of different forms of vasculitis should be a reflection of their etiopathogenesis, wherever this has been determined. In the absence of this, definition must rely on a clear, accurate description of the salient features of the condition. The use of eponyms should be reviewed if a more rational approach to nomenclature can be developed, based on etiopathogenesis, but their retention is necessary at present to avoid confusion.
- Definitions point 15: Age is worthy of inclusion in the definitions of some forms of vasculitis, but its role should not be overstated.
- Research agenda points 16 and 17: Future criteria initiatives should include all forms of vasculitis, providing definitions of less common syndromes not covered by CHCC. The development of a classification tree will provide the foundation for future criteria.

**EPIDEMIOLOGY**

Determining the incidence and prevalence of AAV is challenging given the uncommon occurrence of the disorders, difficulties in case ascertainment, the slow evolution of classification criteria and definitions fit for epidemiologic purposes, and the clinicopathologic overlaps that occur between the component diseases designated ANCA-associated diseases (GPA, MPA, AGPA) and their limited variants, including RLV. Most studies have been carried out in populations of European ancestry using data that were collected...
**Table 89-1  Classification of Vasculitis as Adopted by the Chapel Hill Consensus Conference on the Nomenclature of Systemic Vasculitis**

<table>
<thead>
<tr>
<th>Classification of Vasculitis</th>
<th>Histopathology</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large Vessel Vasculitis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giant cell (temporal) arteritis</td>
<td>Granulomatous arteritis of the aorta and its major branches, with a predilection for the extracranial branches of the carotid artery</td>
<td>Often involves the temporal artery Usually occurs at &gt;50 yr of age Often associated with polymyalgia rheumatica</td>
</tr>
<tr>
<td>Takayasu’s arteritis</td>
<td>Granulomatous inflammation of the aorta and its major branches</td>
<td>Usually occurs at &lt;50 yr of age</td>
</tr>
<tr>
<td>Medium-Sized Vessel Vasculitis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polyarteritis nodosa (classic polyarteritis nodosa)</td>
<td>Necrotizing inflammation of medium-sized or small arteries without glomerulonephritis or vasculitis in arterioles, capillaries, or venules</td>
<td>Coronary arteries often involved Aorta and veins may be involved Usually occurs in children</td>
</tr>
<tr>
<td>Kawasaki disease</td>
<td>Arteritis involving large, medium-sized, and small arteries, associated with mucocutaneous lymph node syndrome</td>
<td></td>
</tr>
<tr>
<td>Small Vessel Vasculitis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Granulomatosis with polyangiitis*</td>
<td>Granulomatous inflammation involving the respiratory tract and necrotizing vasculitis affecting small to medium-sized vessels (capillaries, venules, arterioles, arteries)</td>
<td>Necrotizing glomerulonephritis common</td>
</tr>
<tr>
<td>Allergic granulomatosis with polyangiitis*</td>
<td>Eosinophil-rich and granulomatous inflammation involving the respiratory tract and necrotizing vasculitis affecting small to medium-sized vessels, associated with asthma and blood eosinophilia</td>
<td></td>
</tr>
<tr>
<td>Microscopic polyangiitis†</td>
<td>Necrotizing vasculitis with or no immune deposits affecting small vessels (capillaries, venules, arterioles)</td>
<td>Necrotizing arteritis involving small and medium-sized arteries may be present Necrotizing glomerulonephritis very common</td>
</tr>
<tr>
<td>Henoch-Schönlein purpura</td>
<td>Vasculitis with IgA-dominant immune deposits affecting small vessels (capillaries, venules, arterioles)</td>
<td>Pulmonary capillaritis often occurs Typically involves skin, gut, and glomeruli Associated with arthralgia or arthritis</td>
</tr>
<tr>
<td>Essential cryoglobulinemic vasculitis</td>
<td>Vasculitis with cryoglobulin immune deposits affecting small vessels (capillaries, venules, arterioles) and associated with cryoglobulins in serum</td>
<td>Skin and glomeruli often involved</td>
</tr>
<tr>
<td>Cutaneous leukocytoclastic angitis</td>
<td>Isolated cutaneous leukocytoclastic angitis without systemic vasculitis</td>
<td></td>
</tr>
</tbody>
</table>

*Large arteries include the aorta and the largest branches directed toward major body regions (e.g., the extremities, head and neck). Medium-sized arteries are the main visceral arteries (e.g., renal, hepatic, coronary, mesenteric). Small arteries are distal arterial radicles that connect with arterioles. Note that some small and large vessel vasculitides may involve medium-sized arteries, but large and medium-sized vessel vasculitides do not involve vessels smaller than arteries.

†Preferred term.

‡Strongly associated with antineutrophil cytoplasm antibodies.

retrospectively. Despite these difficulties, collective studies suggest that the AAVs have an incidence of around 10 to 20 per million and prevalence rates that have been increasing over the past two decades (summarized in Reference 14). The peak age of onset is 65 to 74 years; the disease is very rare in childhood, and most studies suggest a slightly higher occurrence in men than in women (1.5:1.0).

Within the spectrum of ANCA-associated disease are interesting geographic and population differences in the relative incidence of GPA versus MPA or AGPA, or between MPO-ANCA and PR3-ANCA positivity. In populations of European ancestry, GPA appears to have an incidence of around 2 to 10 per million, depending on the geographic location, with the higher 8 to 10 per million incidence being reported in more northerly countries and lower incidences of 3 to 6.6 per million in Greece and Spain. A similar inverse relationship between GPA and MPA has been observed in the southern hemisphere, and a possible link to ultraviolet exposure has been proposed. However, in a Japanese population studied for occurrence of primary renal vasculitis, more than 90% had MPO-ANCA, but PR3-ANCA was not observed, and neither GPA nor AGPA was diagnosed clinically. In China, MPA also seems to be more common than GPA; high rates of MPA have also been reported in a Peruvian population and in Kuwait, where the incidence of MPA was reported as 24 per million.

Considering epidemiologic studies for GPA in greater detail, studies from Finland, Norway, and Sweden have suggested an increased incidence over the past two to three decades, but others from Germany and the United Kingdom and a later Swedish study have not. Overall, it is most likely that methodological factors account for reported differences, with no significant increase in incidence occurring during the past two decades. Prevalence figures for GPA have been increasing, which probably reflects improved treatment regimens. In a primary care–based population in the United Kingdom, the prevalence
increased from 28.8 per million in 1990 to 64.8 per million in 2005. Prevalence figures are now available for several populations in Europe, the United States, and the southern hemisphere over various time periods (summarized in Reference 14).

Epidemiologic studies in MPA have variously shown an increased incidence or not over the past two decades. It is possible that earlier suggestions of an increased incidence reflected increasing awareness of MPA and its differentiation from PAN and GPA. An interesting increase in cases of MPA in Japan followed the Kobe earthquake in 1995. AGPA is the least common of the AAVs, with an incidence around 1.0 to 3.0 per million, although this increases in patients with asthma to 34.6 per million person-years. AGPA affects a similar age of population as GPA and MPA, but its occurrence is more common in women than in men. An association has been noted between the development of AGPA and the administration of leukotriene inhibitors or the anti-immunoglobulin (Ig)E monoclonal antibody omalizumab, possibly due to unmasking of previously unrecognized disease with reduction of corticosteroid dose.

Environmental factors are relevant to AAV. There appears to be a higher incidence of GPA in rural as compared with urban areas. No clear seasonal variation trends have been identified. Infection as a trigger for autoimmune disease in general, and for GPA in particular, is often hypothesized. The closest association is between Staphylococcus aureus and GPA, with nasal carriage of S. aureus being linked to a higher incidence of disease relapse. Silica exposure has been linked to AAV in several studies, including a case-control study in the United States. Exposure to several drugs, including propylthiouracil, has also been associated with AAV. Cocaine abuse has been linked to a midline destructive granulomatous disease that mimics GPA.

**GENETICS**

Evidence of a heritable risk for AAV comes from recent studies suggesting a modestly increased risk of disease in first-degree relatives, similar to that found in rheumatoid arthritis (RA). Children of patients with GPA have an increased risk of RA, suggesting familial clustering of genes associated with inflammatory autoimmune disease. Occasional familial cases of GPA are also reported in the literature. Given the relatively modest risk of disease in family members, it is likely that several genes contribute a small effect on disease development. Undertaking genetic studies in AAV is challenging owing to the rarity of the disease as the statistical power of genetic association studies is determined by the number of cases and controls included in the analysis. A number of candidate genes, often involving the immune response, have been identified. However, many association studies have produced inconsistent results because of the small number of cases studied. More recent studies have used larger numbers of patients with evidence of validation of results by duplication in separate cohorts of patients. Large cohorts of patients are being assembled to allow performance of genome-wide association studies, in which thousands of genes across the genome are compared.

Consistent genetic associations with multiple autoimmune conditions have been restricted to three gene regions: human leukocyte antigen (HLA) class II region, CTLA, and the PTPN22 gene. HLA genes are exceedingly polymorphic, and small studies suggest associations between GPA and HLA compared with healthy controls. However, most of these studies are of poor reliability because of their small size and lack of replication in independent cohorts. Recently, a larger study of 150 German patients with GPA identified an association with the HLA-DPB1*0401 allele. In contrast, the *0301 allele frequency was significantly decreased. The extended haplotype DPB1*0401/RXBO3 showed stronger association, suggesting that this genomic region confers significant risk for development of GPA. Among other functions, the retinoid X receptor β protein forms heterodimers with vitamin D receptors. Vitamin D receptor is important for the effects of the active vitamin D metabolite 1,25-dihydroxycholecalciferol. Active vitamin D has potent immunomodulatory properties, which include inhibition of cytokine transcription and differentiation of T cells toward a T regulatory (Treg) cell phenotype. This association has been replicated in a larger cohort of 282 GPA patients. The association with *0401 was present only in ANCA-positive GPA patients. Other studies have shown associations with HLA-DR. A relatively large study (304 AAV patients and 9872 controls) has shown associations with HLA-DRB4, and another showed that HLA-DRB1*04 was increased in frequency in GPA patients with end-stage renal failure.

Different HLA associations have been found in AGPA with disease associating with HLA-DRB4 alleles, while HLA-DRB3 afforded disease protection. No association of AGPA with HLA-DPB has been described. Different HLA associations suggest that AGPA and GPA may be different disease entities despite many similarities. This is supported by observations that showed association of the extended IL-10-1A haplotype with ANCA-negative AGPA but not GPA.

Cytotoxic T lymphocyte–associated antigen 4 (CTLA4) is expressed mainly on activated CD4+ T cells and inhibits T cell function. Polymorphisms in the CTLA4 gene have been associated with several autoimmune diseases. Several studies have shown associations with polymorphisms in the CTLA4 gene. The most widely implicated CTLA4 polymorphism for risk of autoimmunity is the +49 single nucleotide polymorphism (SNP) (alanine-to-threonine substitution in the leader peptide), which appears to affect cell surface expression of CTLA4 in response to T cell activation. The CT60 SNP appears to affect the expression of soluble CTLA4 and to alter the signaling threshold of CD4+ T cells. Both of these SNPs have been associated with disease in patients with AAV. PTPN22 encodes the lymphoid tyrosine phosphatase LYP, which forms a complex with the kinase Csk and is a critical negative regulator of signaling through the T cell receptor. The R620W variation, which has been associated with autoimmunity, disrupts the interaction between Lck and LYP, leading to reduced phosphorylation of LYP, which ultimately contributes to gain-of-function inhibition of T cell signaling. Two studies have shown an association of the PTPN22 620W allele with AAV, suggesting that this
is likely to contribute to the risk of AAV, as in other autoantibody-associated autoimmune diseases.

Numerous other candidate genes have been investigated in patients with AAV. Alphai antitrypsin (A1AT) is the major inhibitor of PR3. The A1AT gene is highly polymorphic. The deficiency alleles S and Z are increased in patients with AAV.60-63 However, most individuals with A1AT homozygous for the Z allele do not develop AAV.64

PR3 is expressed on the membrane of neutrophils and appears to be genetically determined.65 Patients with GPA tend to have higher percentages of neutrophils expressing PR3 than healthy controls,66 particularly those with relapsing disease.67 However, increased expression does not appear related to 564 promoter polymorphism of the PR3 gene.68,69 The percentage of neutrophils expressing PR3 may be influenced by HLA antigens, although the mechanism of this is unclear. In one study, a group of 34 HLA antigens was found to predict 64% of the variability in PR3 membrane expression.70 This study requires replication in an independent cohort of patients. Less support has been found for a genetic association affecting the expression of MPO. A recent meta-analysis showed no association with the functional promoter polymorphism (G-463A) in the MPO gene.71

Many other genes have been investigated, but results are conflicting owing to the small size of the studies and require further investigation. However, it is clear that existing studies suggest a complex model of genetic variability. Future studies need collaboration between groups to increase the number of available cases of these rare diseases.

**CLINICAL FEATURES**

The three types of AAV may all be associated with marked constitutional upset comprising fevers, night sweats, myalgia, and arthralgia. Systemic vasculitis may also lead to certain commonality of features that may occur in some, but not all, patients. Thus nail-fold hemorrhages, purpuric rashes, and cutaneous ulcers may be observed.

**Microscopic Polyangiitis**

The first descriptions of MPA were provided in 1948 by Davson, Ball, and Platt, who described an illness that they termed *microscopic polyarteritis*, distinguishing it from polyarteritis nodosa by its marked segmental necrotizing glomerulonephritis and greater propensity to involve small vessels.72

The pathology of MPA is of a fibrinoid necrotizing vasculitis with few or no immune deposits that primarily affects small vessels such as capillaries, arterioles, or venules, although spread to include small and medium-sized arteries may occur. A focal segmental necrotizing glomerulonephritis is very common and may progress into a full blown crescentic glomerulonephritis.73 However, in view of the sometimes indolent nature of the disease, evidence of chronic damage with obsolete glomeruli and tubulointerstitial fibrosis may be present at the time of the first diagnostic kidney biopsy.74 Examination of the kidney by immunohistology or electron microscopy reveals few or no immune deposits, although careful inspection may indicate that some complement fragments are present.75 Within the lungs, a pulmonary capillaritis may develop, and, following rupture of capillaries, blood can spill into alveolar spaces and thrombosis may occur within capillaries themselves. Marked neutrophilic infiltration of the alveolar wall is usually seen; this ultimately undergoes fibrinoid necrosis. Type II epithelial cell hyperplasia and lymphoplasmacytic infiltration can develop.

The pathophysiologic understanding of MPA advanced considerably after the strong association between MPA and ANCA was recognized, particularly ANCA directed toward myeloperoxidase (MPO).2,76 A small percentage of patients have ANCA directed toward proteinase-3 (PR3). ANCA’s are believed to bind to their target antigen on the surface of primed neutrophils, leading to further neutrophil activation with release of proinflammatory granule contents and reactive oxygen species, as well as increasing adherence and damage to endothelial cells (Figure 89-2) (reviewed in Reference 77). Studies in mouse and rat species have provided direct evidence that anti-MPO antibodies induce systemic vasculitis, including necrotizing glomerulonephritis,76,79 and can promote neutrophil-endothelial cell interactions in vivo.80,81

The clinical features of MPA have been recognized for many years82,83 and are summarized in Table 89-2. The most commonly affected organs are the kidneys and the lungs. Symptoms can involve the ears, nose, or throat, but distinction from GPA then needs to be considered. The presentation can be insidious with symptoms of deteriorating renal function on the background of mild features that can be attributed to a low-grade vasculitis; alternatively, it may be severe and acute with rapidly progressive glomerulonephritis and pulmonary hemorrhage, presenting as a pulmonary renal syndrome; or it may appear to be limited to the kidney, sometimes being described as renal-limited vasculitis.

Renal manifestations include microscopic hematuria, an abnormal renal sediment with red cell casts, proteinuria that is not usually of nephrotic proportions (i.e., is less than 3.5 g per 24 hours), and variable loss of kidney function. The presence of red cell casts is always indicative of an active glomerulonephritis, and the presence of red cells in a patient with long-standing disease may be due to active disease or to damage in the absence of disease activity, or indeed to other conditions affecting the lower urinary tract, including cystitis or bladder malignancy, as a result of cyclophosphamide therapy.86 Loss of renal function can be very rapid, occurring over days or weeks. Dialysis may be required,

<table>
<thead>
<tr>
<th>Clinical Feature</th>
<th>Percentage*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constitutional symptoms</td>
<td>76-79</td>
</tr>
<tr>
<td>Fever</td>
<td>50-72</td>
</tr>
<tr>
<td>Renal disease</td>
<td>100</td>
</tr>
<tr>
<td>Arthralgias</td>
<td>28-65</td>
</tr>
<tr>
<td>Purpura</td>
<td>40-44</td>
</tr>
<tr>
<td>Pulmonary disease (hemorrhage, infiltrates, effusion)</td>
<td>50</td>
</tr>
<tr>
<td>Neurologic disease (central, peripheral)</td>
<td>28</td>
</tr>
<tr>
<td>Ear, nose, throat involvement</td>
<td>30</td>
</tr>
</tbody>
</table>

*Percentage of a population totaling 150 patients from four studies.82,85
but some recovery of renal function may occur following treatment. In patients who remain on dialysis, the likelihood of relapse is less than before dialysis, so the need for immunosuppression may be less. Although kidney involvement is present in most patients, it is not invariable.

Lung involvement occurs in up to a third of patients. Clinical features include cough, dyspnea, pleurisy, and hemoptysis. Onset may be insidious or acute and severe. Widespread pulmonary hemorrhage can occur without overt hemoptysis. Chest radiograph findings may be patchy or diffuse, reflecting alveolar infiltrates. Repeated episodes of lung hemorrhage can lead to pulmonary fibrosis.

**Granulomatosis with Polyangiitis**

GPA is a granulomatous disorder, often associated with fibrinoid necrotizing vasculitis, which was first described in 1931 by Klinger, with pathologic refinement added to the description by Wegener in 1936. The pathophysiology of GPA, as with MPA, is believed to be inherently autoimmune and driven by PR3-ANCA in a manner very similar to that proposed for MPO-ANCA, with PR3-ANCA being highly specific for GPA. However, a good animal model is not available for anti-PR3 antibodies, so direct evidence for a role in development of the disease is not available as with anti-MPO antibodies; this may reflect differences in expression of PR3 on murine neutrophils. However, the ex vivo effects of MPO-ANCA and PR3-ANCA on human neutrophils are very similar (reviewed in Reference 77). B cells and T cells are also acknowledged to play important roles in the AAV diseases, with recent demonstration of responsiveness to anti-B cell therapies placing increased interest in the role that B cells play. T cell subset abnormalities have repeatedly been described in MPA and GPA, but the nature of the factors responsible for these changes has not been defined.

The clinical features of GPA are driven by a predilection for the upper and lower respiratory tracts, as well as for the kidneys. GPA may occur as a disease limited to the respiratory tract without evidence of systemic involvement, when

---

**Figure 89-2**  Schematic representation of the immune mechanisms hypothetically involved with antineutrophil cytoplasm antibody (ANCA) enhancement of vascular injury. An infectious trigger or another environmental stimulus leads to a burst of cytokines, which prime the neutrophils or monocytes, possibly leading to local upregulation of adhesion molecules on endothelium. The priming process within the inflammatory cells leads to enhanced expression of ANCA antigens on the cell surface. Activated neutrophils or monocytes may degranulate and release reactive oxygen species (O) and lysosomal enzymes, leading to endothelial injury and further activation of the endothelial cell surface. The magnitude of this effect is influenced by the specificity of ANCA for proteinase-3 (PR3) or myeloperoxidase (MPO), as well as different epitopes of these respective antigens. The reaction may be further influenced by the immunoglobulin G (IgG) and Fcγ receptor phenotype engaged. Products released from degranulated inflammatory cells become bound to endothelial cells and further serve as targets of ANCA. Release of chemotactic chemokines such as interleukin-8 (IL-8) and macrophage chemoattractant protein-1 (MCP-1), in conjunction with other adhesion molecules, serves to augment chemotaxis and inflammatory cell transmigration. Thus, the scheme provides the prerequisites for endothelial and vascular injury induced by ANCA, that is, the presence of ANCA, expression of target antigens for ANCA on primed neutrophils and monocytes, interaction between primed neutrophils and endothelium via adhesion molecules, and finally, activation of endothelial cells and ultimate efflux of inflammatory cells to extravascular and perivascular tissues. FcγR, Fcγ receptor; ICAM-1, intercellular adhesion molecule-1; PMN, polymorphonuclear leukocyte; TNF, tumor necrosis factor.
it is referred to as limited GPA; such presentation is usually as a granulomatous disorder without vasculitic features. In a recent study, 10% of patients evolved to develop generalized disease within a median time of 6 years.97 Upper airway disease is the most common presenting feature of GPA, occurring in more than 70% and eventually being present in more than 90% of patients.93,96 Serous otitis media and conductive and sensorineural hearing loss may occur; vertigo is rare. Nasal involvement causes mucosal swelling with obstruction, crusting, septal perforations, serosanguineous discharge, and epistaxis; a saddle nose deformity due to collapse of the cartilaginous portion of the nasal septum may develop (Figure 89-4). Sinusitis is common and will occur in more than 80% at some point during the illness (Figure 89-5); bony erosion may develop and can be detected better using CT scanning of the sinuses than by plain radiography. S. aureus may infect sinuses and is commonly carried on the diseased nasal mucosa, where it may be one factor contributing to disease relapse.93 Laryngotraheal disease may cause hoarseness but may also progress to severe stridor and upper airway obstruction, usually following subglottic stenosis (Figure 89-6). Direct laryngoscopy may reveal an ulcerated friable mucosa, and tracheal tomograms, CT, or MRI can help to further delineate the extent of the stenosis. In occasional patients, subglottic stenosis can precede the development of other features of the disease by years; in others, subglottic stenosis may develop despite apparently effective control of the disease.

Pulmonary involvement will affect around 90% of patients at some point during the course of disease.90 Symptoms are similar to those associated with MPA, although some patients may have asymptomatic disease that is detected only after imaging. In addition to the capillaritis (Figure 89-7) and vasculitis that occur in MPA, patients with GPA are burdened with granulomatous disease that causes development of nodules of chronic granulation tissue.
allow the test to be undertaken. Depending on the progress of the disease, pulmonary function tests may show an obstructive or a restrictive component, particularly if fibrosis has developed following repeated episodes of hemorrhage or cyclophosphamide-induced pneumonitis. The features of renal involvement in systemic GPA are similar clinically and pathologically to those in MPA. Renal involvement probably occurs in around 80% of patients at some time during the course of the illness, although determining the precise frequency overall is difficult. However, it should never be assumed that limited disease excludes the future development of renal and systemic disease because, as already noted, a proportion of patients will develop systemic disease at some point. The lower urinary tract may also be affected with necrotizing vasculitis of the bladder, necrotizing urethritis, orchitis, epididymitis, prostatitis, and penile necrosis (reviewed in Reference 103). Urinary obstruction may develop, particularly with involvement of the ureters. Cystoscopy should be undertaken if unexplained persistent hematuria is present, to exclude bladder malignancy or other complications.

The eye may be affected in several ways during the course of GPA in 28% to 58% of patients. Keratitis, conjunctivitis, episcleritis, scleritis, uveitis, retrobulbar granulomatous disease with proptosis, ocular palsies, lacrimal duct obstruction, optic neuritis, and retinal vascular occlusion may all occur. Proptosis and optic neuritis are feared because they are particularly likely to lead to visual loss. Both CT and MRI may be helpful in defining retrobulbar disease (Figure 89-9). In patients treated with high doses of corticosteroids, cataracts may occur as a complication.

Both the peripheral and the central neurologic system can be affected in GPA. Peripheral neuropathy may manifest as a mononeuritis multiplex or, less commonly, as a distal and symmetric polyneuropathy. Nerve conduction

Figure 89-7  Capillaritis. Alveolar septa with congestion, neutrophilic infiltrate, and fibrinoid necrosis of capillary walls. Adjacent alveolar spaces contain hemosiderin-laden macrophages, consistent with a history of pulmonary hemorrhage (hematoxylin and eosin, ×40). (Courtesy Dr. C. Farver.)

Figure 89-8  Computed tomography scan of the lungs, revealing the presence of nodules. The right-sided pulmonary nodule is cavitary.
Increased risk of cardiovascular disease is noted in patients with AAV, including GPA.\textsuperscript{110,111} Symptomatic involvement of the gastrointestinal tract is not usually a major feature of GPA, or indeed of MPA. However, abdominal pain, bleeding, and diarrhea can occur as a result of the disease itself or through the effects of treatment, for example, corticosteroids can cause peptic ulceration and mycophenolate mofetil may cause diarrhea. The vasculitic process itself may lead to ulcerations, or even perforations, in the small or large intestine. A number of unusual presentations include involvement of the tongue and salivary glands (parotic, sublingual, or submandibular),\textsuperscript{112} pancreatic involvement that may mimic pancreatic cancer,\textsuperscript{113} and cholecystitis and hepatic granulomatous disease that can cause liver failure.\textsuperscript{114} Splenic involvement with vasculitis, granulomas, and necrosis is present in many patients in older autopsy series.\textsuperscript{115}

Allergic Granulomatosis with Polyangiitis

This syndrome was described by Churg and Strauss in 1951.\textsuperscript{116} It has three salient histopathologic features, namely, necrotizing vasculitis, tissue infiltration by eosinophils, and extravascular granulomas (Figure 89-11). To improve recognition of this disorder, Lanham suggested that diagnosis be based on clinical features comprising asthma, peak eosinophil count greater than 1500 cells/mL, and systemic vasculitis involving two or more organs.\textsuperscript{117,118} None of these clinical or pathologic features is entirely specific for AGPA, and they may not all be present or detectable concurrently.

The pathogenesis of AGPA is unknown, but a strong association has been noted with allergy and atopic disorders, including allergic rhinitis, nasal polyposis, and asthma. Around 70% of patients have elevated IgE levels\textsuperscript{117} as well as eosinophilia of peripheral blood and tissue. The association with leukotriene inhibitors was described in the epidemiology section. Although ANCAs may be present and, if so, usually are directed to MPO, up to 60% of patients may be negative for ANCA.\textsuperscript{119,120} ANCA positivity is suggested to be associated with a higher incidence of renal disease, alveolar hemorrhage, mononeuritis multiplex, and purpura.\textsuperscript{119,120}

Cardiac involvement through coronary artery vasculitis and involvement of the myocardium by granulomatous disease may occur, and magnetic resonance angiography and contrast-enhanced MRI may be useful in its detection.\textsuperscript{129} Studies may be helpful in verifying the presence and extent of neuropathy, and biopsy of a nerve, usually the sural nerve, may establish the presence of vasculitis. Central nervous system disease occurs in a minority of patients—around 10%—but can be severe, particularly when a cerebral vasculitis is present.\textsuperscript{109} Chronic pachymeningitis, cerebral hemorrhage and thrombosis, pituitary involvement, cerebral nerve involvement, brain stem lesions, spinal cord involvement, and subarachnoid or subdural hemorrhage may occur. Substantiating the presence of cerebral vasculitis may be difficult because small vessels are affected and angiography is of little utility. However, CT and MRI may help to define infarctions, hemorrhages, mass lesions, meningeal involvement, and white matter changes (Figure 89-10). Lumbar puncture may be necessary if subarachnoid hemorrhage is suspected or to exclude the presence of meningeal infection.

Cardiac involvement through coronary artery vasculitis and involvement of the myocardium by granulomatous disease may occur, and magnetic resonance angiography and contrast-enhanced MRI may be useful in its detection.\textsuperscript{129} Increased risk of cardiovascular disease is noted in patients with AAV, including GPA.\textsuperscript{110,111} Symptomatic involvement of the gastrointestinal tract is not usually a major feature of GPA, or indeed of MPA. However, abdominal pain, bleeding, and diarrhea can occur as a result of the disease itself or through the effects of treatment, for example, corticosteroids can cause peptic ulceration and mycophenolate mofetil may cause diarrhea. The vasculitic process itself may lead to ulcerations, or even perforations, in the small or large intestine. A number of unusual presentations include involvement of the tongue and salivary glands (parotic, sublingual, or submandibular),\textsuperscript{112} pancreatic involvement that may mimic pancreatic cancer,\textsuperscript{113} and cholecystitis and hepatic granulomatous disease that can cause liver failure.\textsuperscript{114} Splenic involvement with vasculitis, granulomas, and necrosis is present in many patients in older autopsy series.\textsuperscript{115}

Allergic Granulomatosis with Polyangiitis

This syndrome was described by Churg and Strauss in 1951.\textsuperscript{116} It has three salient histopathologic features, namely, necrotizing vasculitis, tissue infiltration by eosinophils, and extravascular granulomas (Figure 89-11). To improve recognition of this disorder, Lanham suggested that diagnosis be based on clinical features comprising asthma, peak eosinophil count greater than 1500 cells/mL, and systemic vasculitis involving two or more organs.\textsuperscript{117,118} None of these clinical or pathologic features is entirely specific for AGPA, and they may not all be present or detectable concurrently.

The pathogenesis of AGPA is unknown, but a strong association has been noted with allergy and atopic disorders, including allergic rhinitis, nasal polyposis, and asthma. Around 70% of patients have elevated IgE levels\textsuperscript{117} as well as eosinophilia of peripheral blood and tissue. The association with leukotriene inhibitors was described in the epidemiology section. Although ANCAs may be present and, if so, usually are directed to MPO, up to 60% of patients may be negative for ANCA.\textsuperscript{119,120} ANCA positivity is suggested to be associated with a higher incidence of renal disease, alveolar hemorrhage, mononeuritis multiplex, and purpura.\textsuperscript{119,120}

Cardiac involvement through coronary artery vasculitis and involvement of the myocardium by granulomatous disease may occur, and magnetic resonance angiography and contrast-enhanced MRI may be useful in its detection.\textsuperscript{129} Increased risk of cardiovascular disease is noted in patients with AAV, including GPA.\textsuperscript{110,111} Symptomatic involvement of the gastrointestinal tract is not usually a major feature of GPA, or indeed of MPA. However, abdominal pain, bleeding, and diarrhea can occur as a result of the disease itself or through the effects of treatment, for example, corticosteroids can cause peptic ulceration and mycophenolate mofetil may cause diarrhea. The vasculitic process itself may lead to ulcerations, or even perforations, in the small or large intestine. A number of unusual presentations include involvement of the tongue and salivary glands (parotic, sublingual, or submandibular),\textsuperscript{112} pancreatic involvement that may mimic pancreatic cancer,\textsuperscript{113} and cholecystitis and hepatic granulomatous disease that can cause liver failure.\textsuperscript{114} Splenic involvement with vasculitis, granulomas, and necrosis is present in many patients in older autopsy series.\textsuperscript{115}

Allergic Granulomatosis with Polyangiitis

This syndrome was described by Churg and Strauss in 1951.\textsuperscript{116} It has three salient histopathologic features, namely, necrotizing vasculitis, tissue infiltration by eosinophils, and extravascular granulomas (Figure 89-11). To improve recognition of this disorder, Lanham suggested that diagnosis be based on clinical features comprising asthma, peak eosinophil count greater than 1500 cells/mL, and systemic vasculitis involving two or more organs.\textsuperscript{117,118} None of these clinical or pathologic features is entirely specific for AGPA, and they may not all be present or detectable concurrently.

The pathogenesis of AGPA is unknown, but a strong association has been noted with allergy and atopic disorders, including allergic rhinitis, nasal polyposis, and asthma. Around 70% of patients have elevated IgE levels\textsuperscript{117} as well as eosinophilia of peripheral blood and tissue. The association with leukotriene inhibitors was described in the epidemiology section. Although ANCAs may be present and, if so, usually are directed to MPO, up to 60% of patients may be negative for ANCA.\textsuperscript{119,120} ANCA positivity is suggested to be associated with a higher incidence of renal disease, alveolar hemorrhage, mononeuritis multiplex, and purpura.\textsuperscript{119,120}
Asthma and eosinophilia in AGPA

Ear, nose, and throat disease usually favors GPA.

Progressive renal failure uncommon in AGPA.

AGPA - GPA

Comments

+++ , very commonly seen; ++ , seen often; +, uncommon finding; AGPA, allergic granulomatosis with polyangiitis; GPA, granulomatosis with polyangiitis.

**Table 89-3** Differential Diagnostic Features of the Antineutrophil Cytoplasm Antibody–Associated Vasculitides

<table>
<thead>
<tr>
<th>Feature</th>
<th>Microscopic Polyangiitis</th>
<th>GPA</th>
<th>AGPA</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glomerulonephritis</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>Progressive renal failure uncommon in AGPA</td>
</tr>
<tr>
<td>Pulmonary infiltrates or nodules</td>
<td>++</td>
<td>+++</td>
<td>+</td>
<td>Asthma and eosinophilia in AGPA</td>
</tr>
<tr>
<td>Alveolar hemorrhage</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Upper airway disease</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Skin, purpura</td>
<td>+++</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Peripheral nerve involvement</td>
<td>+</td>
<td>++</td>
<td>+++</td>
<td>Often a prominent feature of AGPA</td>
</tr>
<tr>
<td>Central nervous system involve</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

Laboratory results may confirm the presence of an acute phase response, define the nature and extent of individual organ involvement, or contribute to diagnostic evaluation through detection or not of a range of immune antibody products. Leukocytosis, normocytic normochromic anemia, thrombocytosis, and elevated erythrocyte sedimentation rate and C-reactive protein may be indicative of an acute inflammatory state. In AGPA, peripheral eosinophilia in excess of 1500 cells/mL is often present, although occasional patients may have marked tissue eosinophilia without blood eosinophilia.

Detection of PR3- and MPO-ANCA now occurs by antigen-specific enzyme-linked immunosorbent assay (ELISA), widely available as routine laboratory testing. As a screening assay, indirect immunofluorescence (IIF) using ethanol-fixed human neutrophils reveals two major staining patterns of antibody binding to neutrophils comprising a cytoplasmic pattern (cANCA) and a perinuclear pattern (pANCA) (see Figure 89-1). The cANCA pattern usually equates to specificity for PR3-ANCA, and pANCA usually equates to MPO-ANCA. Guidelines for testing for ANCA have been agreed by international consensus, and if these guidelines are adhered to, the chance of missing a case of AAV is low while indiscriminant testing is avoided. The combination of IIF and ELISA testing increases specificity.

Thus, when the IIF test was combined with ELISA testing...
Treatment is tailored according to the disease severity at presentation, and more aggressive immunosuppression is required in the presence of life- or organ-threatening disease. An algorithm for therapy is shown in Figure 89-12.

**Induction of Remission**

Cyclophosphamide, an alkylation immunosuppressant agent, with glucocorticoids is considered the “standard of care” as induction therapy for those with generalized disease. Treatment should be continued for 3 to 6 months with remission rates following induction treatment varying from 35% to 93% in GPA and from 75% to 89% in MPA.

A dose-response relationship between cyclophosphamide and the risk of bladder cancer has been demonstrated, suggesting the need to treat with the lowest possible cumulative dose. There is therefore a need to balance the toxicity of treatment with the morbidity and mortality associated with disease.

One approach to limiting the cumulative cyclophosphamide dose has involved the use of pulsed cyclophosphamide versus the more traditional continuous daily regimens. A meta-analysis of three small, randomized studies suggested that pulsed intravenous cyclophosphamide was preferable (cANCA/anti-PR3 positive, pANCA/anti-MPO positive, diagnostic specificity increased to 99%). The sensitivity of the combination of cANCA plus anti-PR3 or pANCA plus anti-MPO for GPA or MPA was 73% and 67%, respectively. The question of the status of ANCA-negative patients arises also because around 10% of patients may be negative; the reasons for this may include low titer antibody levels, a poorly performing IIF or ELISA test, or genuinely ANCA-negative disease. However, if clinical and pathologic features are consistent with AAV, patients generally should be treated along similar lines.

Both at the initiation of therapy and during monitoring, a number of tools are available to measure the extent of disease activity and the damage caused through effects of disease or treatment. Chief among these following their validation and use in multicentered randomized prospective clinical trials are the Birmingham Vasculitis Activity Score (now in its third iteration) and the Vasculitis Damage Index. Other outcome measures include the Five Factor Score (FFS) and the Disease Extent Index (reviewed in Reference 129).

**TREATMENT**

The natural history of untreated AAV is rapidly progressive and usually fatal, with a 2-year mortality of 85%. The introduction of cyclophosphamide-based regimens to treat AAV in the 1970s has dramatically improved patient survival to around 80% at 5 years. The aim of treatment is to induce and then maintain disease remission and prevent relapse, using agents with the lowest adverse event profile.
to daily oral cyclophosphamide in terms of remission and reduced adverse events.\textsuperscript{135} A more recent randomized controlled trial in 149 patients with generalized disease performed by the European Vasculitis Study Group (EUVAS) showed no statistically significant difference in the time to remission or in the proportion of patients attaining remission at 9 months between those randomized to pulse cyclophosphamide compared with daily oral cyclophosphamide, despite an absolute lower cumulative dose of cyclophosphamide that was associated with a lower rate of leukopenia in the pulsed limb. This study did not have sufficient duration of follow-up and was not powered to detect a difference in relapse rates.\textsuperscript{116}

Other immunosuppressants may be used as alternatives to cyclophosphamide. Rituximab, an anti-CD20 chimeric monoclonal antibody that depletes B cells, has been shown to be effective in inducing remission in AAV in two randomized controlled trials. RITUXVAS compared cyclophosphamide- and rituximab-based regimens in newly diagnosed patients with severe AAV; the median glomerular filtration rate (GFR) at entry was 18 mL/min.\textsuperscript{137} Forty-four patients were randomly assigned to receive two pulses of cyclophosphamide and four of rituximab or six to ten pulses of cyclophosphamide. Both groups received the same prednisolone dose and were followed for a minimum of 12 months. At 12 months, no differences were noted between the two groups in terms of remission rates, times to remission, or severe adverse events. The RAVE trial also supports the use of rituximab as an induction agent in AAV. This multicentered, randomized, double-blind, placebo-controlled trial, involving 197 ANCA-positive patients, of whom 52% had renal disease and 28% had alveolar hemorrhage, compared rituximab with cyclophosphamide during induction. Both limbs received the same reducing glucocorticoid regimen. Rituximab was shown to be noninferior to cyclophosphamide in the rate of disease remission at 6 months in the absence of corticosteroid therapy. In this study, rituximab may have been superior to cyclophosphamide in inducing remission in patients with relapsing disease. The rate of adverse events was the same between the two treatment limbs.\textsuperscript{138} Of course, both trials were limited by the follow-up time, and long-term outcome trials are required.

ANCA levels fall with rituximab treatment, but this fall occurs more slowly than the improvement in renal disease activity.\textsuperscript{95,139} Induction of remission in patients is generally associated with complete peripheral B cell depletion following rituximab. B cell depletion, although to a lesser extent, also occurs with cyclophosphamide.\textsuperscript{138} However, in studies of patients with relapsing disease, it has been shown that measurement of B cells peripherally is not a good surrogate marker for the risk of relapse because relapses may occur before the peripheral B cell population has recovered.\textsuperscript{139}

Methotrexate has been identified as another alternative agent to cyclophosphamide in the induction phase of limited disease in the absence of vital organ involvement and with serum creatinine less than 150 µmol/L. The NORAM trial, involving 100 patients with newly diagnosed AAV, compared methotrexate with cyclophosphamide. Treatment was tapered and withdrawn in both limbs at 12 months. This study demonstrated noninferiority in the remission rate at 6 months between patients treated with oral methotrexate versus those treated with oral cyclophosphamide. If treated with methotrexate, time to remission was longer in those with more extensive disease or lung involvement, suggesting that methotrexate is more appropriate in patients with limited early disease. The rate of leukopenia was significantly lower among those in the methotrexate limb. Relapse rates at 18 months were acceptably high in both limbs, suggesting that immunosuppression should be continued for longer than 12 months.\textsuperscript{140}

A small study of 35 patients recently suggested that mycophenolate mofetil may be useful as an alternative agent to cyclophosphamide for induction therapy in those with generalized disease and moderate renal impairment.\textsuperscript{141} A larger, multicenter, randomized trial is being conducted by EUVAS to address this question (www.vasculitis.org).

### Maintenance of Remission

It has been suggested that cyclophosphamide is used for the induction of remission, and another immunosuppressant agent, with fewer associated adverse events, is used to maintain remission. Four major agents have been used in this way: methotrexate, azathioprine, mycophenolate mofetil, and leflunomide. CYCAZAREM, a multicenter, randomized, controlled trial, compared azathioprine with cyclophosphamide to maintain remission, and compared relapse rates within 18 months of diagnosis. Investigators showed that following the induction of remission with cyclophosphamide for a minimum of 3 months, no difference in relapse rates was observed between the two limbs.\textsuperscript{142} Although in this trial no difference in adverse events between the two treatment limbs was reported, azathioprine is considered to be less toxic than cyclophosphamide. A retrospective study using historical controls also suggested that relapse rates were no different in those treated with azathioprine compared with cyclophosphamide during the maintenance phase of treatment. This study suggested that switching to azathioprine in patients with PR3-AAV, who remained ANCA-positive at the time of remission, was associated with a high risk of relapse.\textsuperscript{143}

Following induction of remission with cyclophosphamide, methotrexate is as effective as azathioprine in maintaining remission and may be used in those with normal renal function.\textsuperscript{144}

Small case series have suggested that mycophenolate mofetil is well tolerated when used as maintenance therapy.\textsuperscript{145,146} However, results from a large study of 154 adult patients randomized to receive azathioprine or mycophenolate for the maintenance of remission (IMPROVE study) suggested that azathioprine may be associated with fewer total relapses than mycophenolate.\textsuperscript{144} However, no difference was observed in the number of major relapses or in damage sustained during the study as measured by the vasculitis damage index. No difference in adverse events was described in this study.

In a multicenter trial, following induction of remission with cyclophosphamide and prednisolone, 54 patients with GPA were randomized to receive leflunomide or methotrexate as maintenance therapy. This study was terminated prematurely owing to the high incidence of major relapse in those treated with methotrexate, but findings suggested that leflunomide was effective in preventing major relapses.
Conversely, leflunomide was associated with more adverse events than occurred with methotrexate. A dose of 30 mg/day of leflunomide was used; therefore, this may not be an appropriate dose.\textsuperscript{48}

Guidelines suggest that patients with AAV should continue on maintenance therapy for at least 24 months, maybe longer, following successful remission.\textsuperscript{149} The IMPROVE study continued therapy for 40 months and reported relapse rates of almost 50%.\textsuperscript{147}

**Adjuvant Therapy**

Adjunct plasmapheresis in adult patients presenting with severe progressive renal failure (creatinine >500 µmol/L) has been shown, in the MEPEX study, to be superior to methylprednisolone in reducing the number of patients who remain dialysis dependent.\textsuperscript{150} The underpinning principle of plasma exchange is the removal of circulating autoantibodies, with the by-product being the removal of other plasma proteins, including coagulation factors. No randomized trials have examined the use of plasma exchange in diffuse alveolar hemorrhage (DAH). A single retrospective review of 20 patients who presented with DAH associated with AAV treated with plasma exchange showed improved outcome with no complications of therapy compared with historical controls.\textsuperscript{151} A randomized trial conducted by EUVAS aims to investigate the efficacy of plasma exchange in AAV with pulmonary hemorrhage or renal involvement (GFR <50 mL/min) (www.vasculitis.org).

**Alternative Agents**

Alternative agents may be used for those who fail to respond to first-line treatment, or when there is cyclophosphamide intolerance. Rituximab has been shown to be effective in patients with refractory disease in whom maximal treatment with conventional therapy had not previously resulted in remission. Antilymphocyte therapies may also be effective in refractory disease. Treatment with antithymocyte globulin, which consists of polyclonal antibodies directed against T cell surface antigens and results in rapid T cell depletion, has been shown to be of benefit in a small pilot study of 15 patients.\textsuperscript{152} Deoxyspergualin is an antiproliferative agent with lymphocyte and macrophage inhibitory functions. In a study of patients with relapsing or refractory AAV treated for 6 months with deoxyspergualin, 42 of 44 patients achieved at least partial remission and 45% achieved full remission. However, relapse occurred in 44% with a median time of 170 days, and 53% developed severe or life-threatening treatment-related adverse events.\textsuperscript{153} Campath, a humanized anti-CD52 antibody that depletes lymphocytes and monocytes, has shown benefit in refractory disease. However, it should be used with caution in those older than 50 years and dialysis dependent at the time of treatment.\textsuperscript{154}

Intravenous immunoglobulin (IVIG) has been used to induce remission in AAV. In a double-blind controlled trial, 34 patients were randomized to receive a single course of IVIG or a placebo, alongside conventional treatment. Those who received IVIG had significantly reduced Birmingham Vasculitis Activity Score (BVAS) at 3 months compared with the placebo limb, but this effect was not maintained beyond 3 months.\textsuperscript{155} Six months of IVIG treatment has been shown to be effective in inducing remission in the context of disease relapse,\textsuperscript{156} with a study of six patients also suggesting effectiveness of IVIG as sole therapy.\textsuperscript{157} IVIG may be useful where significant immunosuppression should be avoided.

EULAR has recently produced guidelines to help guide the management of patients with AAV\textsuperscript{131} (Table 89-4). However, it must be noted that most studies have excluded patients with AGPA.

**Table 89-4  EULAR Recommendations on the Management of AAV (12 of 15 Applicable to AAV)**

| 1. | Manage in collaboration with, or at, centers of expertise |
| 2. | ANCA testing should be performed only in an appropriate clinical context |
| 3. | A positive biopsy is strongly supportive of vasculitis—recommend the procedure to assist diagnosis and further evaluation for patients suspected of having vasculitis |
| 4. | Structured clinical assessment, urinalysis, and other basic laboratory tests at each clinical visit |
| 5. | Categorize according to different levels of severity to assist treatment decisions |
| 6. | Cyclophosphamide (oral or IV) and glucocorticoids for remission induction |
| 7. | Combination of methotrexate and glucocorticoid as a less toxic alternative to cyclophosphamide for the induction of remission in non–organ-threatening or non–life-threatening ANCA-associated vasculitis |
| 8. | Use of high-dose glucocorticoids as an important part of remission induction therapy; high dose continued for the first month (usual practice to start at 1 mg/kg/day). Dosage should not be reduced to <15 mg/day in first month |
| 9. | Plasma exchange for selected patients with rapidly progressive severe renal disease to improve renal survival |
| 10. | Remission maintenance therapy with a combination of low-dose glucocorticoid therapy and azathioprine/leflunomide/methotrexate |
| 11. | Alternative immunomodulatory therapy choices should be considered for patients who do not achieve remission or relapse on maximal doses of standard therapy |
| 12. | Investigation of persistent unexplained hematuria in patients with prior exposure to cyclophosphamide |

AAV, antineutrophil cytoplasm antibody–associated vasculitides; ANCA, antineutrophil cytoplasm antibody; EULAR, European League Against Rheumatism.


**Treatment of Allergic Granulomatosis with Polyangitis**

Patients without poor prognostic disease according to the FFS can be managed by corticosteroids alone.\textsuperscript{158} However, relapse is common, with 35% of patients relapsing with a mean follow-up of 56 months. Long-term steroids were necessary in 79% mainly owing to asthma. In patients with one or more poor prognostic factors, induction therapy with corticosteroids and cyclophosphamide is necessary.\textsuperscript{121} No maintenance therapy was prescribed in this study, and relapses were common, occurring in 74%. As with those patients without poor prognostic factors, long-term use of steroids occurred in the majority (81%), even after a median follow-up of 8 years.
Adverse Events

Despite improvement in survival resulting from improved treatment strategies, a recent EUVAS short-term outcome study, which included 524 patients prospectively recruited into four clinical trials, reported a 1-year mortality probability of 11.1%. The cause of death was related to an adverse event of treatment in 59%, and infection was the overwhelming factor. Active vasculitis accounted for only 14% of deaths. This study highlights the burden of using nonselective immunosuppressant agents, which suppress the whole immune system, not just the production of ANCA. Patients receiving such treatments, including glucocorticoids, are at increased risk for opportunistic infection, with respiratory tract infection and generalized sepsis being the most common infections in the EUVAS study previously discussed. Factors predictive of infection include age, severity of renal dysfunction, leukopenia, and intensity and duration of immunosuppression. Guidelines advocate the use of prophylaxis against Pneumocystis jiroveci in patients receiving cyclophosphamide. Influenza vaccinations have been shown to be safe and effective in AAV without association with relapse in a retrospective study of 230 patients. Rituximab is associated with an impaired secondary humoral response, making immunization ineffective during treatment.

Other adverse events are specific to treatments. Glucocorticoids are known to have a broad adverse event profile, including steroid-associated diabetes, avascular necrosis, and ocular cataract formation; monitoring for all of these should be provided during treatment. Prophylaxis against osteoporosis and peptic ulceration has become routine, especially in those receiving high-dose corticosteroids. The adverse events associated with cyclophosphamide have already been highlighted. Concomitant treatment with mesna, which binds to the toxic metabolite of cyclophosphamide, is advised on the first day of a cyclophosphamide pulse and in oral regimens to reduce the risk of bladder toxicity. Male and female infertility is a recognized complication of cyclophosphamide therapy in autoimmune disease. In females, infertility is related to high cumulative doses and older age at treatment. Minimizing the dose of cyclophosphamide is important, as is the continuing development of methods of fertility preservation. Patients should be counseled as to the risk of infertility, and cryopreservation of sperm and oocytes should be offered if appropriate.

Adverse events can also be attributable to the disease itself. In a study of 198 patients, followed for a median of 6.1 years, it was shown that AAV is associated with increased risk of venous thromboembolism, especially when the disease is active. Avoidance of classic risk factors for venous thromboembolism is therefore important, as is the use of prophylaxis during prolonged immobility. Because patients are now surviving their acute illness, the long-term consequences of disease activity need to be considered, along with the already discussed burden of therapy.

Outcome

Over the past 30 years, treatment has improved outcomes for patients with AAV. Most patients respond to treatment, and 85% achieve remission. Older age may predict treatment resistance. Patient survival is reported as 45% to 91% at 5 years and as 75% to 88% at 10 years compared with 80% mortality at 2 years if left untreated.

Despite advances in therapy, patients continue to have substantially higher mortality than a matched background population, as was shown by a recent study of long-term outcomes of patients recruited to EUVAS studies. The mortality rate ratio was 2.6 compared with the normal population, with advanced renal failure, increasing age, and a high BVAS being the main predictors of an adverse outcome. Several other studies have identified increasing age and worsening renal function as poor prognostic markers. In this study, no difference in survival was reported between patients with GPA and those with MPA in contrast to other studies. Mortality is highest in the first year, with 1-, 2-, and 5-year survival of 88%, 85%, and 78%. Disease and therapy-related deaths, particularly infection, account for most deaths in the first year. Infection remains an important cause of death even beyond 1 year, but malignancy and cardiovascular disease are also common.

End-stage kidney disease (ESKD) is not uncommon in patients with AAV; approximately 20% of those presenting who have evidence of renal involvement will develop ESKD by 5 years. In a multivariate analysis, renal survival was best predicted by presenting serum creatinine and percentage of normal glomeruli in the diagnostic biopsy. Patients who develop ESKD should be considered for transplantation. Using United Network for Organ Sharing (UNOS) data from 1996-2007, 919 patients with ESKD secondary to GPA were identified. Adjusted outcomes for graft loss, death, and functional graft loss were better in those patients with GPA compared with other causes of ESKD. Outcome in those patients who remain on dialysis is similar to that with other causes of ESKD.

Cardiovascular disease is more common in patients with AAV than in the healthy population or in patients with matched levels of chronic kidney disease. It is currently unclear whether this is related to persistent inflammation or to therapy with prolonged steroid use. Malignancy rates in treated AAV patients are also higher than in the healthy population, particularly nonmelanoma skin and bladder cancer and acute myeloid leukemia. The cumulative dose of cyclophosphamide is important in determining risk of cancer; patients receiving more than 36 g are at greatest risk. Preliminary results from the EUVAS group suggest...
that the philosophy of minimizing cyclophosphamide usage may be proving correct; a follow-up study of patients recruited to clinical trials revealed only nonmelanoma skin cancer to be substantially increased compared with the healthy population. The clinical course in patients with AAV is difficult to predict; approximately 50% of patients will relapse over 5 years. Predictors of relapse have included anti-PR3 antibodies, lung and upper respiratory involvement, age, nasal carriage of S. aureus, and absence of severe renal involvement. Patients with ESKD have lower rates of relapse. Daily oral cyclophosphamide may be more effective than intermittent cyclophosphamide but at the expense of higher risk of death and adverse events. Using microarray analysis of purified T cells, a CD8*T cell transcription signature has been identified that can identify patients at risk of relapse. The subset of genes defining the poor prognostic group was enriched for genes involved in the interleukin (IL)-7 receptor pathway and T cell receptor signaling, and for genes expressed by memory T cells. A model using only three genes—ITAG2, PTPN22, and NOTCH1—was predictive of the poor prognostic group. However, this signature was not discernible following treatment. Confirmation of this signature by others raises the possibility of individualized therapy.

**SUMMARY**

Understanding of the causes and pathogenic mechanisms underlying AAV has progressed apace over the past 20 years. The central place of ANCA and immune processes is appreciated. The mortality incurred by patients suffering from AAV has fallen steadily, but the chronic relapsing nature of the disease, without definitive treatments that offer cure, continues to extract a heavy toll on the health of afflicted individuals of all ages. Understanding of the most efficacious ways with which to control disease using corticosteroids, cyclophosphamide, other immune suppressants, and plasma exchange has increased through use of randomized prospective clinical trials. However, although rituximab holds promise, no therapy has yet been licensed for AAV. Despite the rarity of the condition, the large unmet need for safe, effective therapies remains a challenge to pharmaceutical companies, particularly because mechanisms underpinning the disease are understood sufficiently to provide realistic targets for new therapies.

**Selected References**


ANTINEUTROPHIL CYTOPLASM ANTIBODY–ASSOCIATED VASCULITIS


Full references for this chapter can be found on www.expertconsult.com.
References


8. Murty GE: Wegener's granulomatosis: orthonilorangymological manifesta-


12. Thickett DR, Richter AG, Nathani N, et al: Pulmonary manifesta-
tions of anti-neutrophil cytoplasmic antibody (ANCA)-positive vasculi-


14. Pakrou N, Selva D, Leibovitch I: Wegener's granulomatosis: ophthal-
mic manifestations and management, Semin Arthritis Rheum 35:284–


ment in Wegener's granulomatosis: an analysis of 324 consecutive patients at the Mayo Clinic, Ann Neurol 33:4–9, 1993.


24. de Grooth K, Adu D, Savage COS: The value of pulse cyclophospho-

25. de Grooth K, Harper L, Jayne DR, et al: Pulse versus daily oral cyclo-
phosphamide for induction of remission in antineutrophil cytoplas-


27. Stone JH, Merkel PA, Spera R, et al: Rituximab versus cyclophos-


Polyarteritis Nodosa

Definition and Classification

The term periarteritis nodosa was originally introduced in 1866 and subsequently used to describe any form of systemic vasculitis.1 The term has been modified to polyarteritis nodosa (PAN), and the definition has been improved to consist of “necrotizing inflammation of medium-sized or small arteries without glomerulonephritis or vasculitis in arterioles, capillaries, or venules.”2

The American College of Rheumatology (ACR) criteria, which can be used to classify patients as having PAN, in order to distinguish them from patients with other forms of primary systemic vasculitis, do not differentiate between PAN and microscopic polyangiitis (MPA); both conditions are included under the umbrella of PAN. Three of 10 criteria (listed in Table 90-1) are required.3

The sensitivity of the ACR criteria for PAN (or MPA) is 82.2%, with a specificity of 86.6% when used to classify the disease.3 Definitions of vasculitis (including PAN) were published in 1994 by the Chapel Hill Consensus Conference (CHCC).2 Microscopic polyangiitis, granulomatositis with polyangiitis (GPA) (formerly Wegener's granulomatosis), and Churg-Strauss syndrome were distinguished from PAN on the basis of the pathologic findings of small vessel involvement in the former three diseases; by contrast, small vessel involvement is absent in PAN.

There has been, and still is, much confusion in the terms used to describe patients with different forms of vasculitis. It is worth considering this for a moment because, on reviewing patients who have previously been given the label of "PAN," the clinical features at presentation might actually suggest another form of vasculitis, typically MPA or GPA. In an attempt to rationalize the naming of different forms of vasculitis, the European Medicines Agency (EMEA) produced an algorithm to assist in the classification of patients with systemic small and medium vessel primary systemic vasculitis.4 A decision tree approach was used, effectively putting PAN at the bottom of the tree.4 In other words, patients were assigned to any other form of vasculitis, and only as a last resort, if no other diagnosis was made, the term PAN was applied. Perhaps this might seem a rather negative view of the condition, but because of the overuse of the term PAN in preference to others in the past, the true incidence and prevalence of PAN are actually low.

The EMEA algorithm is particularly useful because MPA is a more common condition than PAN but is absent from the ACR classification criteria (it is considered part of PAN); whereas in the CHCC definition PAN and MPA are treated as separate entities. There are important differences between PAN and MPA in terms of pathogenesis, organ involvement, tendency to relapse, and prognosis.3 In the CHCC definition PAN is a medium vessel disease; by comparison, MPA is predominantly a small vessel disease that includes glomerulonephritis and pulmonary capillaritis.2

Patients are assessed using the EMEA algorithm if they have a clinical diagnosis of an antineutrophil cytoplasm antibody (ANCA)-associated vasculitis or PAN. In the algorithm, PAN is regarded as a diagnosis of exclusion. The first part of the algorithm is to determine whether the patient fulfills the Lanham4 or ACR criteria3 for Churg-Strauss syndrome (CSS). If they do, they are classified as having CSS. If not, the next step is to determine whether

**Polyarteritis Nodosa (PAN)**

The term periarteritis nodosa was originally introduced in 1866 and subsequently used to describe any form of systemic vasculitis. The term has been modified to polyarteritis nodosa (PAN), and the definition has been improved to consist of “necrotizing inflammation of medium-sized or small arteries without glomerulonephritis or vasculitis in arterioles, capillaries, or venules.”

The American College of Rheumatology (ACR) criteria, which can be used to classify patients as having PAN, in order to distinguish them from patients with other forms of primary systemic vasculitis, do not differentiate between PAN and microscopic polyangiitis (MPA); both conditions are included under the umbrella of PAN. Three of 10 criteria (listed in Table 90-1) are required.3

The sensitivity of the ACR criteria for PAN (or MPA) is 82.2%, with a specificity of 86.6% when used to classify the disease.3 Definitions of vasculitis (including PAN) were published in 1994 by the Chapel Hill Consensus Conference (CHCC).2 Microscopic polyangiitis, granulomatositis with polyangiitis (GPA) (formerly Wegener's granulomatosis), and Churg-Strauss syndrome were distinguished from PAN on the basis of the pathologic findings of small vessel involvement in the former three diseases; by contrast, small vessel involvement is absent in PAN.

There has been, and still is, much confusion in the terms used to describe patients with different forms of vasculitis. It is worth considering this for a moment because, on reviewing patients who have previously been given the label of “PAN,” the clinical features at presentation might actually suggest another form of vasculitis, typically MPA or GPA. In an attempt to rationalize the naming of different forms of vasculitis, the European Medicines Agency (EMEA) produced an algorithm to assist in the classification of patients with systemic small and medium vessel primary systemic vasculitis.4 A decision tree approach was used, effectively putting PAN at the bottom of the tree.4 In other words, patients were assigned to any other form of vasculitis, and only as a last resort, if no other diagnosis was made, the term PAN was applied. Perhaps this might seem a rather negative view of the condition, but because of the overuse of the term PAN in preference to others in the past, the true incidence and prevalence of PAN are actually low.

The EMEA algorithm is particularly useful because MPA is a more common condition than PAN but is absent from the ACR classification criteria (it is considered part of PAN); whereas in the CHCC definition PAN and MPA are treated as separate entities. There are important differences between PAN and MPA in terms of pathogenesis, organ involvement, tendency to relapse, and prognosis.3 In the CHCC definition PAN is a medium vessel disease; by comparison, MPA is predominantly a small vessel disease that includes glomerulonephritis and pulmonary capillaritis.2

Patients are assessed using the EMEA algorithm if they have a clinical diagnosis of an antineutrophil cytoplasm antibody (ANCA)-associated vasculitis or PAN. In the algorithm, PAN is regarded as a diagnosis of exclusion. The first part of the algorithm is to determine whether the patient fulfills the Lanham4 or ACR criteria3 for Churg-Strauss syndrome (CSS). If they do, they are classified as having CSS. If not, the next step is to determine whether

Table 90-2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Positive Prediction for PAN</th>
<th>Negative Prediction for PAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive hepatitis B serology</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Arteriographic abnormalities</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Mononeuropathy or polynueurophy</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Positive antineutrophil cytoplasm antibody</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Ear, nose, or throat signs</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Glomerulopathy</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Cryoglobulinemia</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>


they fulfill the ACR criteria or CHCC definition for GPA, either by direct evidence with histology or with appropriate surrogate markers and a positive ANCA. If they fulfill these requirements, they are classified as having GPA. If not, they proceed down the algorithm to see whether they fulfill a definition of MPA. This is either by histology showing small vessel vasculitis or glomerulonephritis and no surrogate markers for GPA or with surrogate markers of glomerulonephritis and a positive ANCA. Only when CSS, GPA, and MPA are excluded is a diagnosis of PAN possible in this schema. To fulfill the definition of PAN in this algorithm, there must be histology or angiographic features consistent with the diagnosis. Any remaining patients are determined to be unclassifiable. The initial validation of the algorithm made use of paper cases assessed by a number of experts in vasculitis. In fact, every case was assigned to one of the conditions, without the need to describe any case as unclassifiable. This may not reflect clinical practice, where some patients appear to have an overlap between different forms of vasculitis. This is an important issue and has started to be addressed by a recent study comparing the incidence of PAN in three European regions: 4.4 to 9.7 per million by the ACR criteria versus 0 to 0.9 per million with the CHCC definition.

The incidence of PAN is 2 to 9 per million in Europe and the United States by ACR criteria. A higher incidence of 16 per million (by CHCC definition) is reported in Kuwait and 77 per million described in an Alaskan population endemic for HBV, although it was based on only 13 actual cases of HBV PAN (in a study performed before ACR criteria or CHCC definitions). The prevalence of PAN by ACR criteria is 31 to 33 per million in Western Europe; 2 to 9 per million in Germany by CHCC definition. A small study from Australia assessed the incidence and prevalence of different sorts of vasculitis between 1995 and 2005 in one area, suggesting a fall in the incidence of PAN from 2.3 per million per annum to 1.1 per million per annum. A study from 2009 assessed the incidence and survival of patients with different forms of vasculitis and estimated the incidence of PAN to be 0.9 (between 0 and 1.7) per million per annum. This compares with an almost 10-fold difference in incidence of GPA and MPA (between 9.8 and 10.1 per million per annum).

PAN can occur at any age, but the commonest age range at diagnosis is 40 to 60 years. There is no clear gender difference. In a multiethnic population from France, patients with European ancestry had a higher prevalence of PAN compared with those without.

Etiology and Pathogenesis

PAN is often related to infection with hepatitis B. The incidence of HBV-related PAN follows HBV infection rates, previously accounting for 7% to 38.5% of patients...
diagnosed with PAN.16,22 The prevalence of HBV-related PAN has reduced in recent years as a result of vaccination for HBV and improved screening of blood products.16,23 Pancreatitis develops in 1% to 5% of patients with HBV infection,14 which confers an approximately 1000-fold increase in risk compared with the background population.12 By contrast, in Alaska, an area endemic for HBV, there is a substantial increased annual incidence of HBV PAN (77 per million), primarily due to vertical transmission.14 Documented exposure to HBV occurs from blood products, intravenous drug use, and sexual contact. Screening of blood products for HBV and mass vaccination against the virus has successfully reduced the incidence of HBV PAN, as shown in French patients with the proportion of PAN due to HBV falling from 38.5% in the 1970s to 17.4% by the period 1997-2002.14

In HBV-related PAN the postulated mechanism of vasculitis includes direct vessel injury by replicating virus or deposition of immune complexes. Deposition of immune complexes leads to activation of the complement cascade, resulting in an inflammatory response and subsequent endothelial damage. The vasculitis typically occurs within the first few months following HBV infection and may be the first presenting feature of this infection. Evidence for the pathogenic nature of HBV and immune complexes is supported by the effectiveness of a treatment strategy to eradicate HBV with antiviral therapy and removal of immune complexes by plasmapheresis without the need for long-term immunosuppression.24,25

Saaoun and colleagues16 described 31 patients who had hepatitis C virus (HCV)-associated vasculitis, which they classified as PAN according to ACR criteria and CHCC definitions. This cohort, representing approximately one-fifth of their HCV vasculitis patients, had more frequent fever, weight loss, severe hypertension, gastrointestinal tract involvement, severe acute sensory-motor multifocal mononeuropathy, kidney and liver microaneurysms, and increased C-reactive protein compared with other HCV vasculitis patients. By contrast, their response to therapy was more successful.

In the remainder of patients with PAN, the etiology is unknown. Genetic, infectious, and environmental agents are thought to be important, but there is no conclusive evidence.31 The fact that idiopathic PAN responds to immunosuppressive therapy suggests an immunologic mechanism. In idiopathic PAN the role of immune complexes is unclear. There is evidence of endothelial dysfunction, an increase in inflammatory cytokines, and an increase in expression of adhesion molecules. There is a propensity for the inflammatory lesions to occur at the sites of bifurcation in vessels, where there is most likely to be turbulent flow. Following the inciting event, there is focal and segmental necrotizing inflammation of medium- and small-sized arteries. This leads to intimal proliferation with subsequent thrombosis, resulting in ischemia or infarction of the organ or tissue supplied by these arteries.

There are case reports describing the development of PAN in patients with pre-existing hairy cell leukemia.27,28 Most cases had undergone splenectomy before the development of PAN. Potential mechanisms for the association between hairy cell leukemia and PAN include cross-reactivity of antibodies between the tumor cells and the endothelium, direct damage of the endothelium by tumor cells, and local production of proinflammatory cytokines triggering vessel wall damage.28

Pathologic Features

Due to the protean manifestations of the disease, a number of different sites can be biopsied. Sampling should be directed by the pattern of clinical involvement suggested by the history or physical examination. Muscle, peripheral nerves, kidney, testis and rectum, when involved, provide the best yield. Skin involvement and a positive skin biopsy do not always indicate evidence of systemic involvement (see section on cutaneous PAN). A biopsy of a medium-sized artery will show “focal and segmental” transmural necrotizing inflammation (Figure 90-1). If there are any clinical or laboratory features indicating involvement of small vessels, further assessment should be undertaken because it suggests an alternative form of vasculitis such as MPA.32

Inflammatory change at the bifurcation of vessels is reported to be common. The coexistence of different stages of inflammation and scarring, as well as normal vessel wall, is typical in PAN. Areas of acute inflammation will usually have a pleomorphic cellular infiltrate of lymphocytes, neutrophils, macrophages, and eosinophils. Aneurysms can occur at the site of active lesions, and this morphologic appearance is what led to the term “nodosas” (Figure 90-2). Proliferative scarring in other areas may lead to vessel narrowing.30,31

Clinical Features

PAN can present at any age, but it typically occurs in the age range of 40 to 60. There is no significant gender difference. The diagnosis can be difficult to make because individual features can occur in a variety of other diseases. Table 37-5 describes the disease's clinical manifestations.

Figure 90-1 This cross-section of a medium-sized artery demonstrates transmural inflammation with infiltrating lymphocytes typical of polyarteritis nodosa. The lesion is at an early stage, before occlusion of the whole vessel or weakness of the vessel wall leading to aneurysm formation.
90-3 summarizes the clinical findings in PAN. The presence of nonspecific constitutional symptoms such as fever, weight loss, joint and muscle pains (found in 65% to 80% of cases of PAN), and ischemic symptoms in one or more organ system should raise the possibility of a systemic vasculitis.\(^\text{17,24}\) Organ-specific manifestations may be present at the onset of disease, or there may be a more low-grade course, resulting in the development of specific organ involvement months to years later. Involvement of organs may be in isolation or as multisystem disease. In a Swedish series, the most common manifestations occurred at disease onset: nervous system (55%), skin (44%), abdominal (33%), and renal (11%).\(^\text{17}\) A review of 348 patients diagnosed as having PAN since the 1960s\(^\text{34}\) suggested that their mean age of diagnosis was 51 and the most frequent symptoms were generalized features in more than 90% of cases, neurologic features in 79%, skin involvement in 50%, abdominal involvement in 36%, hypertension in 35%, renal artery microaneurysms in 66%, and histologically proven PAN in 70%. The subset of 123 patients with hepatitis B–associated PAN was more likely to suffer peripheral neuropathy, abdominal pain, cardiomyopathy, orchitis, and hypertension compared with 225 non-HBV PAN patients. The relapse rate was 22% over a follow-up of 6 years (28% for non–HBV-related PAN compared with 11% for HBV PAN). The mortality rate was 25% overall (20% for

---

**Figure 90-2** This celiac axis angiogram from a patient with polyarteritis nodosa demonstrates an irregular vascular pattern with small aneurysms in the hepatic vessels.

---

**Table 90-3** Organ Involvement in Polyarteritis Nodosa (PAN)

<table>
<thead>
<tr>
<th>System</th>
<th>Comment</th>
<th>Frequency</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constitutional</td>
<td>Fever and weight loss (current and previous)</td>
<td>&gt;90%</td>
<td>34</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>Arthritis, arthralgia, myalgia, or weakness; when muscle is involved, it provides a useful site for biopsy</td>
<td>24%-80%</td>
<td>17,24</td>
</tr>
<tr>
<td>Skin</td>
<td>Purpura, nodules, livedo reticularis, ulcers, bullous or vesicular eruptions, and segmental skin edema</td>
<td>44%-50%</td>
<td>17,34-36</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>Cardiac ischemia, cardiomyopathy, hypertension</td>
<td>35%</td>
<td>24,30,34</td>
</tr>
<tr>
<td>Ear, nose, and throat</td>
<td>No involvement. Nasal crusting, sinusitis, and hearing loss suggest an alternative diagnosis such as granulomatosis with polyangiitis</td>
<td>None</td>
<td>24,30,34</td>
</tr>
<tr>
<td>Respiratory</td>
<td>Lung involvement not seen in PAN; abnormal respiratory findings suggest alternative diagnosis</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Abdominal</td>
<td>Pain is an early feature of mesenteric artery involvement. Progressive involvement may cause bowel, liver, or splenic infarction; bowel perforation; or bleeding from a ruptured arterial aneurysm. Less common presentations include appendicitis, pancreatitis, or cholecystitis as a result of ischemia or infarction. The presence of abdominal tenderness or peritonitis and blood loss on rectal examination should be assessed</td>
<td>33%-36%</td>
<td>17,34</td>
</tr>
<tr>
<td>Renal</td>
<td>Vasculitis involving the renal arteries is present in many cases but does not commonly give rise to clinical features. It can present with renal impairment, renal infarcts, or rupture of renal arterial aneurysms. Glomerular ischemia may result in mild proteinuria or hematuria, but red cell casts are absent because glomerular inflammation is not a feature. If there is evidence of glomerular inflammation, then an alternative diagnosis such as microscopic polyangiitis or granulomatosis with polyangiitis must be considered. Hypertension is a manifestation of renal ischemia causing activation of the renin-angiotensin system.</td>
<td>11%-66%</td>
<td>17,34</td>
</tr>
<tr>
<td>Nervous system</td>
<td>Mononeuritis multiplex, with sensory symptoms preceding motor deficits. Central nervous system involvement is a less frequent finding and can present with encephalopathy, seizures, and stroke</td>
<td>55%-79%</td>
<td>17,34</td>
</tr>
<tr>
<td>Ocular</td>
<td>Visual impairment, retinal hemorrhage, and optic ischemia</td>
<td>Rare</td>
<td>24,30</td>
</tr>
<tr>
<td>Other</td>
<td>Breast or uterine involvement is rare; testicular pain from ischemic orchitis is a characteristic feature, albeit an uncommon presentation</td>
<td>Rare</td>
<td>37,38</td>
</tr>
</tbody>
</table>
Clinical Assessment of Patients

In patients where vasculitis is suspected, a thorough history and physical examination is necessary to identify the potential organ systems involved. Use of the Birmingham Vasculitis Activity Score as a checklist of important features of the disease is recommended to guide further targeted investigation and to be used subsequently in devising a management plan for treatment. Table 90-3 lists the typical features to be assessed in suspected PAN.

Laboratory Testing

No specific laboratory tests are available to diagnose PAN, but some may be useful to support the diagnosis, identify organs that may be affected, and rule out alternative diagnoses (Table 90-4).

Radiology

The angiogram is the imaging modality of choice, but increasingly it is being replaced by less invasive, safer techniques. The classic findings are multiple small aneurysms, vessel ectasia, and focal occlusive lesions in medium-sized vessels, typically in the renal and mesenteric arteries. The reported sensitivity of angiography is as high as 89% and specificity 90% when performed in individuals suspected of having vasculitis.

<table>
<thead>
<tr>
<th>Test</th>
<th>Supports Diagnosis of PAN</th>
<th>Supports an Alternative Diagnosis</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevated C-reactive protein</td>
<td>+</td>
<td>+/−</td>
<td>Supports the presence of systemic inflammation</td>
</tr>
<tr>
<td>Elevated sedimentation rate</td>
<td>+</td>
<td>+/−</td>
<td>Supports the presence of systemic inflammation</td>
</tr>
<tr>
<td>Elevated serum creatinine</td>
<td>+/−</td>
<td>+/−</td>
<td>Raised serum creatinine, typically without hematuria or proteinuria on urinalysis, may indicate renal ischemia or infarction. Significant proteinuria or hematuria (especially red cell casts) would suggest glomerular disease, which is not a feature of PAN</td>
</tr>
<tr>
<td>Abnormal liver function tests</td>
<td>+</td>
<td></td>
<td>May suggest hepatitis, either from HBV or as a result of ischemic hepatitis from PAN affecting the hepatic arteries</td>
</tr>
<tr>
<td>Positive HBV serology</td>
<td>+</td>
<td></td>
<td>Seen in HBV PAN</td>
</tr>
<tr>
<td>Anemia</td>
<td>+</td>
<td></td>
<td>Due to chronic inflammation or from gastrointestinal blood loss</td>
</tr>
<tr>
<td>Positive ANCA</td>
<td>+</td>
<td>+</td>
<td>A positive ANCA would suggest an alternative type of vasculitis such as granulomatosis with polyangiitis or MPA</td>
</tr>
<tr>
<td>Elevated creatine kinase</td>
<td>+/−</td>
<td>+/−</td>
<td>Normal or mildly elevated, despite any muscle involvement</td>
</tr>
<tr>
<td>Blood cultures</td>
<td></td>
<td></td>
<td>To exclude endocarditis or other infective mimic of vasculitis</td>
</tr>
<tr>
<td>Positive HCV serology and cryoglobulins</td>
<td>+</td>
<td></td>
<td>HCV associated with a skin limited manifestation of PAN, but typically it is associated with a small vessel vasculitis related to cryoglobulinemia</td>
</tr>
<tr>
<td>Positive rheumatoid factor and ACPA</td>
<td>+</td>
<td></td>
<td>To rule out rheumatoid arthritis, especially in the context of a patient presenting predominantly with arthritis</td>
</tr>
<tr>
<td>Positive ANA and anti-dsDNA</td>
<td>+</td>
<td></td>
<td>In patients with clinical features consistent with SLE or other connective tissue disease</td>
</tr>
<tr>
<td>HIV positive</td>
<td>+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ACPA, anticitrullinated protein antibody; ANA, antinuclear antibody; ANCA, antineutrophil cytoplasm antibody; dsDNA, double-stranded deoxyribonucleic acid; ESR, erythrocyte sedimentation rate; HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; MPA, microscopic polyangiitis; SLE, systemic lupus erythematosus.

Polyarteritis Nodosa in Children

Pediatric cases of polyarteritis are also well described. In a review of 110 patients, with a mean age of 9 (male-to-female ratio equal), one-third were classified as having cutaneous PAN. Only 5% had classic PAN associated with hepatitis B surface antigen, whereas 80% were described as having MPA associated with ANCA and 57% had systemic PAN.

Microscopic Polyangiitis versus Polyarteritis Nodosa

A comparison of patients with PAN and MPA showed a difference in the skin manifestations of the 162 patients...
with MPA and 248 with PAN. There was a greater frequency of purpura (26%) in MPA compared with only 19% in PAN. By contrast, urticaria was more common in PAN (6% vs. 1.2%). The presence or absence of HBV infection—infuenced skin manifestations were less common (only 30%) in HBV-positive patients compared with HBV-negative patients (54%). Histologic examination of the skin, however, was identical in patients with either MPA or PAN. (See Chapter 89 for more details on MPA.)

Cutaneous Polyarteritis Nodosa

The term cutaneous PAN usually implies a separate entity from PAN that is limited to the skin, but there is some uncertainty as to whether these are actually just early cases of PAN and whether progression to PAN will occur. Cutaneous PAN is defined as a skin-limited form of PAN that is usually considered a separate clinical entity to systemic PAN. Pathologically the findings on skin biopsy are indistinguishable between the two. In a retrospective review of cutaneous PAN patients from a Japanese group, 22 patients with histologically proven cutaneous vasculitis were followed: 32% had a peripheral neuropathy, and 27% had myalgia, suggesting a need to revise the current criteria to differentiate between the two entities of cutaneous PAN and PAN. It was suggested that the two conditions were indeed separate, but cutaneous PAN is not confined to the skin. HCV infection has been associated with cutaneous PAN in one retrospective study of 16 patients, in which 5 individuals were found to have HCV infection.

Hepatitis B Virus Polyarteritis Nodosa

Hepatitis B vasculitis occurs in individuals with chronic hepatitis B antigenemia, most of whom have active liver disease. The manifestations vary considerably, from diffuse small vessel vasculitis predominantly in the skin to larger vessel lesions typical of PAN. Clinical symptoms may include the entire spectrum of vasculitic manifestations, from purpura and other rashes to abdominal pain, hypertension, renal disease, and stroke. HBV-associated PAN is an increasingly rare condition as a result of better immunization and widespread use of a live attenuated vaccine. However, the entity is not completely eradicated; cutaneous PAN and PAN are different clinical syndromes with pathologically similar findings. It may be that cutaneous PAN is a separate clinical syndrome that is only skin-limited but has all the hallmarks of PAN. The term cutaneous PAN is used to denote PAN that is limited to the skin, but there is some uncertainty as to whether these are actually just early cases of PAN and whether progression to PAN will occur. Cutaneous PAN is defined as a skin-limited form of PAN that is usually considered a separate clinical entity to systemic PAN. Pathologically the findings on skin biopsy are indistinguishable between the two.

Non–Hepatitis B Virus Polyarteritis Nodosa

The French Vasculitis Study Group has combined data on PAN and microscopic polyangiitis focusing more on the severity of disease manifestations rather than the type of disease in their studies of treatment or outcome. For example, Ribi and colleagues described the outcome of 124 patients with a new diagnosis of either PAN or MPA in whom the outcome was likely to be good (based on a lack of poor prognostic factors as measured by the Five Factor Score). All patients were treated with corticosteroids only followed by either azathioprine or cyclophosphamide if they relapsed. Ninety-eight of these patients achieved remission with steroids alone, but 46 of these patients relapsed. Primary treatment with steroids failed to control the disease in 26 patients, and 49 of the original 124 patients required additional immunosuppression. Similar rates of improvement were documented by using either azathioprine or cyclophosphamide. Therefore despite the absence of poor prognostic factors, only around 50% of patients could be managed with corticosteroids alone. A study of plasma exchange in 62 patients with a mixture of Churg–Strauss syndrome and what was described as PAN suggested that there was no clinical benefit from adding plasma exchange to standard treatment with pulse intermittent high-dose cyclophosphamide and steroid. This study from the mid-1990s describes patients with “PAN,” but most of these patients probably had MPA.

Treatment of Polyarteritis Nodosa

Uncontrolled vasculitis accounts for 58% to 73% of deaths occurring within the first year of diagnosis. Management of non-HBV PAN should be appropriate to the severity of the disease as defined by the five-factor prognostic score and includes aggressive immunosuppression with corticosteroids and cyclophosphamide in patients with Five Factor Score values of at least one. Patients with HBV-related PAN should be treated with high-dose corticosteroids followed by combination plasma exchange and antiviral agents to

Table 90-5 Pulsed Cyclophosphamide and Methylprednisolone for Non–Hepatitis B Virus Polyarteritis Nodosa

<table>
<thead>
<tr>
<th>Phase</th>
<th>Drug</th>
<th>Dose</th>
<th>Route</th>
<th>Frequency</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Induction</td>
<td>Cyclophosphamide</td>
<td>15 mg/kg per pulse</td>
<td>IV</td>
<td>Every 2 wk x 3, then every 3 wk x3-6</td>
<td>3-6 mo</td>
</tr>
<tr>
<td>Induction</td>
<td>Methylprednisolone</td>
<td>10 mg/kg</td>
<td>IV</td>
<td>Every 2 wk x 3, then every 3 wk x3-6</td>
<td>3-6 mo</td>
</tr>
<tr>
<td>Maintenance</td>
<td>Azathioprine</td>
<td>2 mg/kg/day</td>
<td>Oral</td>
<td>Daily</td>
<td>18-24 mo</td>
</tr>
<tr>
<td>Maintenance</td>
<td>Prednisone</td>
<td>7.5 mg/day</td>
<td>Oral</td>
<td>Daily</td>
<td>18-24 mo</td>
</tr>
</tbody>
</table>

reduce early mortality from vasculitis and late mortality from the consequences of chronic hepatitis.\textsuperscript{24}

**Outcome**

Earlier diagnosis and initiation of treatment has improved outcomes. In non-HBV PAN the 7-year survival rate is 79\%, compared with a 5-year survival rate of 72.5\% in HBV-related PAN. This is similar to the rates in vasculitis associated with ANCA.\textsuperscript{53} The relapse rate in HBV-related PAN is low (<10\%) compared with non-HBV-related PAN (19.4\% to 57\% relapse rates).\textsuperscript{24,49} Delayed diagnosis (>3 months) increases the risk of relapse but does not affect mortality risk.\textsuperscript{54} Seroconversion rates from hepatitis B e-antigen positive to hepatitis B e-antibody positive (i.e., eradication of active viral infection) is achieved in just under 50\% of patients treated with glucocorticoids, antiviral agents, plus limited plasmapheresis.\textsuperscript{24}

A prognostic tool designed for use in PAN is effectively used to distinguish patients with higher or lower risk of poor outcome. This simple clinical score (the Five Factor Score), assessed at diagnosis, consists of assessment of the following abnormalities: proteinuria (>1 g/day), elevated creatinine (>1.58 mg/dL), gastrointestinal involvement, central nervous system involvement, and cardiomyopathy.\textsuperscript{55} Allocating 1 point per item, there is a reduced survival for patients with higher scores (86\% survival for 0 points, 69\% survival for 1 point, 47\% survival for 2 or more points) at onset when followed up for 6 years.\textsuperscript{55} Mortality can also be predicted using the Birmingham Vasculitis Activity Score, a clinical index of disease activity.\textsuperscript{56,49} Older age at diagnosis has been shown to be an important adverse predictor of survival in the first year and after 5 years of follow-up.\textsuperscript{47,56}

**COGAN’S SYNDROME**

Cogan’s syndrome\textsuperscript{57} is a rare form of vasculitis. The median age at onset is 25 years.\textsuperscript{38,59} Patients typically present with red painful eyes and/or hearing loss concurrently or within 4 months in 75\% of patients. Most patients do not develop features of more widespread systemic vasculitis, with the exception of aortitis and aneurysm or aortic insufficiency, occurring in about 12\% of patients.\textsuperscript{60} In atypical Cogan’s syndrome, where the ocular manifestation is episcleritis, scleritis, iritis, uveitis, or chorioteratitis rather than interstitial keratitis, there is a worse prognosis and a higher frequency of aortic and other systemic manifestations.\textsuperscript{61} Eye symptoms usually consist of photophobia with red and irritable eyes. The audi vestibular symptoms are rapid in onset with partial or complete hearing loss (often bilateral), vertigo, and ataxia. Although the vertigo and ataxia may improve, the hearing loss is usually permanent. About 50\% of cases will have constitutional features such as weight loss, fever, lymphadenopathy, hepatosplenomegaly, and purpura. Aortic involvement is the most serious manifestation of Cogan’s syndrome and accounts for most deaths. Ophthalmic and audiovestibular features may precede aortic involvement by months to several years. Rarely, patients develop widespread vasculitis, with purpura and gangrene.\textsuperscript{62}

**Pathology**

A mixture of acute and chronic inflammation is present in large blood vessels, especially around the internal elastic lamina. The main findings are an infiltration of neutrophils, eosinophils, mononuclear cells, and fibrosis. Occasionally granulomas containing giant cells can be found in the lesions.

**Clinical Features**

The diagnosis of Cogan’s syndrome is largely based on clinical features, supported by the histologic abnormalities and exclusion of other conditions. Definitive serologic markers do not exist. Most patients have leukocytosis, anemia, thrombocytosis, and an elevated sedimentation rate during active phases of their disease. Patients with aortitis show aortic root dilation and aortic insufficiency on echocardiography or magnetic resonance imaging. Rare case reports demonstrate the presence of autoantibodies such as antiendothelial cell antibodies\textsuperscript{63} and anti-myeloperoxidase (MPO) antibodies,\textsuperscript{64} but their role in the disease is unknown.

**Treatment**

There is no clear evidence base because there are no prospective studies of treatment in this rare disease. Interstitial keratitis usually responds to topical corticosteroids. Most experts manage the acute audiovestibular symptoms with high doses of glucocorticoids, with a prompt response within a month (or none at all in resistant cases). Cytotoxic drugs,\textsuperscript{65} methotrexate,\textsuperscript{66} and cyclosporine\textsuperscript{67} have been used to treat the condition. The course is variable: For some patients there is only a single episode, and they are free of active disease thereafter. The more typical course is one of waxing and waning symptoms for months or years. Most patients sustain permanent hearing loss, and almost half lose hearing entirely; by contrast, few have permanent effects on vision.\textsuperscript{68} Cochlear implantation can successfully restore some hearing in these patients. Aortic valve replacement and surgical repair of aortic aneurysms may be required, but as in Takayasu’s arteritis,\textsuperscript{69} it would probably be best to ensure that there is no active disease at the time of surgery.

**BUERGER’S DISEASE**

Buerger’s disease, also known as thromboangiitis obliterans, is an inflammatory vaso-occlusive disease that predominantly affects the vascular supply to the lower limbs in young adult male tobacco smokers, although women and older adults\textsuperscript{69} can also be affected. The average age at onset is 35 years. Tobacco, especially cigarette smoking, is clearly important, but the mechanism of action remains unknown. There is no clear human leukocyte antigen association.\textsuperscript{70} Antibodies directed against collagen, elastin, and laminin have been reported in some patients.\textsuperscript{71,72} The presence of high levels of antiendothelial cell antibodies has been described in patients with active disease, but not in patients in remission.\textsuperscript{73} ANCA directed against MPO, lactoferrin, and elastase have also been associated with severe cases.\textsuperscript{74} More
recently a link to periodontal disease and anticiardioliopin antibody has been reported.\textsuperscript{73}

**Pathology**

In most cases Buerger's disease affects small arteries and veins in the distal extremities, but there are case reports of visceral arterial involvement including coronary, pulmonary, and mesenteric arteries. There is a transmural infiltration of polymorphonuclear leukocytes and lymphocytes with preservation of the internal elastic lamina.\textsuperscript{76} Thrombosis is prominent, and microabscesses can be found in the vessel wall and surrounding tissue. The infiltrating cells are enriched in CD3\textsuperscript{T} T cells; CD68\textsuperscript{+} macrophages and dendritic cells have been reported to be increased during disease activity.\textsuperscript{77}

**Clinical Features**

Buerger's disease typically begins with bilateral pain and ischemia in both lower extremities, although the upper extremities may be the site of initial symptoms. At first, the symptoms can be mild, with paresthesia or pain on exposure to cold. Most cases evolve rapidly, however, with increasing ischemic (claudicant) limb pain, digital cyanosis, splinter hemorrhages, and skin vesicles. Digital ulcers often occur, especially after minor trauma.\textsuperscript{78} The disease typically begins distally, with symptoms worse in the tips of the toes (and fingers), but it progresses to larger, more proximal vessels over several years. Usually Buerger’s disease does not cause proximal leg claudication. Superficial phlebitis occurs in about one-third of patients and may be the first symptom. Typical angiographic changes include multiple bilateral areas of narrowing or occlusion in the digital, palmar, plantar, ulnar, radial, tibial, and peroneal arteries. Small collateral vessels around the occlusion can appear to be corkscrew shaped. More proximal lesions resemble atherosclerotic occlusion. Although these findings are typical, none are pathognomonic; in the absence of pathologic confirmation, the differential diagnosis includes premature atherosclerosis; hyperviscosity syndrome; scleroderma and other rheumatic diseases; Takayasu's arteritis; embolic disease including cholesterol emboli and atrial myxomas; ergot toxicity; and thoracic outlet syndrome.

**Treatment**

The most important treatment is to stop smoking. In patients who are unable to comply with this, nicotine substitutes may be used at least in the short term.\textsuperscript{79} Affected limbs must be protected from trauma and cold. Ulcers and cellulitis often require antibiotics and local wound management. Anecdotial use of calcium channel blockers and pentoxifylline reportedly has been beneficial in some, but not all, patients. Sympathectomy seems to provide little or no long-term benefit and is not usually recommended.\textsuperscript{80} Intravenous prostacyclin is effective and superior to aspirin\textsuperscript{81} and also superior to lumbar sympathectomy for rest pain or ischemic ulcers\textsuperscript{82} on the basis of two large randomized controlled trials. Recent approaches using bone marrow–derived mesenchymal stem cells, basic fibroblast growth factor, and granulocyte colony-stimulating factor have been effective in treatment of leg ulcers and ischemia in small open-label studies.\textsuperscript{82-85} A radical surgical approach using omental grafting is reported in large open-label surgical studies to offer more than 90% success in improving the severe ischemic complications.\textsuperscript{86}

In patients who continue to smoke, about half require amputation, often multiple times as more proximal vessels are involved.\textsuperscript{87} If they stop smoking, most patients will stop getting worse, but a few still require amputation. However, the ischemic limb may be a source of pain and ulceration for many years.

**SUSAC’S SYNDROME**

This is another rare condition affecting a wide age range from children to older adults.\textsuperscript{88,89} The etiology is unknown. The presentation is usually with sudden onset of sensorineural hearing loss in the context of encephalopathy and branch retinal artery occlusion.\textsuperscript{88} The differential diagnosis would include Cogan’s syndrome and GPA. The pathogenetic mechanisms suggest an endotheliopathy rather than a true vasculitis. Nevertheless, treatment with immunosuppressive agents including glucocorticoids and cytotoxic agents is reported to improve some patients, but there are no controlled trials.\textsuperscript{89} The diagnosis rests on the combination of clinical features and exclusion of other causes. The outcome is variable, but most patients recover some hearing.\textsuperscript{90}

**VIRUS-INDUCED VASCULITIS**

The most common virus-associated vasculitis today is HCV-associated cryoglobulinemia (see Chapter 91). The majority of patients with cryoglobulinemia have concomitant hepatitis C infection (see Chapter 91). Treatment with immunosuppressive drugs has been only moderately successful; many patients die as a result of vasculitis or liver disease. Saadoun and colleagues\textsuperscript{92} describe their cohort of 72 patients treated with combination antiviral therapy. Patients responded well to a combination of interferon-α and ribavirin; however, 40% required glucocorticoids, around 12% had plasmapheresis, and 6% received immunosuppressive drugs for severe complications. The mortality was just over 10%. The vasculitis resulting from human immunodeficiency virus infection is discussed in Chapter 113. Although the most common rheumatic manifestation of parvovirus B19 infection is arthritis (see Chapter 114), some children with chronic parvovirus B19 infection have been reported to develop a vasculitis similar to PAN. It is not clear, given the high incidence of the infection in children, whether the presence of the virus is coincidental or causative in the few cases of vasculitis described.\textsuperscript{93}

**References**

KEY POINTS

Vasculitis mediated by immune complexes (ICs) includes a heterogeneous group of disorders linked by inefficient or dysregulated clearance of ICs.

The most common types of IC-mediated vasculitis are hypersensitivity vasculitis, Henoch-Schönlein purpura (HSP), and mixed cryoglobulinemia. Rarer forms of this condition include hypocomplementemic urticarial vasculitis and erythema elevatum diutinum.

Connective tissue disorders such as systemic lupus erythematosus, Sjögren’s syndrome, and rheumatoid arthritis can be associated with IC-mediated vasculitis.

Cutaneous involvement of small blood vessels is the most prominent feature in the majority of cases, but extracutaneous involvement occurs in some forms.

The classic cutaneous finding in small vessel vasculitis is palpable purpura, but a variety of other skin lesions may be found including pustules, vesicles, urticaria, and small ulcerations.

Direct immunofluorescence studies of involved blood vessels demonstrate characteristic types and patterns of immunoglobulin (Ig) and complement deposition.

Hypersensitivity vasculitis usually results from a reaction to a medication or an infection.

HSP is associated with purpura, arthritis, glomerulonephritis, and colicky abdominal pain. IgA deposition is found within blood vessel walls.

Cryoglobulinemic vasculitis is most often associated with long-standing hepatitis C virus infection. The term mixed cryoglobulinemia is sometimes used for this disorder because the immunoreactants involved in the disease include both IgG and IgM.

The formation of ICs in the Arthus reaction initiates complement activation and an influx of inflammatory cells, followed by thrombus formation and hemorrhagic infarction in the areas of most intense inflammation. ICs, formed by the combination of antibody and antigen, are continuously created (and usually cleared swiftly and efficiently) by the reticuloendothelial system as a means of neutralizing foreign antigens. Under some circumstances, however, ICs escape clearance and become deposited within joints, blood vessels, and other tissues, inciting inflammation and causing disease. ICs deposited in the blood vessel walls lead to vasculitis. Similarly, those deposited within small blood vessels of the kidney—the glomeruli—cause glomerulonephritis.

Immunogenicity

The fate of formed ICs is governed by several major factors including antigen load, antibody response, efficiency of the reticuloendothelial system, physical properties of the blood vessels (including flow dynamics and previous endothelial

This chapter focuses on forms of small vessel vasculitis that are mediated by IC deposition. These disorders include hypersensitivity vasculitis, Henoch-Schönlein purpura (HSP), mixed cryoglobulinemia, hypocomplementemic urticarial vasculitis, and erythema elevatum diutinum. In addition, forms of vasculitis associated with connective tissue diseases, particularly systemic lupus erythematosus (SLE), Sjögren’s syndrome, and rheumatoid vasculitis, are discussed briefly. Anti–glomerular basement membrane disease and the pauci-immune forms of vasculitis such as those associated with antineutrophil cytoplasmic antibodies, are discussed elsewhere (see Chapter 89). Throughout this chapter, the terms vasculitis and angiitis are used interchangeably when referring to inflammation involving small blood vessels (capillaries, venules, arterioles).

Because all forms of IC-mediated vasculitis share certain elements of pathogenesis, have many cutaneous findings in common, and have overlapping differential diagnoses, these aspects of the disorders are considered together. The epidemiology, cause, distinctive pathophysiologic mechanisms, unique clinical features, and approaches to treatment are discussed separately for each condition. Treatments are also summarized in Table 91-1.

PATHOGENESIS

Arthus Reaction

The Arthus reaction, described after the injection of horse serum into rabbits, forms the basis of our understanding of IC-mediated diseases. The formation of ICs in the Arthus reaction initiates complement activation and an influx of inflammatory cells, followed by thrombus formation and hemorrhagic infarction in the areas of most intense inflammation. ICs, formed by the combination of antibody and antigen, are continuously created (and usually cleared swiftly and efficiently) by the reticuloendothelial system as a means of neutralizing foreign antigens. Under some circumstances, however, ICs escape clearance and become deposited within joints, blood vessels, and other tissues, inciting inflammation and causing disease. ICs deposited in the blood vessel walls lead to vasculitis. Similarly, those deposited within small blood vessels of the kidney—the glomeruli—cause glomerulonephritis.

Immunogenicity

The fate of formed ICs is governed by several major factors including antigen load, antibody response, efficiency of the reticuloendothelial system, physical properties of the blood vessels (including flow dynamics and previous endothelial
damage), and solubility of the ICs themselves. The ratio of antibody to antigen determines the solubility of ICs. Large ICs, formed when antibody and antigen are present in approximately equal proportions, are identified and removed easily by the reticuloendothelial system. In contrast, small ICs are formed in conditions of antibody excess. Small ICs remain in the serum and do not elicit an immune response within tissues. However, when there is a slight antigen excess, ICs precipitate from the serum and become trapped within certain vascular beds. Following the deposition of ICs in tissue, a cascade of pathologic events ensues: complement fixation, neutrophil recruitment, local inflammation, lysosomal release, oxygen free radical generation, and tissue injury.

## CUTANEOUS MANIFESTATIONS

Small blood vessels generally include capillaries, postcapillary venules, and nonmuscular arterioles—vessels that are typically less than 50 μm in diameter. These are found principally within the superficial papillary dermis (Figure 91-1). Medium-sized blood vessels, those between 50 and 150 μm in diameter, contain muscular walls and are located principally in the deep reticular dermis, near the junction of the dermis and subcutaneous tissues. Vessels larger than 150 μm in diameter are not commonly found in the skin.

Figure 91-1, which demonstrates the location and size of blood vessels involved in various types of cutaneous vasculitis, illustrates the types of blood vessels affected by several forms of IC-mediated disease. A blood vessel’s size correlates closely with its depth in the skin layers: The larger the vessel, the deeper its location. Although telltale signs of vasculitis may be evident on inspection of the skin’s surface, the epidermis is avascular. Therefore the pathologic findings in cutaneous vasculitides lie within the dermis and subcutaneous tissues.

Palpable purpura, synonymous with small vessel vasculitis, is the most common cutaneous finding in IC-mediated vasculitis (Figure 91-2). Purpuric lesions result from the extravasation of erythrocytes through damaged blood vessel walls into tissue. Many other skin manifestations are possible in these conditions including vesicles, pustules, urticaria, superficial ulcerations, nonpalpable lesions (macules and patches), and splinter hemorrhages (Figure 91-3). These lesions frequently occur in combination, and careful examination usually reveals a purpuric component. Purpuric lesions do not blanch when pressure is applied to the skin. Following resolution, purpuric lesions may leave postinflammatory hyperpigmentation, particularly if repeated bouts occur (see Figure 91-3F).

In IC-mediated vasculitis, purpuric lesions are usually distributed in a symmetric fashion over dependent regions of the body, particularly the lower legs, because of the increased hydrostatic pressure in these areas. Purpuric lesions are not always palpable to the touch, and the

### Table 91-1 Potential Treatment Approaches for Different Forms of Immune Complex–Mediated Vasculitis

<table>
<thead>
<tr>
<th>Disease</th>
<th>Preferred Treatment Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypersensitivity vasculitis</td>
<td>Removal of the offending agent</td>
</tr>
<tr>
<td>Henoch-Schönlein purpura</td>
<td>Brief (2- to 4-wk) course of glucocorticoids for severe cases</td>
</tr>
<tr>
<td></td>
<td>Moderate glucocorticoid doses (prednisone 20-40 mg/day) are of variable utility but may be</td>
</tr>
<tr>
<td></td>
<td>used empirically in patients with disabling symptoms</td>
</tr>
<tr>
<td></td>
<td>Pulse glucocorticoids (e.g., 1 g methylprednisolone/day) employed with anecdotal success for</td>
</tr>
<tr>
<td></td>
<td>refractory purpura when moderate glucocorticoid doses have failed</td>
</tr>
<tr>
<td></td>
<td>Refractory glomerulonephritis may require high-dose glucocorticoids, azathioprine, mycophenolate mofetil, or cyclophoshamide</td>
</tr>
<tr>
<td>Cryoglobulinemia</td>
<td>Combined antiviral therapies and B cell depletion strategies synergistic in treating mixed</td>
</tr>
<tr>
<td></td>
<td>cryoglobulinemia associated with hepatitis C</td>
</tr>
<tr>
<td>Hypocomplementemic urticarial vasculitis</td>
<td>Rituximab possibly effective in treating idiopathic mixed cryoglobulinemia</td>
</tr>
<tr>
<td></td>
<td>Higher doses of glucocorticoids in patients with serious internal organ involvement or</td>
</tr>
<tr>
<td></td>
<td>ulcerative skin lesions</td>
</tr>
<tr>
<td>Erythema elevatum diutinum</td>
<td>Anecdotal reports of success with tumor necrosis factor inhibition</td>
</tr>
<tr>
<td>Connective tissue disease</td>
<td>Dapsone or sulfapyridine</td>
</tr>
<tr>
<td>Rheumatoid vasculitis</td>
<td>Hydroxychloroquine, low-dose prednisone (5-20 mg/day), azathioprine</td>
</tr>
<tr>
<td></td>
<td>High-dose glucocorticoids plus cyclophosphamide for widespread necrotizing vasculitis</td>
</tr>
<tr>
<td></td>
<td>Anecdotal experience suggests that tumor necrosis factor inhibition or rituximab might be</td>
</tr>
<tr>
<td></td>
<td>effective in combination with glucocorticoids</td>
</tr>
</tbody>
</table>

![Figure 91-1](image1.png) Size of the blood vessels involved in forms of cutaneous vasculitis. The types of vasculitis with an immune complex–mediated pathogenesis include Henoch-Schönlein purpura (HSP), cutaneous leukocytoclastic angiitis (CLA), mixed cryoglobulinemia (CRYO), and connective tissue disease/rheumatoid vasculitis (CTD/RV). ANCA, antineutrophil cytoplasmic antibody; PAN, polyarteritis nodosa.
PATHOLOGIC FEATURES

Full pathologic assessment of cutaneous vasculitis involves examination of a skin biopsy specimen by both light microscopy and direct immunofluorescence (DIF). DIF is a particularly critical procedure in the evaluation of small vessel vasculitides. DIF studies must be planned at the time the biopsy is performed because they require a fresh skin biopsy sample.

**Light Microscopy**

Figure 91-4A displays the light microscopy findings of cutaneous vasculitis. The optimal time for skin biopsy is 24 to 48 hours after the appearance of a lesion. Biopsies should be obtained from a nonulcerated site. For ulcerated lesions—usually more of an issue with medium vessel vasculitides—biopsies should be taken from the ulcer’s edge. The cellular infiltrates in cutaneous vasculitis are usually made up of a combination of neutrophils and lymphocytes, but most cases demonstrate a predominance of one cell type or the other. Lymphocyte-rich infiltrates may be seen in specimens taken from either new (<12 hours) or old (>48 hours) lesions, regardless of the underlying type of vasculitis. Even in connective tissue disorders such as Sjögren’s syndrome, the typical finding is a leukocytoclastic vasculitis rather than a lymphocytic vasculitis.²
CHAPTER 91 | IMMUNE COMPLEX–MEDIATED SMALL VESSEL VASCULITIS

ICs, and vasculitis mimickers that involve small blood vessels. A diagnostic algorithm that includes the critical laboratory and radiographic tests is shown in Figure 91-5.

CLINICAL SYNDROMES

Hypersensitivity Vasculitis (Cutaneous Leukocytoclastic Angiitis)

The term hypersensitivity vasculitis (see Chapter 87) generally refers to an IC-mediated small vessel vasculitis of the skin that spares internal organs and usually follows drug exposures or infections. The Chapel Hill consensus conference recommended eliminating the term hypersensitivity vasculitis in favor of cutaneous leukocytoclastic angiitis because of the disorder’s usual confinement to the skin and its predominant cell type, the neutrophil. However, hypersensitivity vasculitis remains firmly embedded in the medical literature. The disease is characterized pathologically by IC deposition in capillaries, postcapillary venules, and arterioles. A similar illness—serum sickness—is a systemic illness that includes rash and prominent arthralgias or arthritis; it occurs 1 to 2 weeks after exposure to a drug or foreign antigen.

In 1990 the American College of Rheumatology (ACR) performed a study designed to identify features that distinguished one form of vasculitis from others. The resulting ACR classification criteria for hypersensitivity vasculitis are shown in Table 91-3. The key historical element in evaluating a patient with possible hypersensitivity vasculitis is identifying exposures that may have triggered the reaction. However, in approximately half of all patients with presumed hypersensitivity vasculitis, no inciting agent can be identified.

DIRECT IMMUNOFLOUORESCENCE

Although the diagnosis of cutaneous vasculitis rests on routine histology, the features revealed by hematoxylin and eosin stains do not distinguish between pauci-immune and IC-mediated disorders. DIF studies complement the histologic information, provide the only way of diagnosing HSP with certainty, and yield important clues regarding the nature of the underlying disease. The performance of separate biopsies for histologic and DIF analyses is recommended if sufficient lesions exist. With DIF studies, frozen sections are incubated with fluorescein-labeled anti-human immunoglobulin (Ig) G, IgM, IgA, and C3. The staining patterns of these immunoreactants may provide insight into not only the diagnosis but also the pathophysiology of certain conditions. Figure 91-4B displays the typical DIF findings in a skin lesion from a patient with IC-mediated vasculitis.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of IC-mediated small vessel vasculitis is shown in Table 91-2. There are three main groups of disorders in the differential diagnosis of IC-mediated small vessel vasculitis: other forms of IC-mediated disorders, forms of small vessel vasculitis that are not mediated through ICs, and vasculitis mimickers that involve small blood vessels. A diagnostic algorithm that includes the critical laboratory and radiographic tests is shown in Figure 91-5.
A long list of medications, infections, and other exposures may lead to the syndrome of hypersensitivity vasculitis. The typical history for a drug-induced hypersensitivity vasculitis is the occurrence of clinical symptoms approximately 7 to 14 days after starting a new medication. Although virtually any medication can be associated with the induction of a hypersensitivity vasculitis, antibiotics (particularly penicillins and cephalosporins) are the most common offenders. Other common culprits are diuretics and antihypertensive agents, but when confronted with a patient with a potential drug-induced vasculitis it is critical to examine and potentially discontinue any medication added within a recent timeframe. Drug-induced hypersensitivity vasculitides begin to resolve within days of removal of the offending agent.

Removal of the inciting agent is the most critical therapy for hypersensitivity vasculitis when the likely agent can be identified. In patients who have been exposed to multiple medications, determining the inciting agent may be difficult and may require the withdrawal of multiple agents simultaneously until the syndrome clears, typically in 1 to 2 weeks.

The prognosis for patients with hypersensitivity vasculitis depends on the inciting cause. Treatment with glucocorticoids is reserved for patients with extensive disease and can usually be discontinued within several weeks. Patients who experience repeated disease flares may need low-dose glucocorticoids to prevent recurrences.

Table 91-2  Differential Diagnosis of Immune Complex–Mediated Vasculitis

<table>
<thead>
<tr>
<th>Immune Complex–Mediated Vasculitides</th>
<th>Pauci-Immune Vasculitides</th>
<th>Miscellaneous Small Vessel Vasculitides</th>
<th>Vasculitis Mimickers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypersensitivity vasculitis</td>
<td>Granulomatosis with polyangitis</td>
<td>Behçet’s disease</td>
<td>Hemorrhage</td>
</tr>
<tr>
<td>Henoch-Schönlein purpura</td>
<td>Churg-Strauss syndrome</td>
<td>Malignancy associated</td>
<td>Pigmented purpuric dermatoses</td>
</tr>
<tr>
<td>Mixed cryoglobulinemia</td>
<td>Microscopic polyangitis</td>
<td>Infection</td>
<td>Scurvy</td>
</tr>
<tr>
<td>Urticular vasculitis</td>
<td></td>
<td>Inflammatory bowel disease</td>
<td>Immune thrombocytopenic purpura</td>
</tr>
<tr>
<td>Erythema elevatum diutinum</td>
<td></td>
<td></td>
<td>Thrombosis</td>
</tr>
<tr>
<td>Connective tissue disease, rheumatoid vasculitis</td>
<td></td>
<td></td>
<td>Antiphospholipid syndrome</td>
</tr>
</tbody>
</table>

Table 91-3  American College of Rheumatology 1990 Criteria* for the Classification of Hypersensitivity Vasculitis

<table>
<thead>
<tr>
<th>Age ≥16 yr</th>
<th>Use of a possible offending medication in temporal relation to symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palpable purpura</td>
<td>Maculopapular rash</td>
</tr>
<tr>
<td>Maculopapular rash</td>
<td>Biopsy of a skin lesion showing neutrophils around an arteriole or venule</td>
</tr>
</tbody>
</table>

*The presence of three or more criteria has a sensitivity of 71% and specificity of 84% for the diagnosis of hypersensitivity vasculitis.

**Henoch-Schönlein Purpura**

HSP is an IC-mediated form of small vessel vasculitis that is strongly associated with IgA deposition within blood vessel walls. Many cases of HSP occur after upper respiratory tract infections. Multiple bacterial, viral, and other infectious agents have been suggested as the cause of HSP, but the true cause remains unknown. The 1990 ACR criteria for the classification of HSP are shown in Table 91-4.8

A major risk factor for HSP (as well as IgA nephropathy, which has renal findings identical to those of HSP) is aberrant glycosylation of O-linked glycans in the hinge region of a fraction of IgA1 molecules.9 Rather than terminating with galactose, the aberrant galactose-deficient O-glycans end with N-acetylgalactosamine (GalNAc) or sialylated GalNAc. The terminal GalNAc moiety on the aberrantly glycosylated IgA1 may in turn be recognized by antiglycan antibodies, leading to the formation of circulating immune complexes that deposit in the skin, joints, kidneys, and other organs. However, a high serum Gd-IgA1 level is not sufficient for the development of clinical symptoms. A “second hit” in the form of another environmental or inherited risk factor is required to produce HSP. This probably explains why multiple types of infections and many different drugs (e.g., antibiotics) have been linked etiologically with HSP.

The hallmarks of HSP include an upper respiratory tract infection followed by a syndrome characterized by a purpuric rash, arthralgias, abdominal pain, and renal disease. HSP is usually viewed as a disease of childhood, and the majority of cases affect children younger than 5 years. However, adults can also be affected by HSP and have a greater tendency toward a prolonged disease course (with recurrent bouts of purpura) than do children.10 Colicky abdominal pain, presumably secondary to gastrointestinal vasculitis, is a common characteristic of HSP and frequently occurs within a week after the onset of rash. Sometimes the gastrointestinal symptoms of HSP precede the onset of purpura, leading to a diagnostic quandary and occasionally to exploratory surgery. Endoscopy may demonstrate purpura in the upper or lower intestinal tract. Mild glomerulonephritis is common and generally self-limited, although some patients develop end-stage renal disease.

In children with mild manifestations, the clinical history alone may be sufficient to confirm the diagnosis. In more serious cases (e.g., in the presence of renal involvement) or when there is sufficient doubt about the diagnosis, biopsy of an involved organ is essential. Unlike in other forms of IC-mediated disease, however, DIF reveals florid IgA deposition. In the proper clinical setting, this finding is diagnostic of HSP. Other forms of small vessel vasculitis may have small quantities of IgA within blood vessels, but IgA is not the predominant immunoreactant in such cases.

In mild cases of HSP, no specific therapy is necessary. Even for patients with glomerulonephritis, it has been difficult to demonstrate that treatment with glucocorticoids or immunosuppressive agents significantly alters the outcome. Despite this, it is prudent to treat aggressive renal involvement with an immunosuppressive regimen including high-dose glucocorticoids and another immunosuppressive agent such as cyclophosphamide, azathioprine, or mycophenolate mofetil, depending on disease severity.

Recurrences of skin disease, often consisting of multiple episodes occurring over many months, are not unusual. However, even in patients with recurrent disease, the rule is for the disorder to subside and to resolve completely over a few months to a year. In a minority of patients, some evidence of permanent renal damage persists in the form of proteinuria and hematuria. Less than 5% of patients develop renal failure as a result of HSP.

### Cryoglobulinemic Vasculitis

Cryoglobulins are immunoglobulins characterized by a tendency to precipitate from serum under conditions of cold.11 Such proteins, detectable to a varying degree in a wide array of inflammatory conditions, do not always cause disease. In some patients, however, cryoglobulins bind to circulating antigen (e.g., portions of the hepatitis C virion), deposit in the walls of small and medium-sized blood vessels, and activate complement, leading to cryoglobulinemic vasculitis.

In contrast to most other forms of IC-mediated vasculitis, cryoglobulinemia tends to involve small- and medium-sized blood vessels. Thus the syndrome of cryoglobulinemic vasculitis can be associated with the development of large cutaneous ulcers, digital ischemia, and livedo racemosa—which they are associated (Table 91-5).1

Three major types of cryoglobulinemia are recognized, defined by the specific kinds of immunoglobulins with which they are associated (Table 91-6). Type I, characterized by a monoclonal gammopathy (generally IgG or IgM), differs substantially from types II and III in its clinical presentation and disease associations. Type I cryoglobulinemia, associated with Waldenström’s macroglobulinemia or, less frequently, multiple myeloma, is more likely to cause syndromes related to hyperviscosity (dizziness, confusion, headache, and stroke) than necrotizing vasculitis. In contrast to the monoclonal nature of type I cryoglobulinemia, types II and III are known as mixed cryoglobulinemias because they are composed of both IgG and IgM. In type II cryoglobulinemia, more than 90% of cases are caused by hepatitis C infection and the cryoproteins consist of monoclonal IgM and polyclonal IgG. Cases of type II cryoglobulinemia not associated with hepatitis C are sometimes termed mixed essential cryoglobulinemia because their cause is not known. Type III cryoglobulinemia, typically associated with polyclonal IgG and polyclonal IgM, is associated with many

---

**Table 91-4**  American College of Rheumatology 1990 Criteria* for the Classification of Henoch-Schönlein Purpura

<table>
<thead>
<tr>
<th>Palpable purpura</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at onset &lt;20 yr</td>
</tr>
<tr>
<td>Bowel angina</td>
</tr>
<tr>
<td>Vessel wall granulocytes on biopsy</td>
</tr>
</tbody>
</table>

*The presence of two criteria identified Henoch-Schönlein purpura with a sensitivity of 87% and a specificity of 88% in a group of individuals with forms of systemic vasculitis.

forms of chronic inflammation including infection and autoimmune disease.

Type II and III cryoglobulinemias often present with a triad of signs and symptoms: purpura, arthralgias, and myalgias. The purpura may be extensive and confluent (Figure 91-6), sometimes involving the trunk, upper extremities, and even the face; in most cases, however, the rash is confined to the lower extremities. Other organ systems commonly involved in mixed cryoglobulinemia are the kidneys and peripheral nerves. Mixed cryoglobulinemia may cause membranoproliferative glomerulonephritis that resembles lupus nephritis histopathologically. It may also cause a vasculitic neuropathy, usually with sensory symptoms predominating over motor symptoms. Finally, in rare cases, cryoglobulinemia is associated with alveolar hemorrhage.

Skin biopsy is the most straightforward method of confirming the diagnosis. Light microscopy of purpuric lesions demonstrates leukocytoclastic vasculitis. In addition, DIF studies reveal various types of immunoglobulin and complement deposition, depending on the type. In type II cryoglobulinemia, for example, DIF reveals IgG and IgM deposition, as well as complement components. Serologic testing may also yield clues to the presence of mixed cryoglobulinemia. To assay for serum cryoglobulins, the blood is collected in a prewarmed apparatus, allowed to clot at 37°C before processing, and then refrigerated at 4°C for several days. The percentage of the serum occupied by the cryoprecipitate is referred to as the “cryocrit.” The difficulties involved in performing cryoglobulin assays often lead to false-negative results. Nonspecific serologic testing may also implicate mixed cryoglobulinemia. As noted, cryoglobulins detected are not always associated with disease.

A strong clue is the presence of an extremely low level of C4, reduced out of proportion to C3. In addition, the monoclonal component of type II cryoglobulins almost invariably has rheumatoid factor activity (i.e., binds to the Fc portion of IgG). Thus essentially all patients with type II cryoglobulinemia have high titers of rheumatoid factor. As markers of clinical disease activity, C4 levels, rheumatoid factor titers, and cryocrits all fare poorly, often remaining abnormal in the face of clinically improved disease.

Treatment of cryoglobulinemia has seen substantial progress in recent years. Until recently, antiviral therapies with the combination of interferon-α and ribavirin were believed to be the optimal treatment for mixed cryoglobulinemia. Dual strategies that combine antiviral therapies and B cell depletion approaches with rituximab appear to be synergistic in treating this condition and leading to long-term treatment responses.12 The rationale behind rituximab use in cryoglobulinemia is that peripheral B lymphocyte depletion will lead to a reduction in plasma cells that produce cryoglobulins. Studies of these treatment modalities have suggested that disease relapses are associated with the absence of virologic control and peripheral B cell recovery, implying

Table 91-6  Types of Cryoglobulins

<table>
<thead>
<tr>
<th>Cryoglobulin</th>
<th>RF Positivity</th>
<th>Monoclonality</th>
<th>Associated Diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td>No</td>
<td>Yes (IgG or IgM)</td>
<td>Hematopoietic malignancy (multiple myeloma, Waldenström’s macroglobulinemia)</td>
</tr>
<tr>
<td>Type II</td>
<td>Yes</td>
<td>Yes (polyclonal IgG, monoclonal IgM)</td>
<td>Hepatitis C, Other infection, Sjögren’s syndrome, SLE</td>
</tr>
<tr>
<td>Type III</td>
<td>Yes</td>
<td>No (polyclonal IgG and IgM)</td>
<td>Hepatitis C, Other infection, Sjögren’s syndrome, SLE</td>
</tr>
</tbody>
</table>

Ig, immunoglobulin; RF, rheumatoid factor; SLE, systemic lupus erythematosus.
the need to combine the two treatment strategies. The optimal timing of antiviral and B cell depletion strategies is not clear. However, one reasonable approach is to initiate antiviral strategies first and then to employ rituximab within several weeks. Patients who present with overwhelming illness such as the unusual patient with alveolar hemorrhage or with symptoms of hyperviscosity are candidates for plasma exchange, with the goal of removing the pathogenic immune complexes as quickly as possible. One approach is to perform plasma exchange every other day with concomitant immunsuppressive therapy or B cell depletion for a total of seven exchanges or until there is sufficient clinical improvement. The prognosis of patients with cryoglobulinemia generally depends on the underlying cause. The outcome of type I cryoglobulinemia relates closely to the success in treating the cause. Type II or III cryoglobulinemia secondary to hepatitis C can be treated effectively if the viral infection is responsive to therapy. If patients do not tolerate antiviral therapy well or if the treatment is ineffective, they may require low to moderate doses of prednisone to control the disease.

**Hypocomplementemic Urticarial Vasculitis**

In contrast to common urticaria, the lesions of urticarial vasculitis (UV) last more than 48 hours, do not blanch when pressure is applied to the skin, and may leave postinflammatory hyperpigmentation. Unlike common urticaria, the lesions of UV are frequently associated with moderate pain, burning, and tenderness in addition to pruritus. Whereas common urticaria typically resolves completely within 24 to 48 hours, the lesions of UV may take days to resolve completely, often leaving residual hyperpigmentation; they may worsen without therapy.

Three different syndromes of UV are recognized: normocomplementemic UV, hypocomplementemic UV, and the hypocomplementemic urticarial vasculitis syndrome (HUVS). Normocomplementemic UV is typically a self-limited subset of hypersensitivity vasculitis. In chronic cases, normocomplementemic UV must be distinguished carefully from neutrophilic urticaria, a persistent form of urticaria not associated with vasculitis. In contrast, hypocomplementemic UV is more likely to be a chronic disorder that has certain overlapping features with SLE: low serum complements, autoantibodies, and an interface dermatitis characterized by immunoreactant deposition (complement and immunoglobulins) at the dermal-epidermal junction in a pattern essentially identical to the lupus band test. Finally, HUVS is a severe form of the disease associated with extra-cutaneous disease and an array of organ system findings atypical of SLE. For example, HUVS may be associated with uveitis, chronic obstructive pulmonary disease (COPD), and angioedema.

The skin lesions in UV tend to be centripetal, favoring the trunk and proximal extremities more than dependent regions. The lesions are painful and associated with a burning sensation rather than the pruritus of common urticaria. Biopsy of an urticarial wheal in UV demonstrates evidence of leukocytoclastic vasculitis including injury to the endothelial cells of the postcapillary venules, erythrocyte extravasation, leukocytoclasia, fibrin deposition, and a perivascular neutrophilic (or, less commonly, lymphocytic) infiltrate. DIF demonstrates IC deposition around blood vessels in the superficial dermis and a striking deposition of immunoglobulins and complement along the dermal-epidermal junction. In the proper setting, these findings (interface dermatitis and immunoreactant deposition within blood vessels) are diagnostic of hypocomplementemic UV. HUVS, in contrast, is a clinical diagnosis based on the presence of UV and the occurrence of typical features in extra-cutaneous organ systems.

Some cases of hypocomplementemic UV respond to therapies commonly used for the treatment of SLE including low-dose prednisone, hydroxchloroquine, dapsone, or other immunomodulatory agents. Serious cases of HUVS, particularly those presenting with glomerulonephritis or other forms of serious organ involvement, may require high doses of glucocorticoids or biologic agents such as inhibitors of tumor necrosis factor. Both COPD and cardiac valve abnormalities are associated with HUVS and may require specific treatment as well.

The prognosis of UV is linked to the disorder with which it is associated. SLE, COPD, angioedema, and valvular abnormalities are all known to occur in association with this disorder and may strongly influence both quality and quantity of life.

**Erythema Elevatum Diutinum**

Erythema elevatum diutinum (EED) is a rare, distinctive form of leukocytoclastic vasculitis limited to the skin. The disorder is distinctive because of the unusual distribution of skin lesions (found symmetrically over the extensor surfaces of joints) and the prompt response to sulfone medications. The cutaneous findings are typical of any small vessel vasculitis, with a predominance of papules, plaques, and nodules. Early lesions are often pink or yellowish and then become red or purple (Figure 91-7). The natural history of untreated lesions is to persist for years, becoming doughy or hard with time. The lesions have a predilection for the skin overlying the small joints of the hands and the knees; they can also affect the buttocks. The trunk is generally spared.

The principal histopathologic findings of EED are leukocytoclastic vasculitis with fibrinoid necrosis. Although an IC basis is suspected in this disorder, DIF studies are not distinctive. EED has been associated with various connective tissue diseases (CTDs), rheumatoid arthritis, other forms of vasculitis such as GPA, human immunodeficiency virus infections, and paraproteinemia (particularly IgA). EED typically responds promptly to dapsone or sulfapyridine, but chronic therapy may be required because skin lesions recur after cessation of treatment.

**Connective Tissue Disease–Associated Vasculitis**

Vasculitis rarely occurs in CTDs without overt manifestations of the underlying disorder. The forms of CTD typically complicated by vasculitis include those related to SLE: lupus itself, mixed CTD, Sjögren’s syndrome, and overlap CTD. Although vasculitis clearly occurs in some CTD settings, it is commonly overdiagnosed to explain perplexing disease features in patients with known rheumatic illnesses. For example, neuropsychiatric SLE is generally not caused...
by a true vasculitis but rather by other mechanisms that remain poorly defined. Whenever possible, the clinical hypothesis of vasculitis should be confirmed by biopsy.

Cutaneous vasculitis in CTDs is associated almost invariably with hypocomplementemia and high titers of antinuclear antibodies (ANAs). DIF examination of skin lesions shows granular IgG and C3 deposition in and around dermal vessels, with or without IgM, reflecting the contribution of ICs to disease pathogenesis. The phenomenon of the “in vivo ANA” is also observed in keratinocytes and dermal cells in DIF studies (Figure 91-8).

Vasculitis in patients with SLE-related disorders is more likely than other forms of vasculitis to be associated with a lymphocytic predominance. One variant of CTD-associated cutaneous vasculitis, the so-called benign hypergammaglobulinemia of Waldenström, is usually a true lymphocytic vasculitis. Patients with this disorder invariably have anti-Ro antibodies, and many have subclinical Sjögren’s syndrome. Lymphocytic vasculitis typically demonstrates less disruption of blood vessel architecture than does leukocytoclastic vasculitis, perhaps because lymphocytes contain fewer of the destructive enzymes found within neutrophil granules. Fibrinoid necrosis, for example, is rare in lymphocytic vasculitis. True lymphocytic vasculitis is nearly always confined to the small blood vessels of the superficial papillary dermis. Even in Sjögren’s syndrome, however, the histopathology encountered in most cases is that of leukocytoclastic vasculitis.

**Rheumatoid Vasculitis**

Rheumatoid vasculitis (RV) must be distinguished from the isolated digital (periungual) vasculitis that, in the absence of severe involvement, does not require intensive, vasculitis-specific therapy. Isolated digital vasculitis in patients with rheumatoid arthritis, characterized by splinter-like lesions in the periungual region (Bywaters’ lesions), is not necessarily associated with a poorer prognosis than rheumatoid arthritis without digital vasculitic lesions and does not require specific therapy for vasculitis. In contrast, RV is a potentially devastating complication that may involve both medium and small blood vessels and requires the most aggressive therapeutic interventions. Many clinical manifestations of RV are indistinguishable from polyarteritis nodosa, although microaneurysms are less common in RV. RV classically occurs in patients with nodular, rheumatoid factor–positive, joint-destructive disease who have few clinical indications of active synovitis at the time vasculitis begins. However, RV occasionally complicates early disease.

The most common presentation of RV includes purpuric lesions with or without evidence of concomitant medium vessel vasculitis. DIF examination of the skin lesions shows granular IgM and C3 deposition in vessels, consistent with an IC-mediated pathophysiology in which rheumatoid factor, complement, and cryoglobulins may all participate. Deep cutaneous ulcers near the malleoli are a hallmark of RV and require scrupulous local care, as well as judicious immunosuppression. Mononeuritis multiplex often complicates RV. High-dose glucocorticoids, tumor necrosis factor inhibitors, rituximab, and cyclophosphamide have been employed in the treatment of rheumatoid vasculitis.

**References**

hyperproteinemia was associated with thrombosis of the retinal veins and symptoms suggesting Raynaud’s disease, Bull Johns Hopkins Hosp 52:156, 1933.


The references for this chapter can also be found on www.expertconsult.com.
Primary Angiitis of the Central Nervous System

RULA A. HAJJ-ALI • CAROL A. LANGFORD

KEY POINTS

Primary angiitis of the central nervous system (PACNS) is defined as vasculitis that is confined only to the brain, meninges, or spinal cord.

PACNS consists of subsets including granulomatous angiitis of the central nervous system (GACNS) and atypical PACNS that are differentiated by unique histologic and imaging characteristics.

The diagnosis of PACNS is based on compatible clinical features together with evidence from spinal fluid, brain and vascular imaging, and often brain biopsy.

Ruling out mimics that have a similar clinical or angiographic appearance is central to the diagnosis of PACNS.

Reversible cerebral vasoconstriction syndrome (RCVS) is among the most important mimics of PACNS that has treatment and outcome implications.

Initial treatment of PACNS includes glucocorticoids either alone or in combination with cyclophosphamide based on the subset and severity of neurologic disease.

Outcome of PACNS is variable, with the highest rate of disability and mortality being seen in GACNS.

GENETICS

The pathogenesis of PACNS is not well understood. To date, there has been no evidence to suggest a genetic predisposition, although this remains under active investigation.

CLINICAL FEATURES

PACNS is divided into subsets that include granulomatous angiitis of the central nervous system (GACNS) and atypical PACNS.

Masslike lesions can occur in PACNS.

Clinical features of GACNS typically include chronic insidious headaches along with diffuse and focal neurologic deficits.

Great progress has been made toward understanding the clinical features of PACNS despite the many challenges that include the lack of highly specific diagnostic modalities, the sparse material for research, and the lack of controlled clinical trials. In the recent era, specific clinical and pathologic subsets of PACNS that have prognostic implications have been identified.

Proposed Criteria for Primary Angiitis of the Central Nervous System

In 1988 Calabrese and Mallek proposed diagnostic criteria for PACNS that emphasized the importance of ruling out entity. Subsequently, different terms surfaced in the literature such as isolated angiitis of the CNS to encompass cases that were characterized by nongranulomatous pathologic findings. Currently, PACNS is a well-accepted terminology of this disease that emphasizes the sole involvement of the CNS.

Following the description of diagnostic criteria for PACNS proposed by Calabrese and Mallek in 1988 and the potential for effective treatment, there was a tremendous increase of published cases in the literature such that more than 500 cases have now been described worldwide.

Because of its rarity and our evolving understanding of PACNS, the true incidence of PACNS is difficult to calculate. In the recent era, the estimated annual incidence rate of PACNS is 2.4 cases per 1 million person-years. Middle-aged men are often affected by PACNS with a median age at onset of approximately 50 years with a male-to-female ratio of around 2 : 1.

EPIDEMIOLOGY

The incidence of PACNS is estimated at 2.4 cases per 1 million person-years.

PACNS occurs more commonly in men than women at a ratio of 2 : 1.

The average age of patients with PACNS is 50 years.

Vasculitis affecting the central nervous system (CNS) most commonly occurs as a manifestation of a primary systemic vasculitis or as a secondary vasculitis in settings such as connective tissue diseases or infections. When the disease is confined only to the CNS (brain, meninges, and the spinal cord), it is referred to as primary angiitis of the central nervous system (PACNS).

PACNS is a rare disease, first reported as a distinct clinical pathologic entity in 1959 by Cravioto and Feigin. The disease was initially described as “noninfectious granulomatous angiitis with a predilection for the nervous system” that is fatal. Other reports of similar clinicopathologic phenotypes emerged in the literature, and “granulomatous angiitis of the central nervous system” was proposed to describe this entity.
mimics when diagnosing PACNS. These criteria include (1) the presence of an unexplained neurologic deficit after thorough clinical and laboratory evaluation; (2) documentation by cerebral angiography and/or tissue examination of an arteritic process within the central nervous system; and (3) no evidence of a systemic vasculitides or any other condition to which the angiographic or pathologic features could be secondary.

In 2009 Birnbaum and Hellmann proposed changes to the criteria described by Calabrese and colleagues, incorporating the levels of diagnostic certainty in their assessment. They proposed the term definite diagnosis of PACNS if there is confirmation of vasculitis on tissue biopsy and a probable diagnosis of PACNS when the diagnosis is based on high probability findings on an angiogram in the absence of tissue confirmation but with consideration of cerebrospinal fluid (CSF) profiles and neurologic symptoms to discriminate between PACNS and its mimics.

Although the original criteria by Calabrese served as a platform of literature-based research, their interpretation has changed fundamentally as advancement of our diagnostic modalities has revealed many unexplained neurologic diseases. The importance of ruling out mimics of PACNS remains essential in the diagnostic approach of patients with suspected PACNS.

**Clinical Subsets**

The original pathologic reports of PACNS as a granulomatous angiitis imposed a histologic-based nomenclature on the disease that led to the original name of *granulomatous angiitis of the CNS* (GACNS). Until the 1980s, PACNS was considered largely a homogeneous entity with a uniform clinical picture and grave prognosis. This paradigm was challenged with the introduction of direct vascular imaging as a diagnostic tool and recognition that nongranulomatous pathologic findings occur in PACNS.

Initial attempts at subclassification of PACNS described three broad subsets: GACNS, benign angiopathy of the CNS (BACNS), and “atypical” PACNS. BACNS has since become recognized to be a part of the reversible cerebral vasconstriction syndrome (RCVS) with the other PACNS subsets now being defined by pathologic or radiographic features.

**Granulomatous Angiitis of the Central Nervous System**

GACNS is a subset of PACNS described as a clinicopathologic entity characterized by granulomatous angiitis confined to the brain. This is a rare subset of PACNS, in which patients clinically present with chronic insidious headaches along with diffuse and focal neurologic deficits. Because the disease is confined to the brain, meninges, or the spinal cord, signs and symptoms of systemic inflammatory diseases are usually lacking. The diagnosis of this subset is confirmed by the findings of granulomatous angiitis on pathology (Figure 92-1). Typically the CSF findings include those of an aseptic meningitis picture with negative staining for microorganisms. GACNS predominantly affects middle-aged men. The most common findings on neuroimaging include infarcts, most often bilateral, as well as high-intensity T2-weighted fluid attenuation inversion recovery (FLAIR) lesions on magnetic resonance imaging (MRI) in the subcortical white matter and deep gray matter. Cerebral angiogram is not the diagnostic modality of choice given its poor spatial resolution of detecting small vessel vasculitis, which mainly occurs in GACNS.

**Atypical Central Nervous System Vasculitis**

This subset of PACNS comprises multiple manifestations of PACNS that are clinically, radiologically, or pathologically distinct from GACNS. Atypical PACNS represents the most frequent and heterogeneous subset of PACNS. Included in this subset are patients with specific presentations such as lesions or those with pathologic findings of lymphocytic infiltration rather than granulomatous angiitis.

**Masslike Presentation.** Masslike (ML) presentation is a rare manifestation of PACNS occurring in less than 5% of the cases. This presentation has gained attention after the recent report of a series of 38 patients with histologically confirmed PACNS that presented with a solitary cerebral mass. Typically the diagnosis is unanticipated and is confirmed after the pathologic examination from either biopsy samples or surgical excision of the mass. Unfortunately, there are no specific features on clinical assessment, neuroimaging, cerebral angiography, or CSF examination that could reliably distinguish ML-PACNS from other, more common causes of a solitary cerebral mass. Appropriate stains and cultures to rule out mycobacterial, fungal, or other infections and immunohistochemistry/gene rearrangement studies to exclude lymphoproliferative disease are essential to secure the diagnosis and exclude concomitant infectious or malignant processes.

**Cerebral Amyloid Angiitis.** Amyloid protein, in particular amyloid-β peptide, a fragment of the amyloid precursor peptide, can deposit in the brain, causing disease ranging from Alzheimer’s disease to cerebral amyloid angiopathy (CAA). CAA-related inflammation and angiocentric inflammatory reaction in CAA is referred to as...
amyloid-β–related angiitis (ABRA). Patients with ABRA tend to be older and more prone to hallucinations and mental status changes than other PACNS patients. MRI cannot distinguish between ABRA and other forms of PACNS, although there is a higher occurrence of cerebral hemorrhage in ABRA. ABRA carries a poor outcome, which could be related to older age and comorbidities.

**Angiographically Defined Central Nervous System Vasculitis.** The poor specificity of the cerebral angiogram poses a major challenge in the diagnosis of PACNS. When the diagnosis of PACNS is based on angiographic findings, a thorough evaluation should be performed to rule out mimics, especially RCVS.

**Spinal Cord Presentation.** Spinal cord presentation of PACNS is a rare subset in which disease is present only in the spinal cord. The diagnosis is usually made by biopsy.

**Nongranulomatous PACNS.** Pathologic findings of lymphocytic infiltration rather than granulomatous findings can occur in PACNS. In this category, the emphasis should be in ruling out secondary causes of CNS vasculitis such as infectious or lymphoproliferative diseases; adequate staining and immunophenotyping should always be carefully carried out in this category.

### DIAGNOSIS AND DIAGNOSTIC TESTS

#### KEY POINTS

- Spinal fluid, brain and vascular imaging, and brain biopsy are central to the diagnosis of PACNS and in ruling out other diseases.
- In PACNS 80% to 90% of patients have abnormal spinal fluid.
- Abnormalities on cerebral angiogram are not specific for PACNS and can be seen in a wide range of other settings.
- Reversible cerebral vasodilation syndrome is an important mimic of PACNS characterized by thunderclap headaches, normal CSF, and abnormal cerebral angiogram in which the changes resolve within 12 weeks.
- Other important entities in the differential of PACNS include infection, lymphoproliferative disease, primary systemic vasculitis, connective tissue disease, and thromboembolic disease.

The diagnosis of PACNS is challenging due to the nonspecific clinical presentation, the lack of highly specific laboratory and radiologic tests, and the difficulty in obtaining pathologic material.

#### Diagnostic Tests

**Laboratory Findings**

Elevation of acute phase reactants, anemia, and thrombocytosis are not typical of PACNS and, if present, should raise the possibility of a primary systemic vasculitis or other underlying disease. Laboratory testing to rule out connective tissue diseases and thromboembolic abnormalities should be performed. Infectious workup with appropriate cultures, serologies, and polymerase chain reaction testing should be performed by clinical and diagnostic findings and host risk factors.

**Cerebrospinal Fluid Analysis**

CSF is an important tool in the evaluation of PACNS. Although CSF findings are nonspecific in PACNS, its value also lies in ruling out other entities. Obtaining appropriate CSF cultures, microbiologic stains, cytology, and flow cytometry are crucial in ruling infectious and neoplastic disease. Elevated protein, modest lymphocytic pleocytosis, and occasionally oligoclonal bands and elevated IgG synthesis characterize the CSF in 80% to 90% of patients with pathologically documented PACNS. The median CSF white blood cell count is around 20 cells/µL, and the median CSF protein is approximately 120 mg/dL.

**Radiologic Evaluation**

MRI is a sensitive modality for the diagnosis of PACNS reaching 90% to 100%. Abnormalities include infarcts in 50% of patients, commonly affecting the cortex and the subcortex bilaterally. Affected areas include subcortical white matter, followed by deep gray matter, deep white matter, and the cerebral cortex. Hyperintense lesions on T2-weighted sequences are common but not specific for PACNS. Other abnormalities include mass lesions in 5% of patients; leptomeningeal enhancement in 8% of the cases; and gadolinium-enhanced intracranial lesions in about one-third of patients.

Cerebral vasculature imaging by catheter-directed dye angiogram or through magnetic resonance angiography (MRA) is an important modality in the diagnosis of PACNS. Alternating areas of dilatation and stenosis characterize the angiographic findings in PACNS and typically involve the vasculature on both sides but sometimes can involve single vessels. Other angiographic features include smooth tapering of one or multiple vessels. Although cerebral angiograms may visualize abnormalities in medium-sized vessels, they have limited sensitivity to detect abnormalities in small vessels that are less than 500 µm in diameter. Although cerebral angiography is valuable, its specificity for the diagnosis of PACNS can be as low as 25%. The reported “typical” angiographic findings for vasculitis are not specific to PACNS and can be encountered in atherosclerosis, radiation vasculopathy, or vascular spasm. Moreover, cerebral angiogram carries a poor positive predictive value in the diagnosis of PACNS in that the angiographic findings seen in RCVS can be consistent with those found in PACNS. Vascular studies should therefore be interpreted with caution and should not be considered the diagnostic “gold standard” in PACNS.

**Brain Biopsy**

Brain biopsy is thought to carry a low morbidity and mortality in patients with PACNS and is an important part of the evaluation to confirm the diagnosis and rule out mimics. When a brain biopsy is performed, an alternative diagnosis is identified in 30% to 40% of the cases. The interpretation of the brain biopsy results should take into consideration the potential for false-negative results because of the
patchy degree of involvement and the small amount of tissue that can often be obtained. The finding of vasculitis on pathologic examination does not exclude the diagnosis of infection or malignancy, and appropriate stains and markers should be pursued for accurate diagnosis.

**Differential Diagnosis**

The differential diagnosis algorithm of PACNS is large due to the lack of highly specific clinical features, laboratory, or imaging findings (Table 92-1). Excluding other entities that can have similar findings is critical in confirming a diagnosis of PACNS. Certain disease categories in particular should be considered and warrant further detailed discussion.

### Reversible Cerebral Vasocostruction Syndromes

RCVS is the leading mimic of PACNS, which is important to distinguish because it carries different implications for treatment and outcome (Table 92-2). RCVS comprises a group of disorders characterized by acute-onset headaches and reversible cerebrovasoconstriction.¹ These disorders include Call-Fleming syndrome, drug-induced angiopathy, migraine angiitis, BACNS, postpartum angiopathy, and drug-induced vasospasm. The clinical features of RCVS include acute onset of severe headaches, with or without neurologic deficit, with evidence of cerebral vasocostruction that is reversible. The headaches are usually recurrent thunderclap headaches and can be precipitated by straining and coughing. RCVS can be associated with strokes (39%), generalized tonic-clonic seizures (17%), convexity subarachnoid hemorrhage (34%), lobar hemorrhage (20%), and brain edema (38%).¹³ RCVS occurs more frequently in women than men. In contrast to PACNS, CSF analysis is generally normal in RCVS except where there is a concomitant disorder or subarachnoid bleeding. The diagnosis of RCVS is usually brought into consideration when cerebral vascular imaging reveals multiple areas of smooth or tapered arterial narrowing followed by segments of normal-caliber or distended arteries. Typically, arterial narrowing involves multiple intracerebral arteries and their branches in both hemispheres resulting in severe arterial narrowing. The diagnosis of RCVS is secured when repeat vascular imaging reveals reversibility of the cerebral vascular abnormalities, often occurring in about 6 to 12 weeks, which is in contrast to the fixed angiographic abnormalities encountered in PACNS. There is a lack of controlled trials to direct treatment of RCVS. Calcium channel blockers are used for symptomatic treatment of the headaches, with no evidence that they alter the clinical outcome. Some experts use glucocorticoid therapy but with no evidence that this improves the outcome.

### Primary Systemic Vasculitides

Primary systemic vasculitides such as granulomatosis with polyangiitis (GPA) (formerly Wegener’s granulomatosis), microscopic polyangiitis, Churg-Strauss syndrome (CSS), polyarteritis nodosa, or Behçet syndrome (BS) can affect the CNS leading to inflammation and vasculitis. In GPA, CNS disease occurs in 7% to 11% of patients and presents

---

**Table 92-1 Differential Diagnosis of Primary Angiitis of the Central Nervous System**

<table>
<thead>
<tr>
<th>Secondary Cerebral Vasculitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Systemic Vasculitides</td>
</tr>
<tr>
<td>Granulomatosis with polyangiitis</td>
</tr>
<tr>
<td>Microscopic polyangiitis</td>
</tr>
<tr>
<td>Churg-Strauss syndrome</td>
</tr>
<tr>
<td>Polyarteritis nodosa</td>
</tr>
<tr>
<td>Behçet syndrome</td>
</tr>
<tr>
<td>Connective Tissue Diseases</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>Sjögren’s syndrome</td>
</tr>
<tr>
<td>Inflammatory myositis</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Vasospastic Disorders</td>
</tr>
<tr>
<td>Reversible cerebral vasocostrictive syndrome</td>
</tr>
<tr>
<td>Drug exposures</td>
</tr>
<tr>
<td>Other Arterial Disease</td>
</tr>
<tr>
<td>Atherosclerosis</td>
</tr>
<tr>
<td>Fibromuscular dysplasia</td>
</tr>
<tr>
<td>Moyamoya</td>
</tr>
<tr>
<td>Dissection</td>
</tr>
<tr>
<td>Hypercoagulable States</td>
</tr>
<tr>
<td>Antiphospholipid antibody syndrome</td>
</tr>
<tr>
<td>Thrombotic thrombocytopenic purpura</td>
</tr>
<tr>
<td>Stroke-like Syndromes</td>
</tr>
<tr>
<td>CADASIL</td>
</tr>
<tr>
<td>Mitochondrial diseases</td>
</tr>
<tr>
<td>Sickle cell disease</td>
</tr>
<tr>
<td>Fabry’s disease</td>
</tr>
<tr>
<td>Sneddon’s syndrome</td>
</tr>
<tr>
<td>Leukoencephalopathies</td>
</tr>
<tr>
<td>Progressive multifocal leukoencephalopathy</td>
</tr>
<tr>
<td>Reversible posterior leukoencephalopathy syndrome</td>
</tr>
<tr>
<td>Cerebral Hemorrhage</td>
</tr>
<tr>
<td>Hypertensive</td>
</tr>
<tr>
<td>Aneurysmal</td>
</tr>
<tr>
<td>Amyloid angiopathy</td>
</tr>
<tr>
<td>Arteriovenous malformation</td>
</tr>
<tr>
<td>Embolic Disease</td>
</tr>
<tr>
<td>Thrombus</td>
</tr>
<tr>
<td>Cholesterol emboli</td>
</tr>
<tr>
<td>Myxoma</td>
</tr>
<tr>
<td>Endocarditis</td>
</tr>
<tr>
<td>Air emboli</td>
</tr>
</tbody>
</table>

CADASIL, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy.
in three different ways: direct invasion of granuloma from extracranial sites, remote intracranial granuloma, and rarely because of CNS vasculitis. The diagnosis of CNS vasculitis is often made on the basis of radiologic findings and by exclusion of other etiologies of the neurologic deficits, especially infection. As in GPA, the diagnosis of CNS vasculitis in CSS is often presumed without a histologic confirmation when other mimics are excluded. Neuroophthalmologic manifestations including amaurosis fugax, superior oblique palsy, ischemic optic neuropathy, fourth cranial palsy, and scattered areas of retinal infarction are common presentations of CNS vasculitis in CSS. In BS, CNS involvement occurs in 14% of cases manifesting as meningoencephalitis involving the brain stem and rarely as a consequence of thrombosis and inflammation of the dural venous sinuses. Pathologic evaluation in neuro-BS typically shows a mononuclear infiltration around the small vessels of the brain including the venous system, which is not a typical finding in PACNS. True CNS vasculitis rarely occurs in BS.

**Connective Tissue Diseases**

The brain is a common target in connective tissue diseases. In systemic lupus erythematosus (SLE), CNS involvement occurs in 14% to 80% of adults and 22% to 95% of children. Multifocal microinfarcts, cortical atrophy, small infarcts, hemorrhage, ischemic demyelination, and patchy multiple-sclerosis-like demyelination are typical findings in neuropsychiatric lupus. The most common microscopic brain finding in SLE is microvascularopathy, described as “healed vasculitis” consistent with hyalinization, thickening, and thrombus formation. Rheumatoid arthritis, Sjögren’s syndrome, and mixed connective tissue disease rarely affect the CNS in a vasculitic pattern. CNS vasculitis is typically a late occurrence in these diseases.

**Infections**

Infections are one of the most important entities that should be ruled out when considering PACNS. Vasculitis can occur in the setting of human immunodeficiency syndrome (HIV), often presenting as multifocal cerebral ischemia with pathologic findings of angiocentric lymphoproliferative lesions. In addition, HIV can result in granulomatous arteritis or can affect the brain secondary to co-infections such as syphilis. Treponema pallidum infection can affect vessels in the subarachnoid space and result in thrombosis, ischemia, and infarction that can mimic PACNS. Other infectious agents that have predilection to brain include varicella-zoster virus (VZV). VZV affects the brain either by involvement of large cerebral vessels, often affecting the proximal segments of the anterior and middle cerebral arteries in immunocompetent individuals, or small vessel disease in immunocompromised patients. History of VZV rash is usually elicited before the CNS involvement but not always. The diagnosis of VZV angiitis is confirmed by finding VZV DNA in CSF or the presence of reduced serum/CSF ratios of VZV immunoglobulin G (IgG). Tuberculosis affecting the CNS is an important cause of CNS vasculitis in endemic areas and should be carefully excluded. Hepatitis C, West Nile virus, parvovirus B19, and rarely herpes simplex virus can cause CNS vasculitis. Cytomegalovirus can cause CNS vasculitis as an opportunistic infection in immunocompromised individuals. Rarely cysticercosis can involve middle-sized cerebral arteries. A travel history and history of exposures are important features of the workup of patients with suspected PACNS.

**Lymphoproliferative Diseases**

Lymphoproliferative disease with predilection to the vascular wall such as lymphomatoid granulomatosis (LG) can cause CNS vasculitis leading to multiple small cortical infarcts and pathologic findings of multifocal angiocentric, angiodestructive lymphoma. Concomitant HIV infection is common with LG and should be ruled out. Rarely, LG is associated with other systemic autoimmune diseases such as Sjögren’s syndrome. Other lymphoproliferative diseases such as CNS lymphoma and intravascular lymphoma can mimic PACNS.

**Miscellaneous**

Atherosclerotic involvement of the intracerebral arteries is common and is often a consideration in the workup of PACNS. The poor specificity of the cerebral angiogram poses a major limitation in differentiating inflammatory causes from other vasculopathies. However, the lack of inflammatory changes of the CSF and the presence of multiple atherosclerotic risk factors should raise the suspicion for atherosclerotic disease. Other entities such as antiphospholipid antibody syndrome, hypercoagulable states, and thromboembolic etiologies should be carefully ruled out. The workup of PACNS should include transesophageal echocardiogram and hypercoagulable profile to rule out thromboembolic etiologies, especially in patients presenting with recurrent strokes. Other rare diseases that mimic the angiographic findings of PACNS include Moyamoya disease, small vessel arterial dissection, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), radiation vasculopathy, and thromboangiitis obliterans.
CHAPTER 92 | PRIMARY ANGIIITIS OF THE CENTRAL NERVOUS SYSTEM 1523

TREATMENT

KEY POINTS

No standardized trials have been conducted on the treatment of PACNS.

GACNS is typically treated initially with glucocorticoids and cyclophosphamide.

The treatment of PACNS subsets besides GACNS is based on the diagnostic subset and severity of neurologic impairment.

The treatment of PACNS is guided by expert opinion and reports of case series, as well as from extrapolation from controlled trials in primary systemic vasculitis. To date, there have been no controlled trials in the treatment of PACNS. Because of this, it is not possible to have a specific treatment algorithm and therapy is based on the subtype and neurologic severity.

The report of successful treatment of GACNS with a combination of cyclophosphamide and glucocorticoids led to the regular use of this regimen. Generally, patients are treated with a combination of cyclophosphamide and high-dose glucocorticoids for 3 to 6 months until remission is achieved. Then cyclophosphamide is stopped and switched to a maintenance agent following the same principles that are used in small vessel vasculitis. Maintenance therapy usually consists of either azathioprine or mycophenolate mofetil and rarely methotrexate given its low CNS penetration. The duration of maintenance therapy in PACNS is not well defined.

The treatment of atypical PACNS varies highly. Multiple factors affect the treatment regimen in atypical PACNS, and the treatment is usually individualized taking into consideration the degree of the neurologic deficit and the features of the diagnosis. All atypical PACNS cases are treated with high dose of glucocorticoids initially, and the decision to add other immunosuppressive agents is individually based. ABRA and ML-PACNS are two PACNS subsets in which cyclophosphamide is considered at the outset of the treatment.

Assessing disease status and defining remission are crucial steps in the course of treatment of PACNS. Permanent deficits occurring after the initial event in PACNS should not be erroneously considered as lack of remission. Disease is considered in remission when there is stability or improvement of the clinical and radiologic features. Serial MRI should be obtained to help in the assessment of disease activity. Adjunctive therapy for osteoporosis prevention and prophylaxis for opportunistic infections should be incorporated into the treatment plan.

OUTCOME

KEY POINTS

PACNS has an estimated mortality rate of 10% to 17%.

Moderate to severe disability occurs in up to 20% of patients with PACNS.

Originally, PACNS was described as a fatal disease by Cravioto1 but an improved outlook became possible following the description of treatment with glucocorticoids and cyclophosphamide. From recent reports, the estimated mortality rate of PACNS varies between 10% and 17%39 with one study finding that survival correlated with the initial findings of infarcts and gadolinium-enhanced lesions on MRI. Salvarani and colleagues9 and Hajj-Ali and colleagues39 reported that 20% of patients with PACNS had moderate to severe disability as assessed by the modified Rankin disability scores and the Barthel index, respectively. Salvarani and colleagues9 reported an improvement in disability scores over time.

SUMMARY

Considerable progress has been made in our understanding of PACNS. Our ability to recognize disease mimics has substantially increased our accuracy in the diagnosis of PACNS. The identification of RCVS in particular has clarified many of the cases that were erroneously diagnosed as PACNS. Effective diagnosis and management of PACNS requires a multidisciplinary team to evaluate the clinical features, obtain and interpret diagnostic tests, and determine the most effective regimen that will treat the underlying disease and minimize therapeutic toxicity.

There remains a great need for controlled trials to further guide treatment for patients with PACNS in their initial and maintenance phases. Prospective evaluation of long-term cohorts is necessary to better determine the rates of disability and long-term outcome of PACNS.

References

Behçet’s Disease

WILLIAM S. KAUFMAN • ELIZABETH KAUFMAN McNAMARA • JOSEPH L. JORIZZO

**KEY POINTS**

- Behçet’s disease is a complex multisystem disease characterized by oral and genital ulcers and other systemic features.
- Diagnosis is based on the International Criteria for Behçet’s Disease including oral aphthae, genital aphthae, ocular lesions, cutaneous lesions, and a positive pathergy test.
- Cutaneous lesions should display a neutrophilic vascular reaction on histopathologic examination.
- Treatment is based on the degree of systemic involvement and ranges from topical corticosteroids to thalidomide to systemic immunosuppressive agents and tumor necrosis factor inhibitors.
- Prognosis is variable, and patients typically have periods of exacerbations and remissions.

Behçet’s disease is a chronic, complex multisystem disease characterized clinically by oral aphthae, genital aphthae, cutaneous lesions, and ophthalmic, neurologic, or rheumatologic manifestations. The first description of Behçet’s disease was probably by Hippocrates in the fifth century BC, and the first modern account was presented in 1937 by the Turkish dermatologist Hulusi Behçet, who reported on a patient with recurrent oral and genital aphthae and uveitis.

**EPIDEMIOLOGY**

Behçet’s disease is seen worldwide, with the highest prevalence reported in Turkey (80 to 370 patients per 100,000 inhabitants) and Japan (13.6 per 100,000). The prevalence and often the severity is increased in the Middle East and the Mediterranean (i.e., the “Silk Route”). The remainder of the Asian continent has a prevalence of 7 to 30 patients per 100,000 inhabitants. It is relatively uncommon in northern Europe (0.27 to 1.18 per 100,000 inhabitants) and the United States (0.12 to 0.33 patients per 100,000 inhabitants). Patients commonly fulfill the diagnostic criteria in their mid-20s to 30s. In the past, Behçet’s disease was thought to predominantly affect males, but current epidemiologic data show a more equal male-to-female ratio. Overall, the male-to-female ratio has decreased over the past 20 years with males being more affected in the Middle East and a female predominance existing in Korea, China, the United States, and northern Europe.

**CAUSE AND PATHOGENESIS**

Although the pathogenesis of Behçet’s disease remains unclear, many factors have been implicated. Heredity, immunologic factors, infectious agents, inflammatory mediators, and clotting factors likely contribute.

**Genetics**

The onset of Behçet’s disease is believed to be sporadic, though familial clustering, families with multiple affected members, has been reported. Individuals with a first-degree relative with the disease are at an increased risk of developing the disease. Additionally, children of individuals with Behçet’s disease may have an earlier age of onset, suggesting genetic anticipation, which is due to a progressive increase in nucleotide repeats through consecutive generations. Familial occurrence differs regionally throughout the world and is more common in Korea, Israel, Turkey, and Arab countries, compared with Japan, China, and Europe.

Studies have shown a significant association between the human leukocyte antigen (HLA)-B51 and Behçet’s disease. HLA-B51–positive patients are at an increased risk of developing Behçet’s disease (odds ratio of 5.9). This association is more common in the Middle East, Mediterranean, and Japan, though not seen as often in Western nations. Disease prognosis also appears to be more severe in HLA-B51–positive patients. The role of HLA-B51 in Behçet’s remains unclear. It may be that HLA-B51 is not directly involved in causing the disease but is closely linked to disease-related genes. Candidate genes have been localized to chromosome 6 and include the major histocompatibility complex class I chain-related gene A (MICA) and, more specifically, the MICA6 allele; perth block (PERB); new organization associated with HLA-B (NOB); and transporter associated with antigen processing genes (TAP). Other hypotheses suggest that HLA-B51 may contribute to the onset of Behçet’s disease by serving as a heterologous antigen either through original antigen presentation or through viral or bacterial molecular mimicry.

A recent genome-wide association study confirmed the HLA-B51 relationship and also identified a second, independent association within the MHC class I region. Although Behçet’s disease has many features in common with the spondyloarthropathies, especially those associated with inflammatory bowel disease (IBD), the disorder in IBD patients generally evolves in a pattern resembling reactive arthritis, with an erosive axial arthritis; erosive arthropathy and HLA-B27 are not associated with Behçet’s disease.

**Immune Mechanisms**

Immune mechanisms play a major role in Behçet’s disease. Heat shock proteins, cytokines, alterations in neutrophil and macrophage activity, and autoimmune mechanisms...
have all been implicated. Heat shock proteins are released in response to stress and may be involved in stimulating a T helper type 1 immune response through interaction with Toll-like receptors. Specifically, immunoglobulin M–type, 47-kD cell surface heat shock protein against α-enolase has been identified in patients with Behçet’s disease. Although most of the T lymphocytes thought to be involved in this reaction are of the γδ type, the diversity of T lymphocytes seen in the disease suggests a response to multiple antigens, which may account for the various symptoms seen in Behçet’s. Cytokines such as interleukin (IL)-1, IL-8, IL-12, IL-17, and tumor necrosis factor (TNF) seem to be involved in the pathogenesis. Although elevated cytokine levels may serve as an indicator of disease severity, it should be appreciated that plasma TNF levels may rise and fall as an acute-phase reactant along with C-reactive protein and the erythrocyte sedimentation rate. The production of these proinflammatory cytokines, which are responsible for the chronic inflammation observed, may be the result of activated macrophages. In addition to macrophage activation, neutrophil chemotaxis and phagocytosis are increased in the lesions of Behçet’s disease. This increased activity of neutrophils leads to tissue injury in the form of the neutrophilic vascular reaction seen in lesions such as aphthae, pustular cutaneous lesions, and erythema nodosum–like lesions. Circulating immune complexes and activated homocysteine levels have been identified in patients with Behçet’s disease. Increased nitric oxide concentrations in the serum of Behçet’s disease patients. Increased nitric oxide concentrations in the lesions, synovial fluid, and aqueous humor of individuals with Behçet’s may play a role in endothelial activation, resulting in vascular inflammation and thrombosis. Elevated homocysteine levels have been cited as the cause of the increased nitric oxide concentrations, suggesting that hyperhomocysteinemia may represent an acquired risk factor for Behçet’s, which is potentially reversible.

**Infectious Agents**

Several studies have suggested a role for various infectious agents in the pathogenesis of Behçet’s disease; however, no organisms have been consistently isolated. Antistreptococcal antibodies have been isolated in the serum of patients with Behçet’s disease. Higher concentrations of Streptococcus sanguis have also been found in the oral flora of patients with Behçet’s disease and may play a role in the development of aphthae, which is often the initial manifestation. In addition to streptococcal antigens, other bacteria such as Escherichia coli and Staphylococcus aureus may have a role in Behçet’s disease through the activation of lymphocytes. Additionally, a lipoprotein of Mycoplasma fermentans has been found in patients with Behçet’s. This lipoprotein (MALP-404) contains the specific peptide motif, which is capable of being presented by HLA-B51. Studies have also indicated that there is a higher rate of Helicobacter pylori cytotoxin–associated gene-A antibodies in Behçet’s patients. These antibodies may cause endothelial damage via cross-reaction with endothelial antigens. H. pylori eradication in these patients has been shown to decrease disease severity.

Herpes simplex virus (HSV) deoxyribonucleic acid (DNA) has been isolated from the nuclei of peripheral blood lymphocytes by polymerase chain reaction (PCR) assay in patients with Behçet’s disease. HSV has also been detected by PCR in biopsy samples of genital and intestinal ulcers of Behçet’s patients. Other studies, however, have shown no difference in the detection of HSV in Behçet’s patients with and without oral aphthae.

In summary, although the cause and pathogenesis of Behçet’s disease are not completely understood, they likely involve an infectious or environmental trigger and subsequent inflammatory response in a genetically predisposed individual. The article by Zouboulis and May provides an excellent overview of the current understanding of the pathogenesis of Behçet’s disease.

**CLINICAL FEATURES**

**Aphthae**

Oral aphthae, or canker sores (Figure 93-1), are often the initial feature of Behçet’s disease and constitute a requisite diagnostic feature (although many believe that Behçet’s occurs in the absence of oral aphthae). Oral ulcerations usually occur in crops of more than 3 to 10s of lesions, but individual lesions may occur on the buccal mucosa, gingiva, lips, and tongue. Aphthae tend to be painful and shallow, and they heal without scarring over 1 to 3 weeks. Genital ulcers typically occur on the scrotum and penis in males and on the vulva or vaginal mucosa in females. These aphthae are similar in appearance to oral lesions, but they have a greater tendency to scar and may recur less frequently. Lesions in the oral mucosa are generally easy to distinguish from oral HSV, but with genital lesions, HSV should be excluded by viral culture or PCR before lesions are accepted as a diagnostic criterion.

**Cutaneous Lesions**

Several cutaneous manifestations of Behçet’s disease have been described: erythema nodosum–like lesions, pyoderma gangrenosum–like lesions, Sweet’s syndrome–like lesions, cutaneous small vessel vasculitis, and pustular vasculitic
lesions (Figure 93-2) including lesions induced by trauma—the so-called pathergy lesion. Pathergy signifies the development of erythematous pustules or papules 24 to 48 hours following puncture of the skin with a 20- to 21-gauge sterile needle. Specimens from all these lesions demonstrate a neutrophilic vascular reaction on histopathologic analysis. Acneiform or pseudofolliculitis lesions should be considered nonspecific, nondiagnostic findings because of their common occurrence in acne vulgaris and folliculitis.

Ophthalmic Features

A variety of ocular manifestations have been reported in Behçet’s patients including anterior and posterior uveitis, retinal vasculitis, and hypopyon, with secondary glaucoma, cataract formation, decreased visual acuity, and synchiae formation. Ocular involvement occurs in 83% to 95% of men and 67% to 73% of women with Behçet’s disease. Although ocular involvement is not commonly the presenting feature of Behçet’s disease, it is a major source of serious morbidity, and close ophthalmologic evaluation and follow-up are critical to prevent blindness in these patients. BenEzra and Cohen suggested that if ocular disease does not present within a few years of diagnosis, it is unlikely to be a major problem.

Arthritis

The arthritis of Behçet’s disease is typically a nonerosive, inflammatory, symmetric, or asymmetric oligoarthritis, although polyarticular and monoarticular forms are also seen. The most commonly involved joints are the knees, wrists, ankles, and elbows. The prevalence of arthritis among different populations ranges from 40% to 60%, and joint erosions are not observed. Dilsen and colleagues reported that 10% of patients with Behçet’s disease had a sacroiliitis. However, HLA-B27–positive patients were not excluded from their series, and occult IBD was not excluded, as required by O’Duffy and Goldstein. Other studies have shown no significant difference in the occurrence of sacroiliitis between patients with Behçet’s disease and the normal population. HLA-B27–positive patients with erosive sacroiliitis should be included in the reactive arthritis or enteroopathic arthritis disease spectrum, given the erosive, axial nature of the arthritis in the HLA-B27 pattern. This contrasts with the classically nonerosive, nonaxial nature of the arthritis in Behçet’s disease. Oral aphthae, ocular lesions, erythema nodosum–like lesions, pusular vasculitis, and pyoderma gangrenosum all occur in patients with IBD.

Other Systemic Manifestations

Central nervous system (CNS) involvement is most commonly characterized by brain stem or corticospinal tract syndromes (neuro-Behçet’s syndrome), venous sinus thrombosis, increased intracranial pressure secondary to venous sinus thrombosis or aseptic meningitis, isolated behavioral symptoms, or isolated headache. Rarely, ruptured aneurysms, peripheral neuropathy, optic neuritis, and vestibular involvement can occur. Poor prognosis is associated with a progressive course, parenchymal or brain stem involvement, cerebellar symptoms, and cerebrospinal fluid abnormalities. Cranial and peripheral nerve involvement may also occur.

Patients with Behçet’s disease may have gastrointestinal lesions resembling orogenital aphthae. These occur most commonly in the ileocecal region and in the ascending colon, transverse colon, or esophagus. Large aphthae may lead to perforation. Presenting symptoms include abdominal pain, diarrhea, and melena. It is important to distinguish IBD from Behçet’s disease. Aphthae may also affect the bladder.

Pulmonary abnormalities are uncommon in Behçet’s disease. Pulmonary artery aneurysms occur most frequently, followed by other complications secondary to vasculitis affecting the small pulmonary vessels. Aneurysm, thrombosis, hemorrhage, and infarction can result and cause death in patients with Behçet’s disease. Renal manifestations are not common. They vary from minimal changes to proliferative glomerulonephritis and rapidly progressive crescentic glomerulonephritis. The pathogenesis likely involves immune complex deposition.

Cardiac complications include myocardial infarction, pericarditis, arterial and venous thromboses, and aneurysm formation. Thromboses more commonly involve the venous system, sometimes leading to superior and inferior vena cava obstruction. Cardiac manifestations, either occlusive or aneurysmal, are postulated to occur due to a vasculitis of the vasa vasorum, which induces a thickening of the media and splitting of elastic fibers. Atherosclerosis does not appear to occur at an increased rate, as is seen in many autoimmune diseases such as systemic lupus erythematosus. Mortality in Behçet’s disease is low and is usually related to pulmonary or CNS involvement or to bowel perforation.

Limited data are available evaluating Behçet’s in pregnancy. One case-control study reported more remissions than exacerbations during and after pregnancy with higher rates of pregnancy complications but no changes in neonatal outcomes.

HISTOPATHOLOGY

Histopathologic analysis of specimens from the cutaneous lesions seen in Behçet’s disease reveals a neutrophilic...
vascular reaction or even fully developed leukocytoclastic vasculitis. Microscopic examination of dermal capillary or venule walls shows neutrophilic infiltrates, nuclear dust, and extravasation of erythrocytes, with or without fibrinoid necrosis. Immune complex–mediated vasculitis is the likely mechanism in the development of Behçet’s disease. A previously reported finding of lymphocytic vasculitis in patients with Behçet’s disease is thought to represent an older lesion.

Biopsy specimens of synovial membranes reveal a neutrophilic reaction, with occasional plasma cells and lymphocytes. Immunofluorescence microscopy may show immunoglobulin G (IgG) deposition along the synovial membrane. Reports of synovial fluid analysis in patients with Behçet’s disease show leukocyte counts ranging from 300 to 36,200 cells/mm³, with a predominance of neutrophils and normal glucose levels. Synovitis is included as one of the O’Duffy-Goldstein criteria for the diagnosis of Behçet’s disease.

**DIAGNOSIS**

The diagnosis of Behçet’s disease can be difficult, particularly in patients with only a limited number of common features of the disease. Clinicians and investigators must rely on clinical criteria because there are no pathognomonic laboratory findings. Several sets of diagnostic criteria have been proposed including those by O’Duffy and Goldstein, Mason and Barnes, and a Japanese study group. In 2008 the International Team for the Revision of International Criteria for Behçet’s Disease revised the established criteria on the basis of the presence of various disease manifestations including oral aphthae, cutaneous lesions, positive pathergy, genital aphthae, and ocular involvement. Each manifestation contributes 1 or 2 points with a diagnosis confirmed in an individual with a score of 3 or more points (Table 93-1). A diagnostic algorithm based on the Revised International Criteria for Behçet’s disease is presented in Figure 93-3.

Although not required by diagnostic criteria, IBD, systemic lupus erythematosus, reactive arthritis, and herpetic infections should first be excluded because the presenting manifestations of these conditions are often similar to those of Behçet’s disease. Recurrent aphthous stomatitis and complex aphthosis, defined as recurrent oral and genital aphthae or almost constant, multiple (three or more) oral aphthae, should also be considered in the differential diagnosis of patients presenting with oral or genital aphthae.

The O’Duffy-Goldstein criteria mandate the presence of recurrent oral aphthae plus at least two of the following: genital aphthae, synovitis, posterior uveitis, cutaneous

---

**Table 93-1 Revised International Criteria for Behçet’s Disease***

<table>
<thead>
<tr>
<th>Manifestation</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral aphthosis</td>
<td>1</td>
</tr>
<tr>
<td>Skin manifestations (pseudofolliculitis, skin aphthosis)</td>
<td>1</td>
</tr>
<tr>
<td>Vascular lesions (phlebitis, large vein thrombosis, aneurysm, arterial thrombosis)</td>
<td>1</td>
</tr>
<tr>
<td>Positive pathergy test</td>
<td>1</td>
</tr>
<tr>
<td>Genital aphthosis</td>
<td>2</td>
</tr>
<tr>
<td>Ocular lesions</td>
<td>2</td>
</tr>
</tbody>
</table>

*Diagnosis of Adamantiades-Behçet’s disease is made with a score of 3 points.


---

**Figure 93-3 Revised international criteria for Behçet’s disease: diagnosis of Adamantiades-Behçet’s disease—classification tree format. (Modified from International Team for the Revision of International Criteria for Behçet’s Disease: Clinical manifestations of Behçet’s disease. The ITR-ICBD report, Clin Exp Rheumatol 26(Suppl 50): S1–18, 2008.)**
pustular vasculitis, and meningoencephalitis. Patients who have only two of these findings, one being recurrent oral aphthae, are considered to have an incomplete form of Behçet’s disease. Another concern with regard to the International Criteria for Behçet’s is the inclusion of acneiform lesions, which are a common nonspecific finding in both adolescents and adults. Therefore our group advocates histologic confirmation of vessel-based histology to exclude acne lesions and the use of both the O’Duffy and the International Study Group criteria to exclude patients with IBD and enteropathic arthritis.62

The initial evaluation should include referral for ophthalmologic consultation to identify insidious ocular involvement. Patients who have arthralgias, gastrointestinal symptoms, or neurologic abnormalities may require radiographic studies and evaluation by appropriate subspecialists. Cutaneous pustular lesions, erythema nodosum–like lesions, and pyoderma gangrenosum–like lesions should be biopsied (for both histologic evaluation and culture) to confirm the clinical diagnosis.

**TREATMENT**

Therapeutic options should be based on the degree of disease involvement (Tables 93-2 and 93-3).63

**Mucocutaneous Disease**

Patients with oral and genital aphthae can be treated with intraleisional, superpotent topical, or aerosolized (not inhaled) corticosteroids. Topical tacrolimus can also be used, often in combination with superpotent topical corticosteroids. Other palliative therapies include oral tetracycline solutions, topical anesthetics, and rinses containing chlorhexidine gluconate. Oral colchicine, 1 to 2 mg daily, can decrease the size and frequency of mucocutaneous lesions.63,64 Doses can be adjusted according to the degree of gastrointestinal upset experienced by the patient. Dapsone in a dose of 50 to 150 mg/day is often helpful alone65 or in combination with colchicine.61 Patients must be monitored for the development of hemolytic anemia and

<table>
<thead>
<tr>
<th>No.</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Any patient with BD and inflammatory eye disease affecting the posterior segment should be on a treatment regimen that includes azathioprine, or equivalent mycophenolate,* and systemic corticosteroids.</td>
</tr>
<tr>
<td>2</td>
<td>If the patient has severe eye disease defined as ≥2 lines of drop in visual acuity on a 10/10 scale and/or retinal disease (retinal vasculitis or macular involvement), it is recommended that either cyclosporine A or infliximab be used in combination with azathioprine and corticosteroids; alternatively, IFN-α with or without corticosteroids could be used instead. *Rituximab is also an effective treatment for severe eye disease.*75</td>
</tr>
<tr>
<td>3</td>
<td>There is no firm evidence to guide the management of major vessel disease in BD. For the management of acute deep vein thrombosis in BD immunosuppressive agents such as corticosteroids, azathioprine, cyclophosphamide, or cyclosporine A are recommended. For the management of pulmonary and peripheral arterial aneurysms, cyclophosphamide and corticosteroids are recommended.</td>
</tr>
<tr>
<td>4</td>
<td>Similarly, there are no controlled data on, or evidence of benefit from, uncontrolled experience with anti-coagulants, anti-platelet, or anti-fibrinolytic agents in the management of deep vein thrombosis or for the use of anti-coagulation for the arterial lesions of BD.</td>
</tr>
<tr>
<td>5</td>
<td>There is no evidence-based treatment that can be recommended for the management of gastrointestinal involvement of BD. Agents such as sulfasalazine, corticosteroids, azathioprine, TNF antagonists, and thalidomide should be tried first before surgery, except in emergencies.</td>
</tr>
<tr>
<td>6</td>
<td>In most patients with BD, arthritis can be managed with colchicine.</td>
</tr>
<tr>
<td>7</td>
<td>There are no controlled data to guide the management of CNS involvement in BD. For parenchymal involvement, agents to be tried may include corticosteroids, IFN-α, azathioprine, cyclophosphamide, methotrexate, and TNF antagonists. For dural sinus thrombosis corticosteroids are recommended.</td>
</tr>
<tr>
<td>8</td>
<td>Cyclosporine A should not be used in BD patients with CNS involvement unless necessary for intraocular inflammation.</td>
</tr>
<tr>
<td>9</td>
<td>The decision to treat skin and mucosa involvement will depend on the perceived severity by the doctor and the patient.</td>
</tr>
</tbody>
</table>

*Our modifications to the recommendations are noted in italics.


<table>
<thead>
<tr>
<th>Table 93-2</th>
<th>Treatment of Behçet’s Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mucocutaneous Disease Only</strong></td>
<td></td>
</tr>
<tr>
<td>Topical, intralesional, or aerosolized corticosteroids</td>
<td></td>
</tr>
<tr>
<td>Topical sucralfate</td>
<td></td>
</tr>
<tr>
<td>Local anesthetics</td>
<td></td>
</tr>
<tr>
<td>Topical tacrolimus</td>
<td></td>
</tr>
<tr>
<td>Colchicine (1-2 mg/day)</td>
<td></td>
</tr>
<tr>
<td>Dapsone (50-150 mg/day)</td>
<td></td>
</tr>
<tr>
<td>Combinations of these agents</td>
<td></td>
</tr>
<tr>
<td><strong>Severe Mucocutaneous Disease</strong></td>
<td></td>
</tr>
<tr>
<td>Thalidomide (50-150 mg/day)</td>
<td></td>
</tr>
<tr>
<td>Methotrexate (2.5-25 mg/wk)</td>
<td></td>
</tr>
<tr>
<td>Prednisone</td>
<td></td>
</tr>
<tr>
<td>Interferon-α (3 million–9 million U/wk)</td>
<td></td>
</tr>
<tr>
<td><strong>Systemic Disease</strong></td>
<td></td>
</tr>
<tr>
<td>Prednisone</td>
<td></td>
</tr>
<tr>
<td>Azathioprine (50-200 mg twice daily)</td>
<td></td>
</tr>
<tr>
<td>Chlorambucil (4-6 mg/day)</td>
<td></td>
</tr>
<tr>
<td>Cyclophosphamide</td>
<td></td>
</tr>
<tr>
<td>Cyclosporine</td>
<td></td>
</tr>
<tr>
<td>Mycophenolate mofetil (1-1.5 g twice daily)</td>
<td></td>
</tr>
<tr>
<td>Intravenous immunoglobulin</td>
<td></td>
</tr>
<tr>
<td>Rituximab (severe ocular disease)*25</td>
<td></td>
</tr>
<tr>
<td>Anti–tumor necrosis factor agents</td>
<td></td>
</tr>
</tbody>
</table>

## Table 93-3

| Modified EULAR 2008 Recommendations for the Management of Behçet’s Disease (BD) |
|---|---|
| No. | Recommendation |
| 1 | Any patient with BD and inflammatory eye disease affecting the posterior segment should be on a treatment regimen that includes azathioprine, or equivalent mycophenolate,* and systemic corticosteroids. |
| 2 | If the patient has severe eye disease defined as ≥2 lines of drop in visual acuity on a 10/10 scale and/or retinal disease (retinal vasculitis or macular involvement), it is recommended that either cyclosporine A or infliximab be used in combination with azathioprine and corticosteroids; alternatively, IFN-α with or without corticosteroids could be used instead. *Rituximab is also an effective treatment for severe eye disease.*75 |
| 3 | There is no firm evidence to guide the management of major vessel disease in BD. For the management of acute deep vein thrombosis in BD immunosuppressive agents such as corticosteroids, azathioprine, cyclophosphamide, or cyclosporine A are recommended. For the management of pulmonary and peripheral arterial aneurysms, cyclophosphamide and corticosteroids are recommended. |
| 4 | Similarly, there are no controlled data on, or evidence of benefit from, uncontrolled experience with anti-coagulants, anti-platelet, or anti-fibrinolytic agents in the management of deep vein thrombosis or for the use of anti-coagulation for the arterial lesions of BD. |
| 5 | There is no evidence-based treatment that can be recommended for the management of gastrointestinal involvement of BD. Agents such as sulfasalazine, corticosteroids, azathioprine, TNF antagonists, and thalidomide should be tried first before surgery, except in emergencies. |
| 6 | In most patients with BD, arthritis can be managed with colchicine. |
| 7 | There are no controlled data to guide the management of CNS involvement in BD. For parenchymal involvement, agents to be tried may include corticosteroids, IFN-α, azathioprine, cyclophosphamide, methotrexate, and TNF antagonists. For dural sinus thrombosis corticosteroids are recommended. |
| 8 | Cyclosporine A should not be used in BD patients with CNS involvement unless necessary for intraocular inflammation. |
| 9 | The decision to treat skin and mucosa involvement will depend on the perceived severity by the doctor and the patient. |

*Our modifications to the recommendations are noted in italics.

methylglubinemia; the glucose-6-phosphate dehydrogenase level should be checked in all patients before beginning therapy with dapsone. Etanercept has also been shown to improve mucosal and skin manifestations of Behçet’s disease.66

Severe Mucocutaneous Disease

Patients who fail to respond to conservative therapy as outlined for mucocutaneous disease may require thalidomide. Its mechanism of action is thought to be mediated by modulation of TNF and other cytokines. Previous studies have shown that thalidomide is an effective agent for the treatment of severe mucocutaneous lesions of Behçet’s disease.67,68 Thalidomide is known to cause severe birth defects, and all patients and prescribing physicians must adhere to the System for Thalidomide Education and Prescribing Safety (STEPS) protocol including monthly follow-up visits.69 Patients receiving thalidomide can be monitored with nerve conduction studies for the development of peripheral neuropathy, if doing so is warranted on the basis of the clinical neurologic evaluation. Of note, the cost of treatment increased drastically when thalidomide was approved for the treatment of multiple myeloma.

Low-dose oral methotrexate (2.5 to 25 mg/wk) and low-dose prednisone are alternatives for patients with severe mucocutaneous involvement.70-72 Patients receiving methotrexate should be monitored for the development of hepatotoxicity and leukopenia. The risk of rebound after the tapering or discontinuation of systemic prednisone greatly limits its use for mucocutaneous disease alone. In addition, interferon-α is effective for severe mucocutaneous lesions and some systemic manifestations.71,72 A review of the safety and efficacy of interferon-α by Zouboulis and Orfanos71 recommended a 3-month high-dose regimen of 9 million units three times per week, followed by a low, maintenance dose of 3 million units three times per week. Patients do experience the usual flulike symptoms associated with this therapy.

Systemic Disease

Patients with systemic disease such as ocular and cardiovascular abnormalities require immunosuppressive therapy, particularly in view of the risk of morbidity and mortality resulting from untreated disease. Systemic corticosteroids may be used alone or in combination with other immunosuppressive agents such as azathioprine, interferon-α, cyclosporine, cyclophosphamide, and chlorambucil.4,73 The standard of care for eye disease is prednisone plus azathioprine.4,21 If this combination is not successful, one of the aforementioned immunosuppressive agents or mycophenolate mofetil can be substituted for azathioprine.21 Rituximab has also been effective in the treatment of severe ocular manifestations of Behçet’s disease.25 Several of these immunosuppressive agents have associated hematologic toxicity, as well as the potential for the development of associated malignancies; therefore close monitoring is essential. There have also been reports of Behçet’s disease treated with anti-TNF agents. A double-blind, placebo-controlled trial showed that etanercept was successful in suppressing most of the mucocutaneous manifestations of Behçet’s disease.76

Other reports support the efficacy of adalimumab and infliximab.77,78 In 2008 the European League Against Rheumatism (EULAR) developed a standard of recommendations for the management of the various systemic manifestations of Behçet’s disease (see Table 93-3).

OUTCOME

Behçet’s disease has a variable clinical course in most patients with a pattern of exacerbations and remissions. A delay in diagnosis after the initial manifestation of Behçet’s disease is common. Most patients present initially with mucocutaneous manifestations, and evidence of ocular and neurologic involvement may appear several years after diagnosis. Patients with the finding of complex aphthous may represent a forme fruste of Behçet’s disease, and they should be monitored for the development of additional abnormalities fulfilling the diagnostic criteria through regular follow-up and referral to appropriate specialists.84 The greatest morbidity is associated with ocular manifestations (two-thirds of patients), vascular disease (one-third of patients), and CNS disease (10% to 20% of patients).79 The most common cause of morbidity is ocular involvement; manifestations such as posterior uveitis and retinal vasculitis can cause blindness. Mortality in Behçet’s disease is low and is usually related to pulmonary or CNS involvement or to bowel perforation.7

References


The references for this chapter can also be found on www.expertconsult.com.
(2) an inflammatory response to the crystals so formed. How these processes occur and under what circumstances they cross from adaptive to pathologic responses are the subjects of this chapter.

EVOLUTIONARY CONSIDERATIONS

The ancient disease gout has a complex pathogenesis, and its modern relevance is underscored by a rise in prevalence by as much as fourfold in the past half century. Indeed, gout is now the most common inflammatory arthritis in the United States. Gout is a disease of both metabolism and inflammation/immunity. Gout pathogenesis requires the intersection of two distinct processes: (1) the intrinsic formation of uric acid, in the form of urate, at levels sufficient to drive the precipitation of urate into crystallized forms and

Uric Acid as a Danger Signal

Uric acid is a breakdown product of purine metabolism. As such, it represents a metabolic waste molecule that might, in theory, be nothing more than a nuisance requiring excretion. However, studies by Shi and colleagues and others suggest that evolution has co-opted this waste-generating process to play an important and perhaps critical role in organismal immunity. It had long been appreciated that the lysates of damaged mammalian cells can serve effectively as adjuvants—that is, can promote immune responses to injected antigens. Recently Shi and colleagues used classical biochemical techniques to demonstrate that the major endogenous adjuvant found in damaged cells was uric acid. These investigators further demonstrated that uric acid had the capacity to promote T cell activation in response to

KEY POINTS

Uric acid is the biologically active end product of human purine metabolism.

Serum urate concentrations are determined by the balance between urate production and elimination; hyperuricemia results from urate overproduction, urate underexcretion, or a combination of both.

Specific organic anion transporters (OATs) have recently been identified as playing a central role in the excretion of urate by the kidney.

Hyperuricemia is defined as serum urate levels greater than 6.8 mg/dL, the limit of its solubility in serum.

Gout pathogenesis requires the accumulation of monosodium urate at levels sufficient to drive the precipitation of crystals, resulting in the initiation of an inflammatory response.

Monosodium urate crystals activate the NLRP3 (NALP3) inflammasome, a multimolecular cytosolic complex that processes and generates interleukin-1β (IL-1β), IL-18, and IL-33.

The initiation of gouty inflammation by local white blood cells induces an influx of neutrophils into the joint; when these neutrophils encounter urate crystals, they become activated and propagate further inflammation.

Low-level inflammation is present in chronic gout and tophaceous gout; macrophages continue to produce cytokines and proteases, thereby facilitating cartilage and bone destruction.

The metabolic production of uric acid is ubiquitous among mammals and many other forms of animal life, and it is important to recognize that urate generation is not a priori pathologic. Indeed, the production of uric acid may serve one, or possibly a multitude of beneficial roles, an area of interest to molecular immunologists and molecular anthropologists alike.

KEY POINT

High baseline levels of serum urate in humans may be the result of evolutionary pressures in an era when hyperuricemia provided survival benefits.
antigen and that aggressive urate-lowering treatment could abrogate murine immune responses. Thus uric acid may serve as a danger signal to promote immune responses. As first proposed by Matzinger, a danger signal is an intrinsically produced molecule, typically issued by an altered or damaged cell in order to alert the immune system to the need for an immunologic response. Viewed from this perspective, the production of uric acid in a virally infected cell, for example, might serve as an upstream “second signal” to promote antigen presentation by a professional antigen-presenting cells such as dendritic cells, macrophages, or B cells. Indeed, although damaged or dying cells tend to have limited ability to manufacture proteins, their output of uric acid characteristically increases during cellular breakdown. The uric acid danger signal might also play an important role in tumor immunity, and at least one mouse model suggests that modulation of uric acid levels may directly affect immune tumor rejection. Although these observations require more study, they are consistent with a paradigm in which urate production at the local cellular level may play an important role in immunity and homeostasis.

**Uric Acid and Human Evolution**

Most mammals have serum urate levels in a range roughly between 0.5 and 2 mg/dL. In contrast, humans and other primates, including some New World monkeys, typically demonstrate serum urate levels between 4 and 6 mg/dL. The genetic and biochemical basis for these increases is well appreciated. During the Miocene era (10 to 25 million years ago), mutations in various primate and some monkey species resulted in inactivation of the uricase gene, which codes for the enzyme that degrades uric acid to allantoic acid. Genetic studies indicate that the uricase gene experienced nonsense mutations during that period, not once but multiple times across multiple hominoid lineages (Figure 94-1). The occurrence of multiple independent loss-of-function mutations has led some biologists to hypothesize that increases in urate generation may have conferred a survival benefit for these particular species. Several compelling, and not necessarily mutually exclusive, hypotheses have been proposed.

Ames and colleagues noted the fact that these same primate species had also been subject to a unique deletion of the gene permitting organisms to produce ascorbic acid, an event that apparently occurred in the Eocene era, some 10 to 20 million years before the uricase deletions. In mammals that do produce ascorbic acid, this molecule serves as the pre-eminent antioxidant in the body. Thus the loss of ascorbate production may have been an evolutionary liability, for which increases in urate provided compensation, particularly as a protectant against aging and cancer. Other authors have suggested that the effects of urate may have been particularly important in the central nervous system and that hyperuricemia provided an evolutionary advantage by promoting hominoid intellectual function, either through its antioxidant effects or via activation of neurostimulatory adenosine receptors (in a manner similar to that of caffeine). Although the antioxidant theory would appear compelling, critics have pointed out that (1) the production of urate itself generates oxidant molecules, diminishing any possible urate benefit; (2) intracellular urate may have pro-oxidant effects; and (3) the human/primate capacity for antioxidation may be large even in the absence of soluble antioxidants; for example, human red cell membranes have extensive antioxidant capacities.

Other investigators have examined the specific evolutionary pressures exerted on primate species during the Miocene era, in an attempt to understand the potential advantages of urate elevations. Johnson and colleagues pointed out that the Miocene era was an important era in hominoid evolution and that the hominoid diet during that period has appeared to have been mainly vegetarian and extremely low in salt. They suggest that hominoids during that period may have experienced a “hypotensive crisis,” particularly in the face of the transition to upright walking. They further postulate that an elevation in serum uric acid levels provided a mechanism for restoring normotension, primarily through urate-induced renovascular injury. To model this hypothesis, these investigators exposed rats to a low-salt diet, resulting in hypotension. When treated with the uricase inhibitor oxonic acid (to mimic the primate uricase loss), the rats’ urate levels rose and their blood pressures normalized. The effects of uricase inhibition on blood pressure could be reversed through the use of the urate-lowering agent allopurinol. These observations imply that what may once have been a homeostatic adaptation could now contribute to essential hypertension in today’s salt-rich era. In support of the latter hypothesis, Feig and colleagues identified adolescents with premature essential hypertension and hyperuricemia and treated them with the urate-lowering agent allopurinol. The result was normalization in blood pressure that reversed after allopurinol discontinuation.

Although the loss of uricase promoted serum urate increases in humans and other hominoids, the levels of serum urate so achieved were not sufficient to promote urate crystallization and gout. Rather, uricase inactivation created the circumstances under which additional increases in urate

---

Figure 94-1 Mutations affecting the uricase gene. During the Miocene era, hominoid species experienced not one but multiple different mutations, resulting in the loss of uricase activity. The survival of multiple species harboring different mutations suggests that loss of uricase may have conveyed a survival advantage, even as it established the necessary precondition for hyperuricemia and gout. *(Modified with permission from Wu XW, Muzny DM, Lee CC, Caskey CT. Two independent mutational events in the loss of urate oxidase during hominoid evolution, J Mol Evol 34:78–84, 1992. © Springer-Verlag GmbH.)*
production, or impairments in urate excretion, can result in serum urate concentrations exceeding the solubility threshold. Accordingly, we next review the mechanisms of urate production and excretion, as well as the events that may tip the scales toward pathologic hyperuricemia.

**URIC ACID PRODUCTION AND EXCRETION: NORMAL LEVELS AND HYPERURICEMIA**

Uric acid is a breakdown product of purines, and uric acid generation therefore depends directly on both intrinsic purine production and purine intake. In humans, uric acid is an end-product metabolite; consequently the depletion of uric acid depends directly on its excretion. The balance between uric acid production and excretion determines the serum urate level. Most individuals maintain a relatively stable uric acid level between 4 and 6.8 mg/dL and a total body uric acid pool of approximately 1000 mg. However, it is increasingly appreciated that individuals with high serum uric acid levels may deposit uric acid either occultly or in the form of appreciable masses (tophi), with the consequence that the total body urate pool may be significantly higher than in nonhyperuricemics. Such occult deposition of uric acid (total body urate burden) may have implications for treatment because they may form a "buffering reservoir" of urate that resists initial treatment with urate-lowering agents.

**Urate Production: Purine Metabolism and Intake**

In addition to the intrinsic synthesis and extrinsic intake of purines, uric acid production also depends on the metabolic processes that convert purines into uric acid. Next, we review purine and uric acid synthesis; purine intake is discussed later following Diet and Uric Acid.

**Purine Biosynthesis**

Purines are heterocyclic aromatic compounds, consisting of conjoined pyrimidine and imidazole rings (Figure 94-2). In mammals, the most common expression of purines is found in the form of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) (containing the purines adenine and guanine), as well as single-molecule nucleotides (adenosine triphosphate [ATP], adenosine diphosphate [ADP], adenosine monophosphate [AMP], cyclic AMP, and to a lesser extent, guanosine triphosphate [GTP] and cyclic glucose monophosphate [GMP]). Purines are also critical elements of the energy metabolism molecules NADH, NADPH, and coenzyme Q. Purines may also serve as direct neurotransmitters; for example, adenosine may interact with receptors.

![Structure of uric acid and common purines](image-url)

*Figure 94-2 Structure of uric acid and common purines. All purine bases may exist in the lactam form in a reversible manner as shown for uric acid.*
to modulate cardiovascular and central nervous system function.\(^{16}\)

Purine biosynthesis is initiated on a core of ribose-5-phosphate (Figures 94-3 and 94-4). The enzyme phosphoribosyl pyrophosphate (PRPP) synthase catalyzes the addition of a pyrophosphate moiety to form the adduct PRPP. This reaction is thought to be rate limiting. Subsequently, the enzyme glutamine-PRPP amidotransferase catalyzes the interaction of PRPP with glutamine to form 5-phosphoribosyl amine, the commitment step in purine biosynthesis. Glutamine-PRPP amidotransferase and PRPP synthase are both subject to feedback inhibition by IMP, AMP, and GMP, providing a mechanism to slow purine biosynthesis in the setting of purine sufficiency. 5-phosphoribosyl amine next forms the backbone for a series of molecular additions, ending in the formation of the purine inosine monophosphate (IMP). IMP is converted into either adenosine monophosphate (AMP) or guanosine monophosphate (GMP), which can then be phosphorylated into higher-energy compounds. Collectively, the process of purine biosynthesis is highly energy dependent, requiring the consumption of multiple molecules of ATP. Thus purine biosynthesis not only directly increases the substrate load for urate generation but also increases the turnover of already-formed purines that contribute to increased urate levels.\(^{17}\)

**Urate Formation and Purine Salvage**

Purines generated by the previously described mechanisms are susceptible to enzymatic catabolism, presumably to maintain purine homeostasis (Figure 94-5). Purines

---

**Figure 94-3** Purine biosynthesis. See text for details. ADP, adenosine diphosphate; AMP, adenosine monophosphate; ATP, adenosine triphosphate; 2,3-DPG, 2,3-diphosphoglycerate; GDP, guanosine diphosphate; GLN, glutamine; GMP, guanosine monophosphate; GTP, guanosine triphosphate; IMP, inosine monophosphate; PRPP, phosphoribosyl pyrophosphate; XMP, xanthine monophosphate.

**Figure 94-4** Formation of phosphoribosyl pyrophosphate by phosphoribosyl pyrophosphate synthase. In some patients, this reaction proceeds too rapidly, promoting hyperuricemia on the basis of primary overproduction of uric acid.
Figure 94-5 Uric acid synthesis and purine salvage. Catabolism of purines, especially inosine monophosphate (IMP) and guanosine monophosphate (GMP), results in urate synthesis via the common substrate xanthine. Xanthine oxidase is necessary for urate synthesis from any purine and so serves as a target for agents that inhibit uric acid synthesis (e.g., allopurinol, febuxostat). Purine salvage via hypoxanthine guanine phosphoribosyl transferase (HGPRT) returns hypoxanthine and guanine to IMP and GMP, respectively. HGPRT deficiencies result in not only increases in hypoxanthine and guanine and subsequent uric acid synthesis but also in the depletion of nucleotides that provide feedback inhibition on purine biosynthesis. Denoted in bold, mammals other than primates and some monkeys possess uricase, which converts uric acid to allantoic acid for further degradation. APRT, adenine phosphoribosyl transferase; AMP, adenosine monophosphate; PNP, purine nucleotide phosphorylase; XMP, xanthine monophosphate.

UREATE OVERPRODUCTION: PRIMARY AND SECONDARY CAUSES

Hyperuricemia may result from urate overproduction owing to primary or secondary causes.

Primary Urate Overproduction

In some patients, inborn errors of metabolism lead to urate overproduction and subsequent hyperuricemia. Several of these deserve mention. First, a small number of individuals possess PRPP synthase enzymes that are hyperactive. The result is increased generation of PRPP. Because under normal circumstances PRPP concentrations are below the Km of glutamine-PRPP amidotransferase for this substrate, increased PRPP levels drive amidotransferase activity and accelerate purine biosynthesis.

The second well-described class of abnormality occurs within the purine salvage pathway. Deficiencies of HGPRT result in impaired purine salvage and increased substrate for uric acid generation. Additionally, because purine salvage normally results in nucleotide monophosphate generation, patients with purine salvage failure experience lower levels of nucleotide monophosphates, as well as loss of feedback inhibition of both PRPP synthase and glutamine-PRPP amidotransferase. As a result, purine salvage failure resulting from HGPRT deficiency is accompanied by purine overproduction.
Two major variants of HGPRT deficiency have been described. Complete HGPRT deficiency, better known as Lesch-Nyhan syndrome, is an X-linked recessive disorder characterized by extremely high levels of serum urate, gouty attacks, nephrolithiasis, mental retardation, movement, and behavioral disorders including self-mutilating behavior. The disorder can occasionally arise by de novo mutation; female carriers are generally asymptomatic but may have elevated serum urate levels. In contrast to the gouty attacks and nephrolithiasis, which are direct consequences of hyperuricemia, the neurologic findings in Lesch-Nyhan syndrome are independent of hyperuricemia and unresponsive to urate-lowering drugs. Life expectancy can be greatly reduced, and these patients rarely come to the attention of adult rheumatologists.20

In contrast to Lesch-Nyhan patients, individuals with Kelley-Seegmiller syndrome have a partial deficiency of HGPRT.21 Kelley-Seegmiller patients typically present with hyperuricemia and gout and have limited or no neurologic symptoms.22 Several variants of Kelley-Seegmiller syndrome have been described, based on the extent of HGPRT inactivity and the presence/absence of neurologic findings. The mutations seen in the Kelley-Seegmiller variants tend to occur in regions of the HGPRT gene other than those identified in Lesch-Nyhan patients (whose mutations typically localize to the PRPP-binding region); whether such differences influence the presence/absence of neurologic symptoms has not been determined.23

Several hereditary defects of energy metabolism also promote hyperuricemia, mainly as a consequence of ATP consumption. Patients with glucose-6-phosphatase deficiency (type 1 glycogen storage, or von Gierke’s disease) demonstrate a high rate of both purine and ATP turnover. The lactic acidemia that secondarily occurs in patients with glucose-6-phosphatase deficiency may also contribute to hyperuricemia, by promoting decreased renal urate excretion (see later).24 In fructose-1-phosphate aldolase deficiency, patients lack the capacity to metabolize fructose-1-phosphate. Fructose-1-phosphate accumulation causes feedback inhibition of fructokinase and fructose accumulation in the blood. As an apparent consequence of these changes, AMP accumulates and promotes hyperuricemia by the mechanisms described earlier.25 The role of fructose intake in patients without inborn errors of fructose metabolism is discussed further later.

**Secondary Urate Overproduction and Hyperuricemia**

A number of secondary causes can lead to urate overproduction and hyperuricemia. In most cases, these conditions induce increased cell turnover, with resultant purine generation and breakdown. Chief among these must be counted diseases of erythropoietic, lymphopoietic, and myelopoietic cell turnover, of both malignant and nonmalignant varieties. Among the erythropoietic diseases causing hyperuricemia, autoimmune, and other hemolytic anemias (red cell destruction with increased red cell generation), sickle cell disease,26,27 polycythemia vera,28 and ineffective erythropoiesis (e.g., in pernicious and other forms of megaloblastic anemia, thalassemia, and other hemoglobinopathies) must be included.29 Patients with myeloproliferative and lymphoproliferative disorders including myelodysplastic syndrome, myeloid metaplasia, leukemias, lymphomas, and paraproteinemic diseases such as multiple myeloma and Waldenström’s macroglobulinemia are also at increased risk for hyperuricemia.30,31 In some cases, particularly in the pediatric setting, hyperuricemia and concomitant renal failure may be the first presentation of these malignancies.32 Indeed, the level of hyperuricemia may correlate with the degree of disease and cell turnover. Patients with essential thrombocytosis may also be at increased risk for hyperuricemia.33 An association between solid tumors and hyperuricemia has been reported34; given the slower turnover of solid tumor cells, solid tumor hyperuricemia tends to be less...
common and less severe than that seen in malignancies of bone marrow–derived cells.

Tumor lysis syndrome represents a unique form of tumor-related hyperuricemia, in which cell death induced by chemotherapy results in not only hyperuricemia but also hyperphosphatemia, hyperkalemia, and hypocalcemia, often resulting in acute renal failure and arrhythmias. Although tumor lysis syndrome occurs most commonly during treatment for hematologic malignancies, it may also occur during treatment of solid tumors. 33 Although not well documented in the literature, the authors have noted hyperuricemia subsequent to the use of granulocyte colony-stimulating factor for myelofibrosis-associated anemia. Such use of colony-stimulating factors may secondarily contribute to the new onset of gout. 36

Although somewhat more controversial, the increased cell turnover in patients with psoriasis has also been associated with elevated levels of serum urate. 37,38 An association between sarcoidosis and hyperuricemia has also been proposed, 33 again presumably relating to increased cell turnover and/or metabolic activity. However, the epidemiologic evidence supporting sarcoidosis as a cause of hyperuricemia is less than convincing. 39

Conditions leading directly to the physiologic consumption/degradation of ATP also contribute to the potential for secondary purine turnover leading to hyperuricemia. Thus strenuous and prolonged exercise, particularly to levels driving anaerobic respiration, may induce transient serum urate elevations. 40,41 Status epilepticus is likely to mimic these events. A number of acute illnesses including myocardial infarction and sepsis are also accompanied by ATP catabolism and may result in transient hyperuricemia. 42 Patients with hereditary myopathies including metabolic myopathies such as glycogen storage disease types III, V, and VII (debranching enzyme deficiency, myophosphorylase deficiency, and muscle phosphofructokinase deficiency, respectively), as well as mitochondrial myopathies (including carnitine palmitoyltransferase deficiency and myoadenylate deaminase deficiency), are susceptible to increases in serum urate levels after even moderate exercise. 43,44,45 In these individuals, a limited ability to synthesize ATP on demand apparently results in a rapid turnover of established ATP pools during exercise, with resultant purine and uric acid formation. Patients with medium-chain acyl-coenzyme A dehydrogenase deficiency, a defect of fatty acid metabolism, have also been shown to have elevated levels of serum urate, although the mechanism for this effect is not entirely clear. 44

**Urate Excretion: Gastrointestinal and Renal Mechanisms**

**KEY POINT**

Urate excretion occurs via the gut and the kidneys. In the kidneys, a complex series of urate transport proteins mediates a net elimination of sodium urate.

In most patients, serum urate level is maintained within a narrow range. Accordingly, mechanisms must exist to ensure disposal of urate, either by metabolism or excretion. As noted earlier, humans possess little or no capacity to metabolize urate; therefore urate excretory mechanisms play a critical role in homeostasis. The gastrointestinal tract and the kidneys each participate in urate excretion.

**Gastrointestinal Excretion of Urate**

Uric acid elimination via the gastrointestinal tract has been recognized for more than 50 years but has been relatively little studied. On the basis of radiolabeled urate tracer studies, Sorensen estimated that in healthy individuals, the gastrointestinal tract is responsible for the excretion of 20% to 30% of the daily uric acid burden. 45-47 Thus gastrointestinal excretion of uric acid may represent a minor pathway for urate excretion under most circumstances. Gastrointestinal uric acid excretion may become more important in settings of renal insufficiency, however, particularly in view of animal studies suggesting that uric acid excretion via the gut may increase in a compensatory manner in the setting of renal failure and decreased renal uric acid excretion. 48

Mechanisms of uric acid transport into the gut appear to include exocrine secretion (saliva, gastric, and pancreatic juices), as well as direct bowel secretory mechanisms. 44 Uric acid is apparently excreted into the gut in its native form and then undergoes degradation by intestinal flora. 47

**Renal Excretion of Uric Acid: Normal Mechanisms**

In all but the most extreme circumstances of renal failure, the kidney comprises the primary organ for uric acid excretion. The mechanisms of renal urate excretion are complex and involve multiple steps. In the bloodstream, uric acid (in the form of urate anion) is considered to be completely or nearly completely unbound to plasma proteins. As a result, nearly 100% of the urate load entering the renal afferent arteriole undergoes ultrafiltration by the glomerulus. As discussed further later, decreases in glomerular function therefore reduce urate filtration and promote rises in serum urate levels.

Subsequent to ultrafiltration, urate (initially in a monovalent ion form) undergoes several distinct handling steps: (1) a resorption step, in which as much as 90% to 98% of the filtered urate undergoes reclamation; (2) a secretion step, in which most of the urate resorbed in step 1 is retransferred back into the tubule lumen; and (3) a possible additional resorption step in which a smaller amount of uric acid is then resorbed. The net result is excretion of approximately 10% of the filtered load. In fully functioning nephrons, these steps are responsive to serum urate levels, such that rises in serum urate induce increased renal excretion and maintenance of total body urate homeostasis. Early studies, particularly in mouse models, suggested that these three steps might occur in anatomic sequence, in the proximal tubule, descending loop of Henle, and distal convoluted tubule, respectively. However, studies in humans over the past decade including both genetic and physiologic approaches suggest that these functions are likely to overlap and to occur mainly in the proximal tubule. Moreover, these same studies have emphasized the importance of organic anion transporters (OATs) and other active and passive transport molecules in the movements of urate in both directions across the proximal tubule (Figure 94-7). 49,50
Urate Resorption. Urate resorption in the proximal tubule depends on the action of several apical surface transporters and at least one resorption transporter at the basolateral surface (see Figure 94-7). In humans, the most important of the apical-surface transporters appears to be URAT1 (gene, SLC22A12). URAT1 acts as a urate/anion exchanger to transfer urate from the tubule lumen to the epithelial cytosol. The major inorganic counter-ion for URAT1 activity is Cl−. However, organic anions such as lactate, pyrazinoate, and nicotinate can substitute for chloride, with potential clinical consequences as discussed later. The importance of URAT1 to renal urate resorption is indicated by the fact that patients with inactivating mutations of URAT1 excrete nearly 100% of their filtered urate and demonstrate low serum urate levels (along with increased urinary uric acid levels and risk for uric acid kidney stones). Moreover, several well-established urate-lowering drugs including probenecid, benz bromarone, and losartan act by inhibiting URAT1. Conversely, other mutations in the URAT1 gene appear to convey a risk for increased urate resorption and hyperuricemia, presumably through a gain-of-function mutation.

OAT4 (gene, SLC22A11) and OAT10 (gene, SLC22A8) are two other apical anion transporters involved in renal uric acid resorption. Like URAT1, OAT10 is an anion exchange transporter; counter-ions that can promote urate transport by OAT10 include lactate, pyrazinoate, and nicotinate, a fact of clinical importance (see later). In contrast, although OAT4 also transports urate from the tubular lumen to the cytosol of renal epithelial cells, the counter-ions it employs tend to be dicarboxylates.

Urate transport by URAT1, OAT4, and OAT10 would lead to accumulation of urate intracellularly and presumably to gradients that would eventually impair further urate uptake if a mechanism did not exist to transport intracellular urate out of the cell at the basolateral surface. This function appears to be served by Glut9a (gene, SLC22A9; also confusingly known as URATv1), which was first identified as a glucose transport family protein but has little or no glucose transport capacity. Instead, Glut9a is an effective transporter of urate from the renal epithelial cell out into the renal interstitium. A number of different inactivating mutations of Glut9a have been identified in both humans and mice; in each case the result is impaired urate resorption, increased urate excretion, and hypouricemia. Glut9a and its splice variant, Glut9b, are also expressed on other cells including chondrocytes, leukocytes, intestinal cells, and hepatocytes; the role of Glut9 proteins on these cells is actively being explored.

Urate Secretion. Other transport proteins in the renal tubule epithelium regulate the excretion of urate from the peritubular fluid into the tubular lumen (see Figure 94-7). At the basolateral surface, OAT1 and OAT3 (genes, SLC22A6 and SLC22A8, respectively) transport urate from the interstitium into the epithelial cell cytosol. These transporters act via exchange with dicarboxylate anions and
transport not only urate, but other organic anions and some drugs. At the apical surface of proximal tubule cells, two proteins have been identified that serve as urate-extruding transporters. The multidrug resistance protein MRP4 (gene, ABCG4) mediates ATP-dependent urate transport. Whether energy failure promotes hyperuricemia by impairing MRP4 has not been established but would seem plausible. ABCG2 (gene, ABCG2) also directly mediates urate excretion. Genetic association studies have implicated two other anion transport proteins as playing a role in apical urate transport outside of the cell, namely, NPT1 (gene, SCL17A1) and NPT4 (gene, SLC17A3). Additionally, genetic studies have implicated nontransporter proteins as playing roles in urate excretion including PDZK1, CARMIL, NHERF1, SMCT1, SLC5A8, SMCT2, and SLC2A12. It is thought that some of these proteins may contribute to a macromolecular complex regulating urate transport.

**Renal Causes of Hyperuricemia**

**KEY POINT**

Decreased renal excretion of urate, owing to intrinsic or secondary causes, results in elevated levels of serum urate.

Many patients with hyperuricemia underexcrete urate; that is, for any given serum urate level, their degree of renal urate excretion is inadequate and less than is seen in normal controls (Figure 94-8). The mechanisms of urate underexcretion are various and stem from either primary or secondary renal effects.

**Primary Urate Underexcretion**

In a subset of patients, hereditary defects of renal tubule urate excretion result in hyperuricemia. With the identification of the aforementioned renal urate transporters, the underlying basis for some of these defects has become apparent. For example, up to 10% of gout cases in white Europeans may be attributable to hyperuricemia induced by defects in the urate exporter ABCG2. Gain-of-function mutations of other transport-related proteins (e.g., PDZK1, CARMIL, NHERF1) may actually promote tubular urate resorption mediated by URAT1 activity. Patients with such intrinsic tubular defects leading to net urate underexcretion frequently display normal renal filtration and normal serum creatinine levels.

Familial juvenile hyperuricemic nephropathy (FJHN), also known as medullary cystic kidney disease (MDCK), constitutes a group of autosomal dominant disorders characterized by early hyperuricemia and progressive chronic kidney disease. The hyperuricemia typically precedes the renal insufficiency and is considered to be primary. Three variants are recognized, designated MDCK 1, 2, and 3. In MDCK2, mutations in the uromodulin gene result in underproduction and/or misproduction of uromodulin (Tamm-Horsfall protein), the most prevalent protein secreted by the kidney. The probable importance of uromodulin deficiencies to FJHN has recently been underlined by the observation that although patients with the MDCK1 and 3 subtypes of FJHN have mutations that are not in the uromodulin gene, they nevertheless display a phenotype of decreased uromodulin expression. The mechanisms through which uromodulin deficiencies predispose to hyperuricemia are not yet understood, but mice transgenic for human uromodulin mutations develop renal tubular abnormalities and concentrating defects.

**Secondary Causes of Renal Urate Underexcretion**

A large number of underlying causes can result in nephrogenic retention of urate and subsequent hyperuricemia. These include acute or chronic renal failure, the effects of toxins and drugs, and systemic illnesses that alter renal urate handling directly or indirectly.

**Age and Gender.** Classic studies by French and colleagues documented that serum urate levels tend to be low in children. In males, urate levels increase precipitously at puberty, a time when females experience only a slight increase in serum urate. For women, a gradual increase over the subsequent years, followed by another rise at menopause, finally brings the serum urate near to that of men, consistent with Hippocrates’ astute aphorisms that “a young man does not take the gout until [around the time] he indulges in coition” and that “a woman does not take the gout, unless her menses be stopped.” This discrepancy between men and women, in the period between puberty and menses, suggests strongly that sex hormones play a role in urate regulation. Indeed, studies suggest that in women, estrogenic hormones may promote renal urate excretion.
Conversely, an active role for androgens in promoting hyperuricemia may be inferred from Hippocrates’ assertion that “Eunuchs do not take the gout” and from studies indicating that androgens and estrogens have opposing effects on renal organic anion transporters. In pregnant women, increases in serum urate levels are characteristic of pre-eclampsia, a reproductive emergency consisting of hypertension and proteinuria. The hyperuricemia of pre-eclampsia is thought to result from this condition’s renal dysfunction; hyperuricemia in pre-eclampsia does not lead to gout but is considered by some investigators to secondarily contribute to pre-eclamptic renal dysfunction.

**Systemic Illnesses** (Table 94-1). Renal insufficiency (i.e., reduction in glomerular filtration rate), whether acute or chronic and for whatever reason, promotes urate underexcretion and hyperuricemia. The mechanisms of hyperuricemia in patients with renal insufficiency are complex but depend first and foremost on decreased delivery of a filtered urate load to the renal tubule. At high levels of azotemia such as are rarely seen in this era of dialysis (blood urea nitrogen >100 mg/dL), hyperuricemia is practically universal. Lesser degrees of renal failure are more variably accompanied by hyperuricemia because decreases in glomerular filtration rate (GFR) promote some compensatory increase in renal tubule urate secretion. The effects of renal insufficiency on hyperuricemia may be more apparent in patients who also possess other risk factors for elevated urate levels. An association between congestive heart failure and hyperuricemia has also been reported. Although the effect direction of this relationship has not been well established, it is likely that the reduced renal perfusion seen in congestive heart failure can promote urate retention.

Various forms of organic (metabolic) acidoses may promote renal underexcretion of urate. Thus patients suffering lactic acidosis (e.g., in hypoxia, sepsis, liver or kidney disease, postsurgery or myocardial infarction, during excessive anaerobic exercise, or in response to certain drugs such as metformin) may become hyperuricemic. Similarly, patients experiencing ketoacidosis (e.g., alcoholic or diabetic ketoacidosis, starvation ketosis) may also develop hyperuricemia. Lactic acidosis may also occur secondarily to ketoacidosis. Although the mechanisms of these effects were previously considered to result from direct competition between the organic acids and urate for tubular excretion, they are probably better considered from the perspective of recent discoveries in renal urate transport (described earlier). In particular, organic acids serve as exchange anions for the apical surface renal urate transporters URAT1 and OAT10. Such acids therefore provide a motive force for increased urate resorption in the proximal tubule.

**Table 94-1** Systemic Conditions Promoting Hyperuricemia

<table>
<thead>
<tr>
<th>Overproduction</th>
<th>Underexcretion</th>
<th>Both Overproduction and Underexcretion</th>
<th>Metabolic States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemolytic anemia</td>
<td>Renal insufficiency</td>
<td>Myocardial infarction</td>
<td>Hyperthyroidism</td>
</tr>
<tr>
<td>Sickle cell disease</td>
<td>Dehydration/volume depletion</td>
<td>Congestive heart failure</td>
<td>Hypothyroidism</td>
</tr>
<tr>
<td>Polycythemia vera</td>
<td>Lactic acidosis</td>
<td>Sepsis</td>
<td>Hyperparathyroidism</td>
</tr>
<tr>
<td>Megaloblastic anemia</td>
<td>Ketoacidosis</td>
<td></td>
<td>Hypoparathyroidism</td>
</tr>
<tr>
<td>Thalassemia</td>
<td></td>
<td></td>
<td>Obesity</td>
</tr>
<tr>
<td>Myelodysplastic syndrome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leukemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymphoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waldenström’s macroglobulinemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essential thrombocytosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solid tumors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor lysis syndrome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psoriasis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarcoïdosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolic myopathies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitochondrial myopathies</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
urate have been performed in patients who have undergone renal filtration, suggesting a probable direct effect on tubular hyperuricemia. Moreover, cyclosporine’s mechanism of immune action is not dissimilar to that of cyclosporine effects. Reciprocally, hyperuricemia may exacerbate the nephrotoxic sporoine effect of decreasing renal glomerular filtration; it is likely that these organic anions act in a manner similar to nicotinate and salicylate. Another antituberculous agent, ethambutol, can also reduce renal tubular urate excretion and promote hyperuricemia. For example, the potassium-sparing diuretics triamterene, amiloride, and spironolactone do not raise urate. Interestingly, some diuretics actually lower serum urate levels, apparently by directly promoting renal urate excretion even as they induce volume depletion. One such drug was tiensin acid, an effective diuretic and anti-hyperuricemic that was withdrawn from use owing to hepatotoxicity. Several drugs that are weak organic acids may raise serum urate by serving as counter-ions to promote urate retention by URAT1 and OAT10. These drugs have also been assumed to inhibit tubular secretion, possibly by acting as urate competitors. Among these agents is the lipid-lowering drug nicotinic acid, which may not only block urate secretion but also promote urate formation. Low-dose salicylates including the low doses of aspirin used for cardioprotection may also raise urate by impairing renal urate efflux. At high doses salicylates become uricosuric, apparently through the inhibition of URAT1. The antituberculous agent pyrazinamide is the most potent urate-retaining agent known. Pyrazinamide is metabolized to pyrazinoate and subsequently to 5-hydroxypyrazinoate; it is likely that these organic anions act in a manner similar to nicotinate and salicylate. Another antituberculous agent, ethambutol, can also reduce renal tubular urate excretion and promote hyperuricemia. However, the mechanism of ethambutol’s action is not well established.

Table 94-2  Drugs Promoting Hyperuricemia

<table>
<thead>
<tr>
<th>Diuretics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Thiazide diuretics</td>
<td></td>
</tr>
<tr>
<td>Loop diuretics</td>
<td></td>
</tr>
<tr>
<td>active</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Cyclosporine</td>
<td></td>
</tr>
<tr>
<td>Ethambutol</td>
<td></td>
</tr>
<tr>
<td>Ethanol</td>
<td></td>
</tr>
<tr>
<td>Colony-stimulating factors</td>
<td>(?</td>
</tr>
</tbody>
</table>

Handling. For example, loop diuretics such as furosemide and bumetanide interact directly with the tubular urate transporter NPT4, and both thiazide and loop diuretics inhibit the renal urate exporter MRP4. Indeed, despite the volume-depleting effects of diuresis, not all diuretics promote hyperuricemia. For example, the potassium-sparing diuretics triamterene, amiloride, and spironolactone do not raise urate. Interestingly, some diuretics actually lower serum urate levels, apparently by directly promoting renal urate excretion even as they induce volume depletion. One such drug was tiensin acid, an effective diuretic and anti-hyperuricemic that was withdrawn from use owing to hepatotoxicity.

Several drugs that are weak organic acids may raise serum urate by serving as counter-ions to promote urate retention by URAT1 and OAT10. These drugs have also been assumed to inhibit tubular secretion, possibly by acting as urate competitors. Among these agents is the lipid-lowering drug nicotinic acid, which may not only block urate secretion but also promote urate formation. Low-dose salicylates including the low doses of aspirin used for cardioprotection may also raise urate by impairing renal urate efflux. At high doses salicylates become uricosuric, apparently through the inhibition of URAT1. The antituberculous agent pyrazinamide is the most potent urate-retaining agent known. Pyrazinamide is metabolized to pyrazinoate and subsequently to 5-hydroxypyrazinoate; it is likely that these organic anions act in a manner similar to nicotinate and salicylate. Another antituberculous agent, ethambutol, can also reduce renal tubular urate excretion and promote hyperuricemia. However, the mechanism of ethambutol’s action is not well established.

The immunosuppressant cyclosporine is well known to promote decreased renal urate excretion and hyperuricemia. The mechanism of cyclosporine-induced hyperuricemia is presumed to depend, at least in part, on the common cyclosporine effect of decreasing renal glomerular filtration; reciprocally, hyperuricemia may exacerbate the nephrotoxic cyclosporine effects. However, tacrolimus, whose mechanism of immune action is not dissimilar to that of cyclosporine, and which also can impair renal function, does not promote similar hyperuricemia. Moreover, cyclosporine’s effect on urate retention appears to exceed its effects on renal filtration, suggesting a probable direct effect on tubular urate transport. Most studies of cyclosporine’s effects on urate have been performed in patients who have undergone renal transplant; cyclosporine’s effects on urate in other settings are less well established.

Toxins. Several toxins can affect the kidney to promote hyperuricemia. Chief among these is lead. Lead exposure is endemic to Western society, but there have been a number of periods during which lead exposure may have been excessive (e.g., during the Roman Empire). A connection between lead exposure, hyperuricemia, and gout has been suspected at least since the 18th century. In the 20th century, a large cohort of patients with lead-induced hyperuricemia (saturnine gout) was recognized during and after the era of Prohibition, predominantly in the Southeastern United States and relating to the home brewing of whiskey (moonshine, or white lightning) using lead-lined vessels (typically, automobile radiators). Lead consumption results in distribution to a reservoir in bone and may have adverse effects on the central nervous system as well. In the kidney, lead poisoning leads to interstitial and perivascular fibrosis, as well as glomerular and tubular degeneration. Although patients with lead nephropathy do experience mild to moderate renal insufficiency, their clearance of urate is excessively limited, indicating a tubular defect in urate excretion. Suggestions that lead exposure may also promote purine turnover are provocative but have been less well supported. Although saturnine gout is currently a relatively rare condition, epidemiologic data suggest that it may be more prevalent than commonly assumed. In patients with moonshine nephropathy, associated lifestyle factors (e.g., alcohol consumption, obesity) may play a significant role in the genesis of the hyperuricemia. Patients with chronic beryllium poisoning, usually as a result of an occupational exposure, may also suffer diminished renal urate excretion and hyperuricemia.

**Diet and Uric Acid**

**KEY POINT**

Diet can have significant effects on serum urate levels, both by serving as a source of dietary purines and by altering the metabolic production and/or renal excretion of uric acid.

A number of dietary components can affect serum urate levels, mainly by leading to urate overproduction. Several other dietary components may have the capacity to lower serum urate levels.

**Purine-Rich Foods**

Purine-rich foods comprise a major source of daily purine load and hence a major source of generated urate. However, not all purine-rich foods convey equivalent risk: seafood and red meat, particularly organ meats, convey an increased risk for hyperuricemia, whereas consumption of purine-rich, leafy-green vegetables apparently does not. Earlier authors emphasized the limited effect of purine intake, with studies suggesting that alterations in purine intake may result in, at most, changes in serum urate of about 1 mg/dL. However, such studies have rarely accounted for the role of renal urate excretion in the context of diet. Thus it is worth considering whether, for a patient who is intrinsically
unable to excrete serum urate, an increased dietary purine load may produce more profound increases in serum urate levels than would be seen in a patient with normal excretory capacities.

In contrast to purine ingestion, protein consumption does not increase the risk of hyperuricemia and/or gout, a point of occasional confusion for practitioners, because many high-purine foods are also high in protein.\(^{108}\)

**Fructose**

Osler recognized the ability of fructose to provoke gouty attacks as early as 1901.\(^{110}\) Little was made of this observation, however, until the 1960s and 1970s, when it was demonstrated that fructose loads, administered orally or intravenously, cause transient rises in serum urate levels, particularly in gout patients.\(^{111,112}\) These effects are reproduced with consumption of sucrose (which contains fructose), but not glucose or galactose. Biochemical analysis of fructose metabolism has provided insight into the mechanisms of fructose-induced hyperuricemia (Figure 94-9). The first step in fructose metabolism (not shared with glucose or galactose) is the donation of a phosphate from ATP to form fructose-1-phosphate (enzyme phosphofructokinase), generating ADP. ADP is then converted to AMP (enzyme adenylate kinase), which in turn can be degraded via several steps to uric acid. In addition, fructose in effect serves as a phosphate "sink" because the donated phosphate is no longer available for the regeneration of ATP from AMP and ADP. Because both Pi and ATP inhibit the purine degradation pathway (inhibition of AMP deaminase and 5' nucleotidase, respectively), depletion of these compounds promotes the formation of uric acid as well.\(^{111}\) Depletion of ADP/AMP may also impair feedback inhibition and promote purine biosynthesis. Epidemiologic studies confirm a role for fructose consumption in hyperuricemia; patients who consume excessive fructose in the form of fructose-sweetened soft drinks or fruit juices demonstrate both higher serum urate levels and increased incidence of gout.\(^{114-117}\) The probable importance of fructose may be underscored by the fact that the rise in gout prevalence over the past several decades has occurred in parallel with an increased industrial use of fructose, rather than dextrose (glucose), as a major additive in soft drinks and prepared foods.

### Alcoholic Beverages

Ethanol ingestion is associated with incidence of gout, and ample physiologic and epidemiologic data confirm that ethanol consumption promotes the development of hyperuricemia.\(^{118-122}\) Ethanol is a particularly effective agent for raising serum urate levels because it works through multiple mechanisms. Chief among these is the requirement for ATP degradation during ethanol metabolism, resulting in increased purine turnover and urate generation.\(^{123,124}\) The ability of binge alcohol consumption to induce increases in lactate levels also contributes to hyperuricemia by decreasing renal urate excretion, likely through effects on URAT1 as discussed earlier.\(^{125,126}\) Ethanol consumption also promotes a diuresis, probably via antidiuretic hormone suppression;\(^ {127}\) as noted earlier, dehydration and volume depletion can promote renal urate retention. At high levels of acute-on-chronic alcohol consumption ketoacidosis may ensue, particularly in the setting of transient starvation and/or vomiting;\(^ {128}\) in such settings uric acid secretion may be inhibited, or uric acid resorption promoted by 3-hydroxybutyrate and acetocacetate, in much the same manner as lactic acid.\(^ {129}\) As a dietary matter, some alcoholic beverages, particularly beers and ales, are high in purines, mainly in the form of guanosine. Indeed, consumption of ethanol-free beer derivatives transiently raises serum urate levels, although not to the extent of beer itself.\(^ {130}\) The importance of purine load in alcoholic beverages may be underlined by the fact that, in contrast to beer, moderate wine consumption (<2 glasses/day) does not appear to increase serum urate levels.\(^{122,131}\) As noted earlier, saturnine gout contracted from lead-adulterated moonshine may also contribute to ethanol-induced hyperuricemia.

![Fructose metabolic pathway](image)

**Figure 94-9** Fructose and the generation of uric acid. Ingested fructose is converted to fructose-1-phosphate, a process that uses adenosine triphosphate (ATP), generates adenosine diphosphate (ADP), and sequesters inorganic phosphate (fructose/phosphate "sink"). The ADP generated then serves as a substrate for urate generation, while the depletion of ATP and Pi results in decreased feedback inhibition of the enzymes that mediate uric acid synthesis. AMP, adenosine monophosphate; IMP, inosine monophosphate.
Other Dietary Components

Recent data, both epidemiologic and physiologic, indicate that consumption of low-fat dairy products is independently associated with reduced serum urate levels and risk of gout. In physiologic studies, consumption of milk or milk proteins has a direct uricosuric effect that results in serum urate lowering. Interestingly, dairy products may also have anti-inflammatory effects. Regular heavy coffee consumption (4 to 6 cups daily) may have urate-lowering properties that are independent of caffeine, these effects appear to be concordant with a reduced risk of incident gout. In contrast, intermittent coffee consumption may be transiently prohyperuricemic, possibly as a result of caffeine-induced diuresis and volume depletion. Increased intake of vitamin C has been associated with decreased serum urate levels, possibly via a uricosuric effect.

CRYSTAL FORMATION: THE TRANSITION FROM HYPERURICEMIA TO GOUT

KEY POINT

Monosodium crystal formation involves physicochemical processes but may be regulated by synovial fluid proteins and immunoglobulins.

Soluble urate does not induce gouty attacks; only crystalized urate promotes acute inflammation. Urate crystallization is therefore a critical step in the progression from hyperuricemia to frank gout. Uric acid is a weak organic acid (pKa1 of 5.75 [position 9] and pKa2 of 10.3 [position 3]). At physiologic pH (7.4), approximately 98% of uric acid exists in the form of monosodium urate (MSU) monohydrate. At concentrations greater than 6.8 mg/dL, MSU exceeds its apparent solubility limit in serum, the definition of hyperuricemia. Exceeding the urate solubility threshold potentiates the precipitation of needle-shaped crystals and results in inflammatory responses. However, Sir William Roberts recognized that not all patients with hyperuricemia developed gout and drew a distinction between hyperuricemia and the crystallization of MSU. Epidemiologic studies also suggest that urate crystal formation and consequent development of acute gout may occur in a minority of subjects with hyperuricemia. Therefore factors other than hyperuricemia must influence the formation of urate crystals.

In vitro models have shed light on the relationship between urate crystallization and environmental factors such as pH, temperature, salt content, vibration, and large molecules. The fact that urate may more readily precipitate at both lower pH and lower temperature, for example, provides a possible rationale for the fact that gout attacks are most commonly seen in the first metatarsophalangeal joint, a joint that is both circumferentially exposed (i.e., relatively low in temperature) and at the farthest point of the systemic circulation.

Given that crystal formation and gouty attacks most commonly occur within the confines of the joint, other investigators have emphasized a possible role for joint biology per se in the formation of MSU crystals. For example, some investigators have suggested that soluble urate is excreted more slowly from the joint than other serum elements, providing a possible mechanism for urate concentration within the joint space. Once urate crystals are formed, the fenestrated endothelium of the synovium might permit the joint space to serve as a crystal “trap,” preventing crystals from being dispersed and/or dissolved in the wider circulation. Other investigators have emphasized the possible role of cartilage itself in the precipitation of urate, particularly in the setting of aging and/or osteoarthritis. Potential mechanisms through which aging cartilage might facilitate MSU nucleation/crystallization include changes in the proportions and/or chemical properties of cartilage glycoaminoglycans and proteoglycans, as well as increases in intracellular and extracellular lipid content of articular cartilage. One recent report suggests that osteoarthritis chondrocytes may actually secrete urate into the joint space, promoting local urate excesses in the synovial fluid.

Several investigators have suggested that urate crystallization may be an immune-assisted process. Kam and colleagues proposed that IgG antibody-binding to MSU monomers, previously assumed to be nonspecific, may permit the stacking of MSU crystals to occur despite dispersion forces that exist in fluid and tissue. More recently, Kanevets and colleagues postulated a role for urate crystal-specific IgM antibodies in the nucleation and formation of urate crystals.

Although urate precipitation is the sine qua non of acute gout, not all urate precipitation may lead directly to acute gouty arthritis. Imaging studies confirm that, in patients with asymptomatic hyperuricemia, urate crystals may nonetheless deposit in both cartilage and synovium. These deposits are of consequence not only for their potential to directly damage tissue (see Chronic Tophaceous Gout later) but also for their role as a reservoir of uncoated and potentially inflammatory urate crystals. Thus local trauma, long recognized as a possible antecedent to acute gouty attacks, may physically release cartilage-deposited crystals into the joint space, where they can initiate inflammation. In addition, the acute lowering of serum and synovial fluid urate levels (e.g., during initiation of urate-lowering drugs) is well documented to precipitate gouty attacks by a mechanism that is most likely akin to that in which glacial melting causes the sloughing off of icebergs and the exposure of previously hidden surfaces.

ACUTE GOUT ATTACKS: THE INFLAMMATORY RESPONSE TO MONOSODIUM URATE CRYSTALS

KEY POINT

Uric acid crystals activate both complement and resident synovial leukocytes, inciting neutrophil influx that itself promotes further inflammation.

Uric acid in its crystalline form is a potent trigger of inflammation. The phlogistic potential of crystalline urate in human gout was established dramatically in the 1960s by Faires and McCarty, who self-injected urate crystals into their knee joints and subsequently experienced attacks of
acute inflammation.\textsuperscript{154,155} In the clinic, the diagnosis of gout is made on joint aspiration, when examination of synovial fluid under polarized light microscopy demonstrates the presence of urate crystals, neutrophils, and particularly intracellular urate crystals, confirming directed neutrophil phagocytic activity.\textsuperscript{156} However, the inflammatory mechanisms of gout are complex and involve not only neutrophils but other cell types, numerous inflammatory mediators, and well-organized sequences of events.\textsuperscript{157}

**Uric Acid Crystals and Complement Activation**

Complement activation by the alternative pathway is a continuous process in body fluids, in which C3 component activation in the fluid phase is followed by rapid C3b deposition onto nearby surfaces.\textsuperscript{158} On most cell surfaces C3 is routinely inactivated by regulatory proteins. In contrast, the polyanionic surfaces of uric acid crystals provide opportunities for unconstrained C3 deposition and subsequent activation of downstream complement components. Weissmann and others\textsuperscript{159,160} have demonstrated the ability of urate crystals to activate complement from C2-depleted serum, confirming activation of the alternative pathway. Interestingly, other groups have confirmed that uric acid crystals also activate complement by the classical pathway (i.e., via C1 activation). Crystal activation of the classical pathway may occur in two ways. First, uric acid crystals may activate the classical pathway by an immunoglobulin-independent, C-reactive protein (CRP)-dependent pathway.\textsuperscript{159,161,162} In addition, it has been shown repeatedly that urate crystals possess the ability to bind antibodies. The specificity of IgG binding, as well as the question of whether antibodies bound to urate crystals can lead to additional activation of the classical pathway, is not fully resolved.\textsuperscript{159,162,163} One result of crystal-induced complement cascades is to produce C5a, a potent vasodilator and chemoattractant for inflammatory cells such as neutrophils.\textsuperscript{164,165} A study by Tramontini and colleagues suggests that urate crystal complement activation may also lead to generation of soluble complement membrane attack complexes, which may activate local cells to promote inflammation.\textsuperscript{166} Other proteins of potential import to inflammation including fibronectin and kininogen may also adhere to urate crystals.\textsuperscript{166}

**Cellular Response to Crystals**

**Cell Recognition of Crystalline Urate**

MSU crystals interact with, and potently stimulate, a range of inflammatory cells. How crystals activate cells remains an open question, but three major mechanisms have been proposed: (1) crystal recognition via Toll-like receptors (TLRs), (2) interactions between urate crystals and cholesterol rafts in the cell membrane, and (3) direct phagocytic mechanisms.

TLRs are critical for innate immunity and permit organisms to rapidly recognize bacteria and viruses on the basis of stereotypical features (pathogen-associated molecular patterns [PAMPs]), rather than characteristics unique to the specific invader. Because crystalline urate might theoretically be treated as a foreign molecule, several investigators have examined whether urate crystals can activate TLRs. In TLR2- and TLR4-knockout mice, decreased IL1-\(\beta\) and tumor necrosis factor (TNF) production, as well as decreased neutrophil influx in the air pouch inflammation model, argue strongly for a role for TLRs in urate crystal responses.\textsuperscript{167} Impairment of urate-driven inflammation in CD14 knockout mice also appears to support a role for TLRs (CD14 is essential for TLR2- and TLR4-dependent signaling).\textsuperscript{168} Interestingly, Joosten and colleagues suggest that TLR2 activation by MSU crystals requires simultaneous exposure of the receptors to C18:0 free fatty acids, suggesting a synergistic effect.\textsuperscript{169} However, other researchers have observed no effect of multiple TLR knockouts on murine models of urate-induced inflammation.\textsuperscript{170} The role of TLRs in urate signaling therefore remains something of an open question.

Other authors have emphasized the ability of MSU crystals to electrostatically interact with cholesterol.\textsuperscript{171} Cholesterol-rich regions of plasma membranes (lipid rafts) are characteristically rich in signaling molecules and represent hot spots for cellular activation. Receptor-independent interactions between crystalline MSU and lipid rafts have been demonstrated in dendritic cells, resulting directly in cell activation. The mechanism behind this effect appears to relate to hydrogen bond–dependent aggregation of lipid rafts; aggregation of transmembrane receptors within the rafts then results in activation of immunoreceptor tyrosine-based activation motifs (ITAMs), followed by activation of the signaling molecule Syk.\textsuperscript{172,173} Syk activation in turn can induce cell activation including phosphoinositol-3 (PI-3) kinase signaling, cytoskeletal rearrangement, and crystal phagocytosis.

As noted earlier, MSU crystals may become coated by immunoglobulins and other serum proteins and may serve as a substrate for complement activation. Thus protein-coated urate crystals may also activate cells via their ability to engage immunoglobulin, complement, and possibly other cell surface receptors. Consistent with this model, IgG- but not IgM-coated urate crystals have been observed to incite greater inflammatory responses than those that are uncoated.\textsuperscript{174}

**Intracellular Responses to Urate Crystal Encounters**

Activation of cells by urate crystals results in the activation of a number of intracellular signaling molecules associated with inflammatory responses. In addition to Syk, these include PI-3 kinase; the ERK, JNK, and p38 mitogen-activated protein (MAP) kinases; phospholipases C and D; rho-family proteins; and nuclear factor \(\kappa\)B (NF\(\kappa\)B). Depending on the cell type in question, activation of these molecules will result in specific phenotype alterations including cytoskeletal alterations, production of cytokines and lipid mediators, and induction of phagocytosis and superoxide generation.\textsuperscript{175-180} Additional responses, secondary to these initial ones, include vasodilation and vascular leakiness, the upregulation of adhesion molecules on endothelial and inflammatory cell surfaces, and cellular chemotaxis, discussed later.

It has long been appreciated that urate crystals can induce macrophages to produce IL-1\(\beta\) and that IL-1\(\beta\) so produced is central to the development of gouty inflammation.\textsuperscript{176} However, the mechanism of urate-induced IL-1\(\beta\)
upregulation remained, until recently, unknown. The NLRP3 (formerly NALP3, CIAS1, or cryopyrin) inflammasome is a multimolecular, cytosolic complex whose primary purpose is to generate IL-1β, as well as IL-18 and IL-33 (see also Chapter 18).\textsuperscript{181} The pro form of IL-1β is cleaved into activated IL-1β by the inflammasome-associated enzyme caspase-1, which may also play a role in IL-1β secretion. In a seminal study, Martinon and colleagues\textsuperscript{182} documented the ability of MSU crystals to activate the NLRP3 inflammasome and stimulate IL-1β generation (Figure 94-10). Subsequent inflammatory responses (e.g., TNF production) appear to occur secondary to autoengagement of cell surface IL-1β receptors. The importance of the inflammasome in crystal inflammation has been well documented in mouse models: Macrophages from mice lacking inflammasome components generate diminished levels of IL-1β on MSU crystal exposure, and intraperitoneal injection of urate crystals into inflammasome-deficient mice, as well as mice deficient in IL-1β receptors, leads to significantly decreased neutrophil recruitment.\textsuperscript{182} Moreover, the ability of IL-1β-directed therapies to abrogate both human gout and murine crystal-induced inflammation speaks to the centrality of IL-1β in the inflammatory response to urate crystals.\textsuperscript{183}

To date, there is no definitive explanation as to how urate crystals activate the NLRP3 inflammasome. Leading theories center on the inflammasome as an internal sensor for cell stress resulting from either (1) oxidative stress or shifts in ion concentration, and/or (2) lysosomal disruption.\textsuperscript{184,185} In one proposed model, crystal-induced damage to the plasma membrane promotes cellular potassium efflux and the resultant hypokalic state activates the inflammasome directly.\textsuperscript{186} Additionally, crystal activation of the phagocyte NADPH oxidase leads to the generation of reactive oxygen species (ROS), which may then be directly or indirectly sensed by the inflammasome. The ability of large urate crystals to provoke incomplete or frustrated phagocytosis—in which a phagocyte surrounds but cannot fully engulf a target—may lead to an activated state (including oxidase activation) that also contributes to inflammasome activation.\textsuperscript{184} The second proposed mechanism of inflammasome activation by MSU crystals is based on the ability of phagocytosed urate crystals to rupture the membranes of phagolysosomes, either by mechanical or physicochemical effects.\textsuperscript{187} Phagolysosome rupture would lead to cytosolic acidification and intracellular release of cathepsin B, each of which has been proposed as a NLRP3 inflammasome activator.\textsuperscript{185} These various mechanisms may not be mutually exclusive and may act synergistically (e.g., lysosomal disintegration and cathepsin-B release may themselves promote the production of ROS).

**Initiation and Propagation of the Acute Gouty Attack**

The clinical picture of acute gout is one of rapid, almost explosive development of an inflammatory response. Accordingly, cellular changes during the acute attack must reflect the accelerating nature of the inflammation.

*Figure 94-10* Activation of the NOD-like receptor family, pyrin domain containing 3 (NLRP3) inflammasome by monosodium urate crystals. Activation of the inflammasome by urate crystals results in activation of caspase-1, which cleaves and activates interleukin-1β (IL-1β) and promotes IL-1β secretion. In turn, IL-1β can engage its receptors to secondarily promote the synthesis and secretion of other cytokines such as tumor necrosis factor (TNF) and IL-6. Three possible mechanisms of inflammasome activation by urate crystals are illustrated: (1) production of reactive oxygen species (ROS), (2) activation of Toll-like receptors 2 and 4 (TLR2/4), and (3) potassium depletion, which may be sensed by the inflammasome. The precise roles of these and other mechanisms of inflammasome activation by urate crystals remain a matter of investigation. ASC, apoptosis-associated speck-like protein containing a caspase-recruitment domain.
Beginning with the appearance/release of urate crystals, the initial phase of gouty inflammation must depend on already-available local mediators, as well as cells that are both (1) capable of urate responses and (2) already in place within the joint. Clearly, immunoglobulin binding and the activation of complement on the crystal surface must be one such early response. Interaction of the crystal with local tissue cells, facilitated in part by IgG and complement opsonization, is also likely to play an early role. Best studied among the early cell type responses is that of the synovial macrophage, a resident cell in the synovial membrane. Macrophage activation in response to crystals results in synthesis and secretion of important cytokines such as IL-1β, TNF, IL-6, and the chemokine IL-8 (CXCL8), as well as release of potentially tissue-destroying matrix metalloproteinases and toxic oxygen radicals. Activation of synovial macrophages is contemporaneous with their phagocytosis of urate crystals. Human synovial fibroblasts are also capable of responding to urate crystals, leading to the generation of both inflammatory mediators and metalloproteinases.

Finally, mast cells are also resident within the synovium and appear to be important in the early phases of acute gouty responses, as their depletion diminishes inflammation in murine models of crystal-induced inflammation. Mast cells increase in number and activity in the lining of experimental air pouches after MSU crystal injection, suggesting that additional mast cells may be secondarily attracted to the inflammatory site. These early events promote the influx of both additional bloodstream monocyte/macrophages and polymorphonuclear neutrophils, the predominant cell in the inflamed gouty joint. Cytokines and other inflammatory mediators produced in the early phase of the crystal response act on the vascular endothelium to promote vasodilation and leakiness and to upregulate the expression of adhesion molecules such as selectins and ICAMs (intercellular adhesion molecules) on the vascular surface of endothelial cell. These same cytokines also promote neutrophil activation within the bloodstream, particularly the upregulation of the integrin adhesion molecule CD11b/CD18. A role for the crystal-generated complement component C5a in the activation of bloodstream neutrophils is also likely because C5a can potently stimulate CD11b/CD18 activation and neutrophil adhesiveness. The result is that neutrophils and (in far fewer numbers) monocytes first adhere tightly to the endothelium, resulting in neutrophil influx and the further propagation of inflammation as neutrophils undergo direct activation by urate crystals. See text for additional details.

Figure 94-11 Initial phases of urate crystal–induced activation. The presence of “fresh” urate crystals, resulting either from spontaneous precipitation or the liberation of crystals from established pools, results in direct complement activation and activation of resident cells in the synovium including macrophages, fibroblasts, and mast cells. Activated cells produce interleukin-1β (IL-1β) and other cytokines, as well as multiple other mediators (not all illustrated) that in turn activate both bloodstream neutrophils and endothelial cells. Not illustrated, these responses permit neutrophils to adhere to and traverse the endothelium, resulting in neutrophil influx and the further propagation of inflammation as neutrophils undergo direct activation by urate crystals. See text for additional details.
importantly include interactions with CD11b/CD18 (which is also a complement receptor) and the IgG receptor FcyRIIIb. Subsequent neutrophil intracellular signaling events include phosphorylation and/or activation of a number of tyrosine kinases including Lyn, Syk, Tec, and Src; activation of PI-3 kinase; and activation of phospholipase C. Activation of MAP kinases is probably also involved because these kinases regulate neutrophil adhesion and superoxide generation.

Neutrophils encountering urate crystals can rapidly promote additional inflammation and additional neutrophil influx via several mechanisms (Figure 94-12). First, neutrophils stimulated by urate crystals generate a number of inflammatory mediators and potent neutrophil chemoattractants including IL-1β, IL-8, leukotriene B4 (LTB4), S100A8/A9, prostaglandin E2, and crystal-induced chemotactic factor. Of these, IL-8 and LTB4 are potent chemoattractants. They play a particularly important role in amplifying the chemoattractant gradient and promoting additional neutrophil influx. Second, neutrophils stimulated by crystals release a wide range of products capable of directly damaging local tissues including oxygen radicals and metalloproteinases such as MMP-8. Although these mediators are intended for digestion of foreign particles within phagolysosomes, the large size of many urate crystals results in incomplete (“frustrated”) phagocytosis of the crystals, with the result that lysosomal contents are released into an unsealed phagolysosome and can escape to promote extracellular tissue damage (“regurgitation during feeding”). This process appears to be facilitated by crystals coated with immunoglobulins. Alternatively, when uncoated crystals are phagocytosed, their ability to interact with the cholesterol-rich bilayers of the phagolysosome results in phagolysosome rupture, cellular necrosis, and direct release of toxic neutrophil contents. In either case, the result is the potential for tissue damage and rapidly accelerating inflammation.

Resolution of the Acute Gouty Attack

One of the more interesting features of the acute gouty attack is that, particularly in the early years of the disease, most such attacks are self-limited. Multiple effects have been invoked to explain this phenomenon. Urate crystals may become coated with synovial fluid proteins (e.g., apolipoproteins B and E), inhibiting their ability to provoke inflammation. In this context, it is worth noting that between acute attacks, some crystals may be present in the joint despite the absence of inflammation, suggesting the potential for such crystals to become inactive. Additional research has been focused on the clearance of crystals from the joint during the resolution phase of acute attacks, when the joint fluid crystal burden typically falls. Some authors have emphasized the ability of macrophages to clear urate crystals and of macrophages to clear apoptotic neutrophils that have ingested crystals, resulting in decreased crystal burden. The enzymatic products of inflammatory cells may degrade proinflammatory cytokines, and persistent receptor stimulation may result in receptor downregulation. The stress of the gouty attack may promote adrenocorticotropic hormone (ACTH) secretion; in addition to its ability to induce glucocorticoid release, ACTH can bind directly to melanocyte-stimulating hormone receptors to provoke anti-inflammatory effects.

More recent investigations have emphasized the local, active events through which inflammation may be resolved. For example, in vivo studies using the urate crystal-induced mouse air pouch model of inflammation suggest that local

**Figure 94-12** Propagation of the acute gouty response by activated neutrophils. Neutrophils that enter the joint migrate toward and phagocytose crystals. In the case of crystals coated with immunoglobulins and complement, the resultant activation results in synthesis and/or release of inflammatory mediators such as interleukin (IL)-1β, IL-8, and tumor necrosis factor, as well as proteases and reactive oxygen species. In the case of uncoated crystals, the crystal frequently interacts with, and lyses the membrane of the phagolysosome, spilling toxic contents and leading to cell lysis. In both cases, the result is local tissue damage and recruitment of additional neutrophils from the bloodstream in an explosive inflammatory cycle. LTB4, leukotriene B4.
upregulation of expression of PGD₂ and 15d-PGJ₂ (a spontaneous dehydration product of PGD₃) contributes to the resolution of crystal-mediated inflammation, probably through the ability of these molecules to activate the peroxisome proliferator activating receptor-γ (PPAR-γ).³,²⁻¹⁴ After crystals also stimulate the expression of PPAR-γ itself, setting the stage for enhanced suppression of inflammation. 15d-PGJ₂ inhibits secretion of IL-1β, IL-6, and IL-12, as well as TNF from macrophages, and downregulates the expression of inducible nitric oxide synthase. Other targets of 15d-PGJ₂ include PGD receptors and the NFκB pathway. 15d-PGJ₂ also inhibits CXC chemokine production, alters cellular adhesion molecules, and stimulates apoptosis of endothelial cells. Other anti-inflammatory molecules (e.g., IL-10) are also upregulated, and the more recently identified and potent anti-inflammatory resolvins and lipoxins are almost certainly involved.²⁵

In considering the resolution of acute gouty inflammation, it is important to recognize the biologic programs that delay the production of anti-inflammatory molecules until the appropriate phase in the sequence. For example, PGD₂ and 15d-PGJ₂ production initially declines on exposure to urate crystals, apparently facilitating the inflammatory response, but then rebounds during the resolution phase.²⁶ Similarly, the generation of anti-inflammatory lipoxins requires the accumulation of at least two active cell types (either neutrophils and activated endothelium, or neutrophils and activated platelets), an evolutionary adaptation that creates an intrinsic delay before the anti-inflammatory effects are initiated.²⁷ Finally, a late-phase, anti-inflammatory release of TGF-β by macrophages that have phagocytosed urate crystals appears to require cellular maturation because monocytes that are newly arrived to the gouty joint generate proinflammatory rather than anti-inflammatory mediators.

**Chronic Gouty Arthritis and Tophaceous Gout**

The natural history of gout may include eventual progression to a state characterized by chronic inflammation and/or the establishment of macroscopic urate deposits known as tophi. Even during the asymptomatic, intercritical periods of gout, low-grade chronic inflammation can persist, with continuous phagocytosis of crystals by leukocytes.²⁸ In patients with longstanding gout, intercritical inflammation can become frankly apparent and the cytokines, chemokines, proteases, and oxidants that participate in acute inflammation can contribute to chronic synovitis, cartilage loss, and bone erosion, signaling the progression to chronic gouty arthritis.²⁹

Although tophi are composed primarily of MSU crystals, they are complex structures in which urate (1) is embedded in a matrix of lipids, proteins, and mucopolysaccharides and (2) drives a persistent inflammatory state.³⁷ A tophus can alternatively be conceived of as a granuloma of mononucleated and multinucleated macrophages arranged in three distinguishable zones. MSU crystals and debris constitute the central zone. Surrounding the central zone is the corona zone, consisting of macrophages, mast cells, and plasma cells. This biologically active corona is thought to account for the ultrasonographic finding of an anechoic rim that circumscibes a tophus.³⁰ Eventually, the corona and central zones may become encased by a connective tissue layer, the fibrovascular zone. Within this zone, macrophages express surface markers of recent migration, maturation, apoptosis, continuous recruitment, and proinflammatory activation.²¹,²²

Tophi are important not only for their role as reservoirs of crystalline urate but also for their ability to damage the tissues in which they reside (see Figure 95-6). Mechanical factors, induction of lytic enzymes, and synthesis of proinflammatory cytokines all contribute to the tophus’s ability to promote erosion and joint destruction.³⁶ Tophus macrophages produce IL-1β, TNF, IL-6, IL-17, matrix metalloproteinases (MMP)-2, and MMP-9, and macrophage colony-stimulating factor (M-CSF).²¹ These molecules promote further inflammation and tissue damage, as well as promote the maturation and activation of osteoclasts that actively resorb bone. For example, M-CSF interacts with M-CSF receptors on osteoblast progenitor cells to promote osteoclastogenesis. Tophus expression of IL-1β and other cytokines has been shown to decrease the anabolic effects of osteoblasts by decreasing the 1,25-dihydroxyvitamin D₃-dependent activity of alkaline phosphatase and osteocalcin.²⁷ Both IL-1β and TNF may directly promote bone erosions through the elucidation of matrix metalloproteinases and may also upregulate the RANK/RANK ligand system, the major promoter of osteoclastogenesis and osteoclast activation. Additionally, MSU crystals on the cartilage surface activate chondrocytes to release IL-1β, nitric oxide, and MMPs, causing further joint damage and cartilage destruction. Even in asymptomatic patients, the inflammatory and erosive processes persist; damage may not be apparent until late in the process.

**Nongout Effects of Hyperuricemia**

**KEY POINT**

Even in the absence of gout, hyperuricemia and soluble serum urate are biologically active and may have previously unappreciated clinical effects.

In addition to its roles in acute and chronic gouty arthritis, hyperuricemia may have other adverse as well as beneficial effects. Several investigators have demonstrated that soluble urate is biologically active, with effects on renal and vascular function. Among its mechanisms of action, soluble urate inhibits synthesis of the potent vasodilator nitric oxide; induces smooth muscle cell proliferation by activating mitogen-activated protein kinases; and stimulates cyclooxygenase-2 and platelet-derived growth factor synthesis, all contributing to arterial vasoconstriction.²¹,²² Soluble urate has also been shown to directly stimulate the renin-angiotensin system in the kidney and to induce renal interstitial and tubular inflammation.²³,²⁴ As noted earlier (see Evolutionary Considerations), these effects may promote hypertension; other studies suggest that hyperuricemia may also contribute to the risk for both renal insufficiency and myocardial infarction, though additional research is necessary to better assess the direction of causality.²⁵ Several recent clinical trials suggest that lowering serum urate levels may reduce the risk of myocardial
infarction and slow the progression of renal failure.\textsuperscript{218,219} Hyperuricemia may also promote insulin resistance in adipose cells, potentially serving as a risk factor for diabetes and metabolic syndrome.\textsuperscript{220,221} Urate may also play a role and/or serve as a biomarker in osteoarthritis (OA).\textsuperscript{14,222} In contrast to the adverse effects of serum urate, accumulating evidence suggests that hyperuricemia may protect against neurologic diseases such as dementia, multiple sclerosis, and Parkinson’s and Huntington’s diseases.\textsuperscript{223-226} Recognition of uric acid’s biologic complexity will likely lead to a better understanding of its impact on the immune, cardiovascular, endocrine, neurologic, and musculoskeletal systems.

**Selected References**


Full references for this chapter can be found on www.expertconsult.com.
References


6.8 mg/dL at 37° C. Thus a value greater than 6.8 mg/dL indicates supersaturation of body fluids. The serum urate concentration is elevated in a relative sense when it exceeds the upper limit of an arbitrary normal range, which is usually defined as the mean serum urate value plus two standard deviations in a sex- and age-matched healthy population. In most epidemiologic studies, the upper limit has been rounded off at 7 mg/dL in men and 6 mg/dL in women. A serum urate value in excess of 7 mg/dL begins to carry an increased risk of gouty arthritis or renal stones.

**EPIDEMIOLOGY**

Hyperuricemia is fairly common, with prevalence ranging between 2.6% and 47.2% in various populations. A variety of factors appears to be associated with high serum urate concentrations. In adults, serum urate levels correlate strongly with the serum creatinine and urea nitrogen levels, body weight, height, age, blood pressure, and alcohol intake. In epidemiologic studies, body bulk (as estimated by body weight, surface area, or body mass index) has proved to be one of the most important predictors of hyperuricemia in people of many different races and cultures, with rare exceptions.

Serum urate concentrations vary with age and sex. Children normally have a concentration in the range of 3 to 4 mg/dL because of high renal uric acid clearance. At puberty, serum urate concentrations increase by 1 to 2 mg/dL in males, and this higher level is generally sustained throughout life. In contrast, females exhibit little change in the serum urate concentration until menopause, when concentrations increase and approach those seen in adult men. The mechanism of lower serum urate levels in women is a consequence of sex hormones and is related to a higher fractional excretion of urate secondary to lower tubular urate postsecretory reabsorption.

The incidence of gout varies among populations, with an overall prevalence ranging from less than 1% to 15.3%.

**KEY POINTS**

- Hyperuricemia is common and directly associated with serum creatinine, body mass index, age, blood pressure, and alcohol intake.
- Serum urate levels are low in childhood and increase in men at puberty and in women at menopause.
- The prevalence of gout ranges from 1% to 15% in populations with a clear increase in incidence in recent years, perhaps related to the assumption of a Western diet and the epidemic of obesity.

Gout has been called the “king of diseases” and the “disease of kings.” Today, the term gout is used to represent a heterogeneous group of diseases found exclusively in humans that include the following characteristics:

- Elevated serum urate concentration (hyperuricemia)
- Recurrent attacks of acute arthritis in which monosodium urate monohydrate crystals are demonstrable in synovial fluid leukocytes
- Aggregates of sodium urate monohydrate crystals (tophi) deposited chiefly in and around joints, which sometimes lead to deformity and crippling
- Renal disease involving glomerular, tubular, and interstitial tissues and blood vessels
- Uric acid nephrolithiasis

These manifestations can occur in various combinations. Hyperuricemia denotes an elevated level of urate in the blood. This occurs in an absolute (or physiochemical) sense when the serum urate concentration exceeds the limit of solubility of monosodium urate in the serum, which is
This upper limit appears to be increasing. The prevalence increases substantially with age and with increasing serum urate concentration. The annual incidence rate of gout is 4.9% for urate levels greater than 9 mg/dL, 0.5% for values between 7 and 8.9 mg/dL, and 0.1% for values less than 7 mg/dL. For serum urate values greater than 9 mg/dL, the cumulative incidence of gout reaches 22% after 5 years.

**ENVIRONMENTAL FACTORS**

**KEY POINTS**

- Many environmental factors are associated with gout including alcohol consumption, particularly beer, and diet.
- Certain foods clearly promote hyperuricemia and gout including alcohol, seafood, and red meat.
- The consumption of some foods may be protective, especially milk and yogurt.

An association between alcohol consumption and gout has been recognized for centuries. The risk of developing gout varies by the type of alcohol ingested. Beer, which is purine rich, carries the highest risk; this risk is substantially greater than that for liquor. Moderate wine drinking does not increase the risk of gout. The quantity of alcohol also strongly correlates with gout. Compared with men who did not consume alcohol, the relative risk of gout was 1.32 for an alcohol intake of 10 to 14.9 g/day, 1.49 for 15 to 29.9 g/day, 1.96 for 30 to 49.9 g/day, and 2.53 for 50 g/day and higher, with calculations based on 12.8 g of alcohol per 12 oz serving of regular beer, 11.3 g per 12 oz serving of light beer, 11 g per 4 oz serving of wine, and 14 g per shot of liquor.

Diet also influences hyperuricemia and gout. Serum urate levels increase with meat or seafood intake and decrease with dairy intake. Men in the highest quintile of seafood consumption have a 51% higher risk of developing gout, and those in the highest quintile of meat intake have a 41% higher risk. However, consumption of oatmeal and purine-rich vegetables (e.g., peas, mushrooms, lentils, spinach, cauliflower) is not associated with an increased risk for gout. The consumption of milk one or more times a day or yogurt consumption at least once every other day is associated with lower serum urate levels.

**GENETICS**

**KEY POINTS**

- Rare forms of early hyperuricemia and gout have a clear genetic and metabolic basis.
- Gout often runs in families, probably because of inherited factors affecting serum urate levels through renal urate clearance.
- Recent genome-wide association studies have identified polymorphisms in several candidate genes encoding urate transporters in the renal proximal tubules as determinants of serum urate levels and the risk of gout.

Since antiquity, gout has been recognized as a familial disorder. The familial incidences reported range from 11% to 80%. In two large series, one English and one American, about 40% of gouty subjects gave a positive family history of gout. These wide discrepancies may be attributed in part to variations in diligence and pursuit of genealogic data. When all available data are considered, they suggest that serum urate concentrations are controlled by polygenic traits. Several rare forms of hyperuricemia and gout such as hypoxanthine phosphoribosyltransferase deficiency, phosphoribosyl-1-pyrophosphate synthetase overactivity, and familial hyperuricemia nephropathy have a clear genetic basis, most presenting in childhood or early adulthood (see later).

But the vast majority of gouty patients do not have one of these discrete inborn errors of metabolism as explanations for their disease. In most gout patients, the mechanism of hyperuricemia is simply inefficient renal excretion of uric acid. It stands to reason then that some of the familial predisposition to gout is related to inherited variations in renal urate handling. With the recent advances in our understanding of the transport of uric acid in the renal proximal tubule, some of this variability is being clarified. Genome-wide association studies (GWAS) have identified genetic variations of the SLC2A9/GLUT9 and ABCG2 genes as important determinants of serum urate levels (see Chapter 94). Curiously, the data for the role of polymorphisms in URAT1 (SLC22A12), a key urate transporter involved in renal proximal tubule urate reabsorption, are thus far inconsistent. Glucose transporter 9 (GLUT9, SLC2A9) is an electrogenic hexose transporter whose splicing variants mediate reabsorption of uric acid, along with glucose and fructose, at the renal proximal tubule epithelial cell, first at the apical membrane, then through the basolateral membrane, and on into the circulation. Polymorphisms in SLC2A9 are associated with a lower serum uric acid, with stronger effects seen in women, possibly accounting for 0.5% to 2% of the variance in serum urate concentration in men and 3.4% to 8.8% in women. In five of six populations studied by GWAS, SLC2A9 polymorphisms were associated with a decreased incidence of self-reported gout, particularly in women. Indeed, three patients with renal hypouricemia have been identified with loss-of-function mutations in SLC2A9. The implication is that certain polymorphisms in SLC2A9 enhance renal uric acid excretion, lower serum urate levels, and are therefore protective against gout, particularly in women.

A gene product of ABCG2, human adenosine triphosphate (ATP)-binding cassette, subfamily G, 2, appears to be a secretory urate transporter located in the renal proximal tubule apical border. Polymorphisms in ABCG2 are associated with a higher serum urate concentration in humans, with a more potent influence in men, potentially accounting for 1.6% to 2.1% of the variance in serum urate concentration in men and 0.5% to 0.8% in women. One ABCG2 polymorphism, rs2231142, has been associated with a higher incidence of gout only in males (odds ratio [OR], 2.03 for self-reported gout). In a transfection model in Xenopus oocytes, the common mutation in human ABCG2 encoded by rs2231142, Q141K, resulted in a 53% decrease in urate transport compared with the wild-type...
gene product. The implication is that certain polymorphisms in ABCG2 result in reduced proximal tubule urate secretion, increased serum urate concentrations, and a higher incidence of gout, at least in men.

Because the impact of comorbidities and environmental influences such as diet and lifestyle is so significant in gout, it remains to be seen whether risk contribution from genetic variation is clinically relevant, either directly or due to an unexpected interaction with these other factors. An example of the latter is the potential influence the urate-fructose/glucose exchange transporter SCL2A9 polymorphisms may have on the well-described association of heavy consumption of soft drinks with elevated serum urate levels and gout.

### CLINICAL FEATURES

<table>
<thead>
<tr>
<th>KEY POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>The three stages of gout are asymptomatic hyperuricemia, acute and intercritical gout, and chronic gouty arthritis.</td>
</tr>
<tr>
<td>A period of asymptomatic hyperuricemia lasts up to 20 years before the initial attack of gout or nephrolithiasis.</td>
</tr>
<tr>
<td>The first gout attack generally occurs at age 40 to 60 in men and after age 60 in women.</td>
</tr>
<tr>
<td>Many drugs raise serum urate levels and predispose to gout attacks, especially diuretics.</td>
</tr>
<tr>
<td>Most attacks of gout, especially early in the course, are monoarticular, with a predilection for the first metatarsophalangeal joint (podagra) and have a characteristic abrupt and painful onset.</td>
</tr>
<tr>
<td>The differential diagnosis for acute gout is usually infectious arthritis or other crystal-induced synovitis, particularly pseudogout.</td>
</tr>
<tr>
<td>Ultrasonography appears to be a useful adjunct in the diagnosis of acute and chronic gout.</td>
</tr>
<tr>
<td>In untreated or undertreated individuals, chronic gout is characterized by the development of tophi and progressive joint damage.</td>
</tr>
</tbody>
</table>

Throughout its natural history, gout passes through three stages: (1) asymptomatic hyperuricemia, (2) episodes of acute gouty arthritis separated by asymptomatic intervals (termed intercritical or interval gout), and (3) chronic gouty arthritis, the period when tophi often become apparent.

The basic pattern of clinical gout begins with acute attacks of intensely painful arthritis. The first attack is usually monoarticular and associated with few constitutional symptoms. Later, attacks may become polyarticular and are associated with fever. Attacks vary in duration but are time limited. Over time, attacks recur at shorter intervals, last longer, and eventually resolve incompletely. This leads to the development of chronic arthritis that slowly progresses to a crippling disease on which acute exacerbations are superimposed.

### Asymptomatic Hyperuricemia

Asymptomatic hyperuricemia is a condition in which the serum urate level is high, but gout—manifested by arthritis or uric acid nephrolithiasis—has not yet occurred. Most people with hyperuricemia remain asymptomatic throughout their lifetimes. The tendency toward acute gout increases with the serum urate concentration. The risk of nephrolithiasis increases with the serum urate level and with the magnitude of urinary uric acid excretion. The phase of asymptomatic hyperuricemia ends with the first attack of gouty arthritis or urolithiasis. In most instances, this occurs after at least 20 years of sustained hyperuricemia. Between 10% and 40% of gouty subjects have one or more attacks of renal colic before the first articular event.

### Acute Gouty Arthritis

The first attack of acute gouty arthritis usually occurs between age 40 and 60 years in men and after age 60 in women. Onset before age 25 should raise the possibility of an unusual form of gout, perhaps one related to a specific enzymatic defect that causes marked purine overproduction, an inherited renal disorder, or the use of cyclosporine.

A single joint is involved in about 85% to 90% of first attacks, with the first metatarsophalangeal joint being the most commonly affected site. The initial attack is polyarticular in 3% to 14%. Acute gout is predominantly a disease of the lower extremities, but eventually, any joint of any extremity may be involved. Ninety percent of patients experience acute attacks in the great toe at some time during the course of their disease. Next in order of frequency are the insteps, ankles, heels, knees, wrists, fingers, and elbows. Acute attacks rarely affect the shoulders, hips, spine, sacroiliac joints, sternoclavicular joints, acromioclavicular joints, or temporomandibular joints. Acute gouty bursitis, tendinitis, or tenosynovitis can also occur.

Urate deposition and subsequent gout appear to have a predilection for previously damaged joints such as in Heberden’s nodes of older women. The differential diagnosis is usually septic arthritis or other crystal-induced arthritis, but a broader differential should be considered in confusing cases (Table 95-1).

Some patients report a history of short, trivial episodes of “ankle sprains,” sore heels, or twinges of pain in the great toe before the first dramatic gouty attack. In most patients, however, the initial attack occurs with explosive suddenness and commonly begins at night after the individual has gone to sleep feeling well. Within a few hours of onset, the affected part becomes hot, dusky red, swollen, and extremely tender. Occasionally, lymphangitis may develop. Systemic signs of inflammation may include leukocytosis, fever, and elevation of the erythrocyte sedimentation rate. Radiographs usually show only soft tissue swelling during early episodes.

The course of untreated acute gout is highly variable. Mild attacks may subside in several hours or persist for only a day or two and never reach the intensity described for the classic attack. Severe attacks may last days to weeks. The skin over the joint often desquamates as the erythema subsides. With resolution, the patient becomes asymptomatic and enters the intercritical period.
The use of ultrasonography as a means of diagnosing acute and chronic gout is gaining favor. The characteristic finding is a superficial, hyperechoic, irregular band on the surface of articular cartilage, the so-called “double contour sign” or “urate icing,” in one study seen in 92% of gouty joints and in no joints of patients with other types of arthropathy. Also characteristic is nonhomogeneous topophageous material surrounded by an anechoic rim. Further support for ultrasound comes from the demonstration of resolution of these findings in a small number of gouty patients on urate-lowering therapy who continued a serum urate less than or equal to 6 mg/dL for at least 7 months. Recent studies presented in abstract form have indicated excellent concordance between ultrasound readers in identifying changes, the presence of these findings even in those with asymptomatic hyperuricemia, and the superiority of ultrasound over conventional radiography in detecting gouty erosions. Ultrasound now appears to be a technology that will be an important ancillary approach to the diagnosis and treatment of gout. Magnetic resonance imaging is much more sensitive than even ultrasound at detecting gouty erosions in patients with gout and normal plain radiographs. Computed tomography (CT) scan is also sensitive for the detection of erosions and tophi, and threedimensional CT may have utility in the quantitation of the size of tophi during clinical trials in gout.

Intercritical Gout

The terms intercritical gout and interval gout have been applied to the periods between gouty attacks. Some patients never have a second attack. However, most patients suffer a second attack within 6 months to 2 years. In Gutman’s series, 62% had recurrences within the first year, 16% in 1 to 2 years, 11% in 2 to 5 years, and 4% in 5 to 10 years; 7% had experienced no recurrence in 10 or more years. The frequency of gout attacks usually increases over time in untreated patients. Later attacks have a less explosive onset, are polyarticular, become more severe, last longer, and abate more slowly. Nevertheless, recovery is complete.

### Table 95-1 Differential Diagnosis of Gout

<table>
<thead>
<tr>
<th>Acute Gouty Arthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other crystal arthritis, especially CPPD, or pseudogout, but also basic calcium phosphate (hydroxyapatite) and others</td>
</tr>
<tr>
<td>Septic arthritis including gonorrhea</td>
</tr>
<tr>
<td>Trauma</td>
</tr>
<tr>
<td>Cellulitis</td>
</tr>
<tr>
<td>Lyme arthritis</td>
</tr>
<tr>
<td>Reactive arthritis</td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
</tr>
<tr>
<td>Sarcoidosis</td>
</tr>
<tr>
<td>Unusual presentations of other inflammatory arthropathies including rheumatoid arthritis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chronic Gouty Arthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatoid or other chronic inflammatory arthritis</td>
</tr>
<tr>
<td>CPPD (Inflammatory) osteoarthritis</td>
</tr>
<tr>
<td>Lyme disease</td>
</tr>
<tr>
<td>Indolent infections including mycobacterial</td>
</tr>
</tbody>
</table>

CPPD, calcium pyrophosphate disease.

Drugs may precipitate acute gout by either increasing or decreasing serum urate levels acutely. The occurrence of gout after the initiation of antihyperuricemic therapy is well established. In fact, the more potent the urate-lowering effect, the more likely there is to be an acute attack. Drug-induced gout secondary to increased serum urate levels occurs on occasion with diuretic therapy, intravenous heparin, and cyclosporine. Diuretic therapy in the elderly appears to be a particularly important precipitating factor for gouty arthritis. Other provocative factors include trauma, alcohol ingestion, surgery, dietary excess, hemorhage, foreign protein therapy, infections, and radiographic contrast exposure. The risk of a patient with gout developing an attack during hospitalization is 20%. The definitive diagnosis of gout is best established by aspiration of the joint and identification of intracellular needle-shaped crystals that have negative birefringence with compensated polarized light microscopy. However, various alternatives have been proposed for a presumptive diagnosis. These include the triad of acute monoarticular arthritis, hyperuricemia, and a dramatic response to colchicine therapy, a set of criteria proposed by the American College of Rheumatology (Table 95-2) in 1977, and 10 “propositions” for diagnosis by a European League Against Rheumatism (EULAR) panel of experts in 2006 (Table 95-3). There are limitations to using any of these schemes. First, although the diagnosis of acute gouty arthritis can be strongly suggested by the typical presentation, not all inflammation of the great toe (podagra) in hyperuricemic patients is caused by gout. Second, some patients with gout are normouricemic at the time of an acute attack, a phenomenon related to alcohol use or a consequence of interleukin (IL)-6 generation by the acute inflammatory process. Third, diseases other than gout can occasionally improve with colchicine therapy; these include pseudogout, hydroxyapatite calcific tendinitis, sarcoid arthritis, erythema nodosum, serum sickness, rheumatoid arthritis, and familial Mediterranean fever. Finally, the simultaneous presence of both gout and septic arthritis can be confusing clinically, with the former masking the latter.
Radiographic changes may develop during the intercritical period despite no sign of tophi on physical examination. These changes are more likely in patients with more severe hyperuricemia and more frequent acute attacks. The diagnosis of gout in a hyperuricemic patient with a history of acute attacks of monoarthritis may be difficult or inconclusive during the intercritical period. Aspiration of an asymptomatic joint, however, can be a useful adjunct in the diagnosis of gout if urate crystals are demonstrated. Joint fluids obtained from gouty patients during the intercritical phase revealed monosodium urate crystals in 12.5% to 90% of joints. Such crystals in asymptomatic joints are often associated with mild synovial fluid leukocytosis, which suggests the potential to contribute to joint damage even in the intervals between attacks.

Chronic Gouty Arthritis

Eventually, the patient may enter a phase of chronic polyarticular gout with no pain-free intercritical periods. At this stage, gout may be easily confused with other types of arthritis or other conditions. The time from the initial attack to the beginning of chronic symptoms or visible tophaceous involvement is highly variable in studies of untreated patients. Hensch reported intervals ranging from 3 to 42 years, with an average of 11.6 years between the first attack and the development of chronic arthritis. Ten years after the first attack, about half the individuals were still free of obvious tophi and most of the remainder had only minimal deposits. Thereafter, the proportion of those with nontophaceous involvement slowly declined, to 28% after 20 years. Two percent of the patients had severe crippling disease some 20 years after the initial attack.

The rate of formation of tophaceous deposits correlates with both the degree and the duration of hyperuricemia. The principal determinant is the serum urate level. Gutman found the mean serum urate concentration to be 9.1 mg/dL in 722 patients without tophi, 10 to 12 mg/dL in 456 patients with minimal to moderate tophi, and greater than 11 mg/dL in 11 patients with extensive tophaceous involvement. The rate of tophus formation also increases with the severity of renal disease and the use of diuretics. Tophaceous gout is the consequence of the chronic inability to eliminate urate as rapidly as it is produced. As the urate pool expands, deposits of urate crystals appear in cartilage, synovial membranes, tendons, soft tissues, and elsewhere. Tophi are rarely present at the time of an initial attack of primary gout; they are more likely to be present in gout secondary to myeloproliferative diseases, in juvenile gout-complicating glycogen storage diseases (GSDs), in Lesch-Nyhan syndrome, or after allograft transplantation in patients treated with ciclosporin.

Tophi can occur in a variety of locations. Tophaceous deposits may produce irregular, asymmetric, moderately discrete tumeescence of the fingers (Figure 95-1), hands, knees, or feet. Tophi also form along the ulnar surfaces of the forearm, as saccular distentions of the olecranon bursa (Figure 95-2), in the antihelix of the ear (Figure 95-3), or as fusiform enlargements of the Achilles tendon (Figure 95-4). The process of tophaceous deposition advances insidiously. Although the tophi themselves are relatively painless, acute inflammation can occur around them. Eventually, extensive destruction of the joints and large subcutaneous tophi may lead to grotesque deformities, particularly of the hands and feet, and to progressive crippling (Figure 95-5).
Figure 95-1  Tophus of the fifth digit, with a smaller tophus over the fourth proximal interphalangeal joint.

Figure 95-2  Saccular tophaceous enlargements of the olecranon bursae, with small cutaneous deposits of urate.

Figure 95-3  Tophus of the helix of the ear adjacent to the auricular tubercle.

Figure 95-4  Tophi of Achilles tendons and their insertions in a patient with gout.

Figure 95-5  A and B, Radiographs demonstrating severe destructive changes in tophaceous gout.
ASSOCIATED CONDITIONS

KEY POINTS

- Gout is associated with obesity, hypertriglyceridemia, glucose intolerance and the metabolic syndrome, hypertension, atherosclerosis, and hypothyroidism.
- Renal insufficiency is frequently associated with hyperuricemia and gout.
- Hyperuricemia is a common cause of nephrolithiasis, and rarely, chronic hyperuricemia may cause urate nephropathy and acute hyperuricemia may lead to uric acid nephropathy in the tumor lysis syndrome.
- Alcohol use, lead intoxication, and cyclosporine treatment are associated with hyperuricemia and gout.
- A diagnosis of gout should prompt a search for the coexistence of these associated conditions.

The association of gout with obesity and overeating is well recognized. In 6000 subjects, hyperuricemia was found in only 3.4% of those with a relative weight at or below the 20th percentile, in 5.7% of those between the 21st and 79th percentiles, and in 11.4% of those at or above the 80th percentile.

Hypertriglyceridemia has been reported in 75% to 80% of patients with gout, and hyperuricemia is found in more than 80% of patients with hypertriglyceridemia. However, studies have been unable to show a correlation between serum urate and cholesterol values or a unique lipid phenotype. Gouty patients who drink alcohol excessively have mean serum triglyceride levels that are higher than those of their obesity-matched controls and of non-alcohol-drinking gouty patients.

Hyperuricemia has been reported in 2% to 50% of patients with diabetes mellitus, and gouty arthritis has been reported in less than 0.1% to 9%. Abnormal glucose tolerance tests have been noted in 7% to 74% of patients with gout, depending, in part, on the criteria used. Hyperuricemia may be an indication of a potential risk for hypertension in adolescent males. Hypertension is present in one-fourth to one-half of patients with classic gout, but the presence of hypertension is unrelated to the duration of gout. Elevated serum urate concentrations are associated with increased tubular reabsorption of sodium. The serum urate concentration also correlates inversely with renal blood flow and urate clearance and correlates directly with both renovascular and total resistance. Therefore the association between hypertension and hyperuricemia may be related to the reduction of renal blood flow in hypertension. In addition, uric acid causes smooth muscle proliferation in vitro and vascular disease in animal models through a mechanism that involves complex intracellular signaling, mitogen-activated protein kinase activation, and platelet-derived growth factor expression.
The association between hyperuricemia and the manifestations of atherosclerosis has led to speculation that hyperuricemia is a risk factor for coronary artery disease. Some studies show no clear associations among blood pressure, blood glucose, or serum cholesterol and serum urate concentration when adjustments are made for the effects of age, sex, and relative weight; the serum urate concentrations of persons with coronary heart disease are not significantly different from the mean levels of the population. Other studies, however, maintain that hyperuricemia is an independent risk factor for coronary artery disease.

The term metabolic syndrome has been applied to a cluster of abnormalities including resistance to insulin-stimulated glucose uptake, hyperinsulinemia, hypertension, and dyslipoproteinemia that are characterized by high levels of plasma triglycerides and high-density lipoprotein cholesterol. Hyperuricemia closely correlates with the degree of insulin resistance and therefore is a likely feature of metabolic syndrome. Metabolic syndrome has been associated with coronary artery disease, and hyperuricemia as a component of metabolic syndrome may explain the previously recognized association between coronary artery disease and hyperuricemia. A recent study concluded that the relationship between hyperuricemia and acute myocardial infarction is independent, but patients who experience gouty arthritis are at an increased risk for myocardial infarction. This association could not be explained by renal function, metabolic syndrome, diuretic use, or traditional cardiovascular risk factors.

Alcohol consumption has long been associated with hyperuricemia and gout. In susceptible persons, alcohol use can precipitate acute gouty arthritis. An epidemiologic study in Saudi Arabia, where alcohol consumption is rare, revealed an 8.42% prevalence of hyperuricemia but no cases of gout among the study group. Both a decrease in the renal excretion of uric acid and an increase in uric acid production seem to be important factors in this association. Ethanol increases uric acid production by accelerating the turnover of ATP. Among alcoholic beverages, beer may have more potent effects on uric acid production because of its high guanosine content.

There appears to be a significant increased prevalence of hypothyroidism among both female and male patients with gouty arthritis. Hyperuricemia may also be more prevalent in patients with hypothyroidism. Thyroid replacement therapy is associated with a decrease in serum urate concentration caused by an increased uric acid diuresis—a change not explained solely by a change in creatinine clearance. Although the cause of hyperuricemia and gout in patients with hypothyroidism is unknown, it is speculated that urate metabolism is mediated by thyroid-stimulating hormone receptors in extrathyroidal tissues including the kidney and that these modulate urate homeostasis.

Studies of acutely ill patients in intensive care units indicate that markedly increased serum urate concentrations, in the vicinity of 20 mg/dL, are associated with hypertensive events and a poor prognosis. This finding may be related to two factors. First, ischemic tissue may foster the degradation of ATP to purine end products, thereby enhancing the production of urate. The finding of increased plasma ATP degradation products associated with hyperuricemia and adult respiratory distress syndrome supports this possibility. Second, the conversion of hypoxanthine to uric acid by xanthine oxidase during ischemia produces oxidant radicals, which are themselves associated with tissue injury. It is possible that inhibition of xanthine oxidase with allopurinol may be a useful therapy in this setting.

Maternal serum urate concentrations normally decrease during pregnancy until the 24th week and then increase until 12 weeks after delivery. An increase in the serum urate level occurs in preeclampsia and toxemia of pregnancy, owing to a decrease in the renal clearance of urate. Perinatal mortality is markedly increased when maternal plasma urate levels are raised, usually in association with early-onset preeclampsia. The highest mortality rate is seen with serum urate concentrations higher than 6 mg/dL and diastolic blood pressures greater than 110 mm Hg. Labor itself is associated with an increased serum urate level, and it remains elevated for 1 to 2 days after delivery.

Gout is rarely seen in patients with rheumatoid arthritis, systemic lupus erythematosus, or ankylosing spondylitis. The basis for the decreased concurrence of these disorders is unclear, although the long-term use of nonsteroidal anti-inflammatory drugs (NSAIDs) or corticosteroids may mask the clinical features of gout in some of these patients.

Renal Disease

After gouty arthritis, renal problems appear to be the most frequent complication of hyperuricemia. Twenty percent to 40% of patients with gout have albuminuria, which is usually mild and often intermittent. Hyperuricemia alone may be implicated as the cause of chronic kidney disease only when the concentration of urate chronically exceeds 13 mg/dL in men or 10 mg/dL in women. Before the routine treatment of asymptomatic hypertension, renal failure accounted for 10% of the deaths in patients with gout. Whether moderate hyperuricemia has a direct harmful effect on renal function is unclear. Some evidence suggests that urate damages the kidneys and leads to hypertension.

The term urate nephropathy is used to describe the deposition of urate crystals in the interstitium of the medulla and pyramids, with a surrounding giant cell reaction—a distinctive histologic finding characteristic of the gouty kidney (Figure 95-7). Factors such as coexistent hypertension, chronic lead exposure, ischemic heart disease, and primary pre-existing renal insufficiency probably play important roles in the pathogenesis of this pathology. Although urate nephropathy appears to exist as a distinct entity, it is not believed to be an important contributor to renal function in most gouty patients.

In contrast, uric acid nephropathy is the term used to describe acute renal failure resulting from the precipitation of large quantities of uric acid crystals in the collecting ducts and ureters. This complication most commonly occurs in patients with leukemia and lymphoma as a result of rapid malignant cell turnover, often during chemotherapy. This syndrome (also termed acute tumor lysis syndrome) has...
been more clearly defined as hyperuricemia, lactic acidosis, hyperkalemia, hyperphosphatemia, and hypocalcemia and is most commonly observed in patients with aggressive, rapidly proliferating tumors including lymphoproliferative disorders and metastatic medulloblastoma. Uric acid nephropathy is less commonly found with other neoplasms, after epileptic seizures, after vigorous exercise with heat stress, and after angiography and coronary artery bypass surgery.\(^{16}\)

In the tumor lysis syndrome, the large amount of nucleic acid in nucleotides liberated with massive cytolysis is converted rapidly to uric acid. Typically, there is marked hyperuricemia, with a mean serum urate level of 20 mg/dL (range, 12 to 80 mg/dL). The pathogenesis of acute renal failure in uric acid nephropathy is related to the precipitation of uric acid in the distal tubules and collecting ducts, the sites of maximal acidification and concentration of urine. Oliguria, or even anuria, and azotemia may occur. There may be “gravel” or “sand” noted in the urine. The ratio of urinary uric acid to creatinine in these patients typically exceeds 1; in patients with most other causes of acute renal failure, the ratio is 0.4 ± 0.3.\(^{120}\)

Nephrolithiasis occurs in 10% to 25% of patients with primary gout, a prevalence greater than that in the general population. The likelihood of stones in a given patient with gout increases with the serum urate concentration and with amounts of urinary uric acid excretion.\(^{121,122}\) It exceeds 50% with a serum urate value above 13 mg/dL or with urinary uric acid excretion rates in excess of 1100 mg every 24 hours.

Uric acid calculi account for approximately 10% of all stones in patients in the United States; elsewhere, rates range from as low as 5% up to 40% in Israel and Australia, respectively.\(^{16}\) Uric acid stones can occur in patients with no history of gouty arthritis, and only 20% in this group are hyperuricemic. Other renal stone disease is associated with hyperuricemia and gout. Gouty subjects also have an increased incidence of stones that contain calcium. In addition, about 30% of patients with recurrent calcium stone disease have either an increased urinary uric acid excretion rate or hyperuricemia. A causative link between uric acid and recurrent calcium oxalate stones is provided by reports of reduced stone frequency in patients treated with allopurinol.

Finally, the report of uric acid as the major constituent of a stone obtained from a patient with no apparent abnormalities of uric acid metabolism should suggest the possibility that the constituent is actually 2,8-dihydroxyadenine and that the patient has adenine phosphoribosyltransferase deficiency.\(^{123}\) This is because x-ray diffraction is required to distinguish uric acid from 2,8-dihydroxyadenine.

Familial juvenile hyperuricemic nephropathy (FJHN), sometimes called familial juvenile gouty nephropathy, was first described in 1960.\(^{124}\) This disorder is inherited as an autosomal dominant trait with a high degree of penetrance and is usually associated with gout. Renal disease typically develops in the second decade of life and progresses to end-stage renal failure by midlife.\(^{125,126}\) Histologic examination of kidney tissue reveals tubulointerstitial inflammation and splitting of thickened tubular basement membranes. The primary diagnostic criterion is a reduced fractional excretion of urate (defined as uric acid clearance factored by creatinine clearance × 100 equal to 5% or less; normal is 8% to 18%).\(^{127}\) The dramatically low fractional excretion of urate and the early onset of disease are conspicuous characteristics of FJHN and distinguish it from other autosomal dominant hyperuricemic disorders that usually appear later in life (Table 95-4). Genotype mapping has linked the gene for FJHN to chromosome 16p12-p11.\(^{128}\)

Autosomal dominant medullary cystic kidney disease (ADMCKD) is another hereditary nephropathy that usually includes gout among its constellation of symptoms. The

---

**Table 95-4  Genetics of Renal Diseases Associated with Gout**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Inheritance</th>
<th>Chromosomal Location</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>FJHN</td>
<td>AD</td>
<td>16p12.3</td>
<td>Uromodulin</td>
</tr>
<tr>
<td>MCKD type 1</td>
<td>AD</td>
<td>17cenq21.3</td>
<td>Hepatic nuclear factor 1β</td>
</tr>
<tr>
<td>MCKD type 2</td>
<td>AD, AR</td>
<td>1q21</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16p12.3</td>
<td>Uromodulin</td>
</tr>
</tbody>
</table>

AD, autosomal dominant; AR, autosomal recessive; FJHN, familial juvenile hyperuricemic nephropathy; MCKD, medullary cystic kidney disease.
onset of renal dysfunction occurs later than in those with FJHN. Renal histology reveals numerous corticomedullary and intramedullary cysts in the kidneys and increased medullary connective tissue. At least two loci appear to be responsible for ADMCKD. One, termed ADMCKD1, is located on chromosome 1; the other, ADMCKD2, is a 16p locus. ADMCKD2 and FJHN loci map to approximately the same region of chromosome 16.129

The chromosome 16p12 locus that harbors the candidate interval for FJHN and ADMCKD2 contains six candidate genes, including the uromodulin gene (UMOD).130-132 UMOD encodes the Tamm-Horsfall protein, a glycosylphosphatidylinositol-anchored glycoprotein localized to the thick ascending limb of the loop of Henle. Amorphous deposits of uromodulin are present in the renal interstitium of patients with medullary cystic kidney disease.133 Four different mutations in exon 4 have been identified in the UMOD gene. Because mutations in the same gene are responsible for both FJHN and ADMCKD, the two entities appear to be allelic variants in UMOD that cause decreased urinary concentrations of Tamm-Horsfall protein, with resulting hyperuricemia and progressive renal failure.134

**Lead Intoxication**

Hyperuricemia and gout are well-recognized complications of chronic lead intoxication, with the prevalence of gout in patients with plumbism ranging between 6% and 50%.135 Although a renal defect is recognized, it has not been well defined.136,137 Some patients with primary gout have increased blood lead levels compared with age- and sex-matched controls, despite the absence of a history of overt lead exposure.138 This suggests that occult chronic lead intoxication may play a causative role in some cases of primary gout (up to 36% of some gout populations).139 In addition, patients with gout who have renal impairment seem to have an increased quantity of mobilized lead compared with gouty patients with normal renal function.139 These observations suggest an important role for lead in the pathogenesis of gouty nephropathy.

**Cyclosporine-Induced Hyperuricemia and Gout**

Cyclosporine interferes with the renal excretion of uric acid. Hyperuricemia and gout occur with increased frequency among transplant recipients treated with cyclosporine and are even more common when diuretics are used concomitantly.138,140 However, serum urate levels do not correlate directly with cyclosporine levels or with the degree of hypertension or renal insufficiency. The onset of gout may occur soon after transplantation, with a mean of about 17 months. Gouty attacks may be typical and monarticular, or they may affect unusual sites such as the shoulder, hip, or sacroiliac joints. Polyarticular attacks and an accelerated course, with early development of tophi, may also be observed. Nephrolithiasis develops in about 3% of renal transplant patients. All calculi from azathioprine-treated patients are composed of calcium compounds, whereas 60% of the calculi from those treated with cyclosporine contain uric acid.141

---

**CLASSIFICATION OF HYPERURICEMIA AND GOUT**

**KEY POINTS**

Primary hyperuricemia and gout are caused by decreased renal uric acid excretion in more than 90% and overproduction of urate in less than 10% of affected patients.

Secondary hyperuricemia and gout are usually related to decreased renal urate clearance as a direct or indirect consequence of the primary disease process.

Four known specific inborn errors of purine metabolism with overproduction of urate account for less than 1% of cases of secondary hyperuricemia and gout.

The concentration of urate in body fluids is determined by the balance between production and elimination. Accordingly, hyperuricemia may be caused by an excessive rate of urate production, a decrease in the renal excretion of uric acid, or a combination of both events.

Hyperuricemia and gout may be classified as follows (Table 95-5):

- **Primary**: These cases appear to be innate, neither secondary to an acquired disorder nor the result of a subordinate manifestation of an inborn error that leads initially to a major disease unlike gout. Some cases of primary gout have a genetic basis; others do not.
- **Secondary**: These cases develop in the course of another disease or as a consequence of drug use.
- **Idiopathic**: In these cases, a more precise classification cannot be assigned.

Further subdivisions within each major category are based on the identification of overproduction, underexcretion, or both, as responsible for the hyperuricemia. Evidence of overproduction of urate is provided by determination of the 24-hour urinary uric acid excretion. For adults ingesting a purine-free diet, a total excretion of up to 600 mg/day is considered within the normal range.142 For patients on regular diets, a value in excess of 1000 mg/day is clearly abnormal and an indication of overproduction, and values between 800 and 1000 mg/day are considered borderline. It has been suggested that overproduction of uric acid can be assessed simply by determining the ratio of uric acid to creatinine in the urine or the $C_{urate}/C_{creatinine}$ ratio. However, comparison of these two ratios with the 24-hour urinary uric acid excretion reveals a poor correlation in most patients.143,144 Exceptions include patients with specific enzymatic deficiencies or with rapid cell lysis during chemotherapy for leukemia or lymphoma.

**Primary Gout**

Renal mechanisms are responsible for the hyperuricemia in most cases of gout. Genetic factors exert an important control in the renal clearance of urate.145 A careful comparison of uric acid clearances and excretion rates over a wide but comparable range of filtered loads of urate indicates that most gouty subjects have a lower ratio of urate-to-inulin clearance ($C_{urate}/C_{inulin}$ ratio) than do
The excretion rates and the capacity of the excretory mechanism for uric acid are the same for gouty subjects and nongouty individuals (see Figure 94-8). The excretion curve, however, is shifted. Gouty subjects require serum urate values 2 or 3 mg/dL higher than those of controls to achieve equivalent uric acid excretion rates. Theoretically, the shift in the excretion curve in gouty subjects may result from reduced filtration of urate, enhanced reabsorption, or decreased secretion. Patients classified as exhibiting an overproduction of uric acid represent less than 10% of the gouty population.

Secondary Gout

Numerous secondary causes of hyperuricemia and gout can be attributed to a decrease in the renal excretion of uric acid. A reduction in the glomerular filtration rate leads to a decrease in the filtered load of urate and, consequently, to hyperuricemia. Patients with renal disease are hyperuricemic on this basis. Other factors such as decreased secretion of urate have been postulated in patients with some types of renal disease (e.g., polycystic kidney disease, lead nephropathy). Gout is a rare complication of the secondary hyperuricemia that results from renal insufficiency. When it occurs in this setting, there is likely to be a positive family history.

Diuretic therapy currently represents one of the most important causes of secondary hyperuricemia in humans. Diuretic-induced volume depletion leads to a decreased filtered load and enhanced tubular reabsorption of urate. A number of other drugs lead to hyperuricemia by a renal mechanism. These agents include low-dose aspirin, pyrazinamide, nicotinic acid, ethambutol, ethanol, and cyclosporine.

Decreased renal excretion of uric acid is thought to be an important mechanism for the hyperuricemia associated with several disease states. Volume depletion may be an important factor in patients with hyperuricemia associated with adrenal insufficiency or nephrogenic diabetes insipidus. An accumulation of organic acids leads to hyperuricemia. This is the case in starvation, alcoholic ketosis, diabetic ketoacidosis, maple syrup urine disease, and lactic acidosis of any cause (e.g., hypoxemia, respiratory insufficiency, chronic beryllium disease, acute alcohol intoxication). The renal basis of the hyperuricemia in conditions such as chronic lead intoxication, hypoparathyroidism, pseudohypoparathyroidism, and hypothyroidism remains unclear.

Secondary gout can also result from urate overproduction. Four specific defects cause urate overproduction as a consequence of accelerated de novo purine biosynthesis: hypoxanthine phosphoribosyltransferase (HPRT) deficiency, phosphoribosylpyrophosphate synthetase overactivity, glucose-6-phosphatase deficiency, and fructose-1-phosphate aldolase deficiency.

TREATMENT

KEY POINTS

Asymptomatic hyperuricemia is generally not treated, but its identification should lead to a search for the cause and/or associated conditions.

Episodes of acute gouty arthritis can be treated with colchicine, NSAIDs, adrenocorticotropic hormone, and systemic or intra-articular steroids.
The therapeutic aims in gout are as follows (Figure 95-8):

1. To terminate the acute attack as promptly and gently as possible
2. To prevent recurrences of acute gouty arthritis
3. To prevent or reverse complications of the disease resulting from the deposition of sodium urate or uric acid crystals in joints, kidneys, or other sites

4. To prevent or reverse associated features of the illness that are deleterious such as obesity, hypertriglyceridemia, and hypertension

Asymptomatic Hyperuricemia

The presence of hyperuricemia is rarely an indication for specific antihyperuricemic drug therapy. Rather, the finding of hyperuricemia should cause the following questions to be addressed:

1. What is the cause of the hyperuricemia?
2. Are associated findings present?
3. Has damage to tissues or organs occurred as a result?
4. What, if anything, should be done?

Hyperuricemia may be the initial clue to the presence of a previously unsuspected disorder. In 70% of hyperuricemic patients, an underlying cause can be readily defined by history and physical examination. The nature of the underlying cause may be useful in predicting the potential consequences, if any, of the elevated serum urate concentration. Therefore an underlying cause should be sought in every patient with hyperuricemia.

Whether to treat hyperuricemia uncomplicated by articular gout, urolithiasis, or nephropathy is an exercise in clinical judgment, and universal agreement is lacking. When considering whether to treat asymptomatic hyperuricemia with urate-lowering agents, the following data are pertinent:

• Although there are intriguing data from animal models to the contrary, there is no good evidence that renal function is adversely affected by elevated serum urate concentrations.
• The renal disease that accompanies hyperuricemia is most often related to inadequately controlled hypertension.

Purine nucleotide metabolism

Xanthine oxidase inhibitors: allopurinol, oxypurinol, febuxostat

Uricosurics: probenecid, benz bromarone, sulfinpyrazone

Inflammation inhibitors: NSAIDs, colchicine, steroids, ACTH, IL-1 inhibitors

Figure 95-8  Treatment options for gout, including inhibitors of inflammation used for acute attacks and the available approaches for lowering serum urate when appropriate. Some inflammation inhibitors may also be used for prophylaxis when initiating urate lowering therapy. See text for details. Red arrows indicate inhibitory effects; green arrows indicate promoting effects. ACTH, adrenocorticotropic hormone; GI, gastrointestinal; IL-1, interleukin-1; NSAID, nonsteroidal anti-inflammatory drug.
• Although debate exists regarding whether hyperuricemia is an independent risk factor for coronary artery disease, there is no evidence that correction of hyperuricemia has an effect on the development of heart disease.

Thus it seems prudent not to treat hyperuricemia with specific antihyperuricemic agents until symptoms develop. Rare exceptions include individuals with a known hereditary cause of uric acid overproduction or patients at risk for acute uric acid nephropathy.

It is, however, strongly recommended that the cause of hyperuricemia be determined and any associated factors related to the process such as obesity, hyperlipidemia, alcoholism, and, especially, hypertension, be addressed. Fenofibrate and losartan might be appropriate agents for the treatment of hypertriglyceridemia and hypertension, respectively, in hyperuricemic individuals because each has modest uricosuric effects.148,149

Acute Gouty Arthritis

The acute gouty attack may be successfully terminated by any of several drugs. For practical purposes, the choice in most situations is among colchicine, an NSAID, a corticosteroid preparation, or adrenocorticotropic hormone (ACTH). In response to the recognition of the critical role IL-1β plays in the initiation of gout attacks, trials of biologic inhibitors of IL-1β have been undertaken with promising results.150-154 Although approved for the treatment of tumor necrosis factor (TNF) receptor–associated periodic fever syndromes, these expensive agents remain unapproved for gout at this writing. The timing of therapy initiation is more important than the choice of drug for acute gout.155 With any of these agents, the sooner the drug is started, the more rapidly a complete response will be attained. Generally, colchicine is preferred for patients in whom the diagnosis of gout is not confirmed, whereas NSAIDs are preferred when the diagnosis is secure. If a patient cannot take medications by mouth or has active peptic ulcer disease, the choice is among intra-articular glucocorticoid, parenteral glucocorticoid, or ACTH. Local application of ice packs may help control the pain of an acute attack.156 In some cases, analgesics including narcotics may be added as well. Drugs that affect serum urate concentrations including antihyperuricemic agents should not be changed (either started or stopped) during an acute attack. Just as sudden fluctuations in serum urate levels tend to precipitate an acute attack, an inflammatory reaction that is already in progress may be substantially worsened by a major change in the serum urate concentration.

Colchicine

Traditionally, the oral dosing schedule for colchicine was 0.5 or 0.6 mg taken hourly until one of three things occurred: joint symptoms eased; nausea, vomiting, or diarrhea developed; or the patient had taken a maximum of 10 doses. If 10 doses were taken without benefit, the clinician questioned the accuracy of the diagnosis. Many clinicians have recommended that doses be taken every 2 to 6 hours to reduce side effects.157 In 2009 the U.S. Food and Drug Administration (FDA) approved Colcrys (URL Pharma, Inc.) as the first single-ingredient oral colchicine product for the treatment of acute gout flares and the prophylaxis of gout flares, as well as for familial Mediterranean fever in adults and children 4 years of age or older. The FDA ordered a halt to the marketing of unapproved single-ingredient colchicine preparations other than Colcrys. The new recommendation for colchicine is the use of a low-dose regimen of 1.2 mg followed by 0.6 mg in 1 hour, for a total of 1.8 mg per attack. That protocol was based on the results of a randomized controlled trial of 184 patients with acute gout flares of less than 12 hours’ duration that compared low-dose (74 patients) with high-dose colchicine (4.8 mg total over 6 hours; 52 patients) and with placebo (58 patients) with a primary endpoint of at least a 50% reduction in pain within 24 hours.158 The low-dose approach resulted in comparable peak plasma concentrations and early gout flare control with much fewer side effects, particularly diarrhea and vomiting, compared with the high-dose approach. It should be noted that these were patients who started treatment within 12 hours of attack onset, and there was still a greater than 50% failure to achieve the primary endpoint in both treatment arms. The appropriate dosing of colchicine for attacks of greater duration remains unclear. The low-dose recommendations are in line with those of a EULAR panel of experts.159 On the basis of two randomized trials, the recommended dosing for prophylaxis is 0.6 mg once or twice a day.160,161 Patients with severe renal insufficiency should be started at 0.3 mg a day. The marketing of unapproved, injectable colchicine was prohibited by the FDA in 2008 due to an unacceptable safety profile, with the potential for serious adverse events including death with intravenous administration.

Peak plasma concentrations of colchicine occur within 2 hours of oral administration. Although its plasma half-life is 4 hours, levels can be detected in neutrophils 10 days after ingestion. Colchicine has a low therapeutic index, with steady-state plasma concentrations after acute treatment ranging from 0.5 to 3 ng/mL and with toxic effects occurring at approximately 3 ng/mL.162 Therefore in most patients, the side effects precede or coincide with the improvement in joint symptoms. These side effects develop in 50% to 80% of patients using the old, high-dose regimen and include increased peristalsis, cramping abdominal pain, diarrhea, nausea, and vomiting. The drug must be stopped promptly at the first sign of gastrointestinal side effects.163

Colchicine derives its effectiveness from its ability to interfere with acute inflammatory reactions in a variety of ways. Colchicine blocks the processing of IL-1β and inhibits E-selectin–mediated adhesiveness to neutrophils. Its action diminishes neutrophil L-selectin expression, random motility, chemotaxis, phospholipase A2 activation, and IL-1 expression, as well as the stimulated elaboration of platelet-activating factor, crystal-induced chemotactic factor, and leukotriene B4. Colchicine also inhibits endothelial cell ICAM-1 expression and mast cell histamine release and downregulates TNF receptors on macrophages and endothelial cells.
**Nonsteroidal Anti-inflammatory Drugs**

In a patient with an established diagnosis of uncomplicated gout, the preferred agent is an NSAID and indomethacin has been the traditional choice. Although this drug may be effective in doses as low as 25 mg four times a day, an initial dose of 50 to 75 mg, followed by 50 mg every 6 to 8 hours, with a maximum dose of 200 mg in the first 24 hours, has generally been recommended. To prevent relapse, it is reasonable to continue this dose for an additional 24 hours and then taper to 50 mg every 6 to 8 hours for the next 2 days. Clinical trials have shown that oral naproxen, fenoprofen, ibuprofen, sulindac, piroxicam, and ketoprofen, as well as intramuscular ketorolac, are also effective. In fact, all members of this family of drugs can be highly effective in the treatment of acute gouty arthritis including the cyclo-oxygenase-2 (COX-2) selective agents.\(^1\)

**Corticosteroids**

Intra-articular glucocorticoids are useful in the treatment of acute gout limited to a single joint or bursa.\(^5\)\(^5\)\(^,\)\(^16\) Oral, intramuscular, or intravenous glucocorticoids can also provide relief, but these agents are usually reserved for patients who are intolerant of colchicine or NSAIDs or who have medical conditions such as peptic ulcer disease or renal disease that contraindicate their use. Doses of glucocorticoids have been systematically studied, and generally, high doses (prednisone 20 to 60 mg/day) are necessary. Lower doses may not be effective, as evidenced by gout flares occurring in organ transplant patients who are taking maintenance prednisone at doses of 7.5 to 15 mg a day.\(^16\) Anecdotally, rebound attacks were reported as steroids were withdrawn.

**Adrenocorticotropic Hormone**

Single injections of intramuscular ACTH gel (25 to 80 IU) can terminate an acute gout attack.\(^6\)\(^9\) More often, however, repeated administration is required every 24 to 72 hours. This treatment is effective postoperatively and may be more effective than glucocorticoids, possibly related to the mechanism of action. In addition to stimulating the adrenal cortex to produce corticosteroids, ACTH interferes with the acute inflammatory response through activation of melanocortin receptor-3.\(^5\)\(^0\) Unfortunately, ACTH preparations, when available at all, have become prohibitively expensive.

**Prophylaxis**

The practice of giving small daily doses of colchicine as prophylaxis to prevent acute attacks is up to 85% effective.\(^5\)\(^1\) Colchicine 0.6 mg one to three times a day is generally well tolerated, although the drug may produce a reversible axonal neuromyopathy.\(^5\)\(^2\) This complication causes proximal muscle weakness with or without painful paresthesia and elevated serum levels of creatine phosphokinase. This is most often seen in patients with hypertension, renal dysfunction, or liver disease who are also using diuretics. Rhabdomyolysis may also occur in these settings and is more common in individuals who are also taking a statin (HMG-CoA reductase inhibitor) or cyclosporine.\(^5\)\(^3\)

In patients who are unable to tolerate even one colchicine tablet per day, indomethacin or another NSAID has been used prophylactically at low doses (e.g., 25 mg indomethacin twice a day or naproxen 250 mg/day), with some success.\(^5\)\(^4\) Maintenance doses of colchicine or an NSAID may make the difference between frequent incapacitation and uninterrupted daily activities. Prophylaxis is usually continued until the serum urate value has been maintained well within the normal range and there have been no acute attacks for 3 to 6 months. It is important to warn patients that colchicine discontinuation may be followed by an exacerbation of acute gouty arthritis and advise them what to do should an attack occur. Finally, prophylactic treatment is not recommended unless the clinician also uses urate-lowering agents. Prophylactic colchicine may block the acute inflammatory response but does not alter the deposition of crystals in tissues. When deposition continues without the warning signs of recurrent bouts of acute arthritis, tophi and destruction to cartilage and bone can occur without notice.

**Control of Hyperuricemia**

Elimination of hyperuricemia with anti-hyperuricemic agents can prevent and reverse urate deposition. Today, opinion differs as to when in the course of gout the clinician should start antihyperuricemic therapy. Some physicians regard the first gouty attack as a late event in a disorder marked by years of antecedent silent deposition of urate crystals in cartilage and other connective tissue. Others believe that because tophi and symptomatic chronic gouty arthritis develop in only a minority of cases and ordinarily develop slowly after many years of recurrent acute attacks, unnecessary or premature medication can be avoided without demonstrable penalty. In practice, it is rare to have a patient who never experiences a second attack.\(^5\)\(^9\) The probability of such a benign course is greatest in patients who have only minimally elevated serum urate concentrations and normal 24-hour urinary uric acid values. Arguably, a case can be made for initiating antihyperuricemic therapy after the second attack in most patients.\(^5\)\(^2\)

Antihyperuricemic drugs provide a definitive method for controlling hyperuricemia. Although it is important to treat and prevent acute attacks of gouty arthritis with anti-inflammatory agents, it is the long-term control of hyperuricemia that ultimately modifies the manifestations of the gouty diathesis. Once started, treatment with specific urate-lowering agents is lifelong and the dose must be sufficient to maintain the serum urate level below 6.8 mg/dL, preferably between 5 and 6 mg/dL. Lowering the serum urate level from 11 mg/dL to 7.5 mg/dL may seem encouraging, but this change does not reverse the process. It merely slows the rate at which crystals continue to deposit. Generally, the lower the serum urate level achieved during antihyperuricemic therapy, the faster the reduction in tophaceous deposits.\(^5\)\(^6\) The 5 to 6 mg/dL target is recommended because it is far enough below the saturation level of 6.8 mg/dL that it provides some margin for fluctuations in serum levels and
avoids excessive exposure to the medication, which might increase the chance of toxicity.

Reduction to target levels may be achieved pharmacologically by the use of xanthine oxidase inhibitors, uricosuric agents, or uricases. Xanthine oxidase, the enzyme that catalyzes the oxidation of hypoxanthine to xanthine and xanthine to uric acid, is inhibited by allopurinol, oxypurinol, and febuxostat. Probenecid, sulfinpyrazone, and benz bromarone are uricosuric agents that reduce serum urate concentrations by enhancing the renal excretion of uric acid. Rasburicase and pegloticase nzymatically convert urate to allantoin, which is much more soluble and readily excreted in the urine. These antihyperuricemic drugs do not have anti-inflammatory properties.

For those patients with gout who excrete less than 800 mg of uric acid per day and have normal renal function, reduction of the serum urate concentration can be achieved equally well with a xanthine oxidase inhibitor or a uricosuric drug. These agents are equally effective in preventing the deterioration of renal function in patients with primary gout. In most cases, allopurinol is the drug of choice because it can be used with fewer restrictions compared with uricosuric agents.

In general, the ideal candidate for uricosuric agents is a gouty patient who is younger than 60 years and has normal renal function (creatinine clearance >80 mL/min), uric acid excretion of less than 800 mg per 24 hours on a general diet, and no history of renal calculi. Patients prescribed uricosuric agents should be counseled to avoid salicylate use at doses greater than 81 mg/day.

In certain situations, an inhibitor of xanthine oxidase is clearly the drug of choice in a gouty patient. Gouty individuals who excrete larger quantities of uric acid in their urine or who have a history of renal calculi of any type should be treated with a xanthine oxidase inhibitor (Table 95-6). The incidence of renal calculi is about 35% in patients with primary gout who excrete more than 700 mg/day of uric acid. There is also a greater risk for uric acid stones on initiation of uricosuric therapy. In addition, patients with tophi generally should receive a xanthine oxidase inhibitor to decrease the load of urate that must be handled by the kidney. Patients with gout and mild renal insufficiency can be given either type of agent, but benpenecid and sulfinpyrazone would not be expected to work when the glomerular filtration rate is less than 50 mL/min. Allopurinol is effective in the presence of renal insufficiency, but doses may have to be decreased in that situation. Febuxostat needs no adjustment in moderate renal insufficiency. A final indication for a xanthine oxidase inhibitor is the failure of uricosuric agents to produce a serum urate concentration lower than 6 mg/dL or patient intolerance of the uricosuric agent. A xanthine oxidase inhibitor and a uricosuric drug may be used in combination for a patient with tophaceous gout in whom it is not possible to reduce the serum urate below 6 mg/dL with a single agent. In most settings, if allopurinol does not cause the serum urate to drop below 6 mg/dL, it is the result of insufficient dosing or poor patient compliance.

**Table 95-7  Maintenance Doses of Allopurinol Based on Creatinine Clearance Measurement**

<table>
<thead>
<tr>
<th>Creatinine Clearance (mL/min)</th>
<th>Allopurinol Dose (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100 every 3 days</td>
</tr>
<tr>
<td>10</td>
<td>100 every 2 days</td>
</tr>
<tr>
<td>20</td>
<td>100 daily</td>
</tr>
<tr>
<td>40</td>
<td>150 daily</td>
</tr>
<tr>
<td>60</td>
<td>200 daily</td>
</tr>
<tr>
<td>80</td>
<td>250 daily</td>
</tr>
<tr>
<td>100</td>
<td>300 daily</td>
</tr>
<tr>
<td>120</td>
<td>350 daily</td>
</tr>
<tr>
<td>140</td>
<td>400 daily</td>
</tr>
</tbody>
</table>

*These doses represent those that might be selected when initiating therapy with allopurinol. However, urate levels should be checked and the dosage adjusted so that the patient is taking the lowest dose that maintains the serum urate level below 5 to 6 mg/dL.*


**Xanthine Oxidase Inhibitors**

Allopurinol is a substrate for xanthine oxidase and is converted to oxypurinol by that enzyme activity. Oxypurinol is also an inhibitor of xanthine oxidase. Allopurinol is metabolized in the liver and has a half-life of 1 to 3 hours, but oxypurinol, which is excreted in the urine, has a half-life of 12 to 17 hours. Because of these pharmacokinetic properties, allopurinol is dosed on a daily basis and the dosage required to reduce serum urate levels is lower in patients with decreased glomerular filtration rates.

In 1984 guidelines for allopurinol dosing based on creatinine clearance were published (Table 95-7). It is now apparent that these guidelines are useful for selecting the initial dosage of allopurinol, but they do not provide the effective maintenance dose for many individuals and following them does not protect against cutaneous hypersensitivity reactions.

Allopurinol should be used at the lowest dose that lowers the serum urate level below 5 to 6 mg/dL. The most commonly prescribed dose is 300 mg/day, but this is insufficient to adequately reduce serum urate to the target level in 21% to 55% of individuals. Thus higher doses, with a maximum of 800 mg/day, may be required. The sudden lowering of serum urate concentrations that accompanies the initiation of allopurinol therapy may trigger acute gout attacks. This risk can be minimized by beginning prophylactic colchicine or an NSAID (see the previous discussion) 2 weeks before the first dose of allopurinol. Alternatively, the clinician can start allopurinol at a dose of 50 to 100 mg/
day and increase it by similar increments weekly until the desired target is reached.

About 20% of patients who take allopurinol report side effects, with 5% discontinuing the medication. Common side effects include gastrointestinal intolerance and skin rashes. The occurrence of a rash does not necessarily mean the drug should be discontinued. If the rash is not severe, the allopurinol can be withheld temporarily and resumed after the rash has cleared. Oxytpurinol has been tried in patients who are sensitive to allopurinol, but its use is limited by poor gastrointestinal absorption and a high prevalence of cross-reactivity with allopurinol. Oral and intravenous protocols for desensitization to allopurinol have been successful in some patients following cutaneous reactions.  

Other adverse reactions include fever, toxic epidermal necrolysis, alopecia, bone marrow suppression with leukopenia or thrombocytopenia, agranulocytosis, aplastic anemia, granulomatous hepatitis, jaundice, sarcoid-like reaction, and vasculitis. The most severe reaction is the allopurinol hypersensitivity syndrome, which may include fever, skin rash, eosinophilia, hepatitis, progressive renal insufficiency, and death. Autopsies reveal diffuse vasculitis involving multiple organs. This is most likely to develop in individuals with pre-existing renal dysfunction and in those taking diuretics.

Allopurinol is involved in relatively few drug-drug interactions. Its use potentiates the actions of other agents that are inactivated by xanthine oxidase. The most important of these are azathioprine and 6-mercaptopurine. In addition, allopurinol can reduce the activity of hepatic microsomal drug-metabolizing enzymes and prolong the half-lives of warfarin and theophylline. Rash may be more common in patients using allopurinol and ampicillin, and bone marrow suppression may be increased in those also taking cyclophosphamide.

Febuxostat (Uloric, Takeda Pharmaceuticals America, Inc.) was recently approved for use in the treatment of gout on the basis of one phase II, three phase III, and two open-label extension trials. Febuxostat is a potent xanthine oxidase inhibitor that differs from allopurinol in that it is of another chemical class and is a selective inhibitor of enzyme activity. These properties indicate that it would be an excellent alternative for individuals who are intolerant of or hypersensitive to allopurinol. In fact, febuxostat appeared to outperform allopurinol in achieving target serum urates in the clinical trials. Importantly, though, allopurinol doses were fixed and clearly too low in these trials, and therefore it is impossible to make firm conclusions about the comparative efficacy of these two agents. In the United States, febuxostat doses of 40 or 80 mg a day have been approved, whereas in Europe, doses of 80 and 120 mg a day are recommended. The dosage of febuxostat does not need to be adjusted in individuals with mild to moderate renal insufficiency. The same precautions apply as with allopurinol in terms of use with other agents metabolized by xanthine oxidase including azathioprine and 6-mercaptopurine. Diarrhea, dizziness, headache, liver function test abnormalities, and altered thyroid function tests were the most common side effects noted in the clinical trials. Cardiovascular events were numerically higher with febuxostat than with other treatments, but when total drug exposures were taken into account, the incidence was the same. In Europe, the use of febuxostat in patients with ischemic heart disease or congestive heart failure is not recommended. After the FDA-required CONFIRMS trial did not detect increased cardiovascular risk with febuxostat, the FDA approved dosing of 40 mg and 80 mg daily in the United States. Takeda, under guidance from the FDA, has committed to cardiovascular postmarketing. Gout flares were higher early on with febuxostat compared with allopurinol, emphasizing again the need for prophylaxis of up to 6 months when initiating urate-lowering therapy.

Febuxostat and allopurinol have similar safety profiles. Patients with hypersensitivity reactions to allopurinol who were given febuxostat were able to tolerate the new drug. Febuxostat should be considered mainly for patients intolerant to allopurinol, for those whose gout is not controlled with other urate-lowering treatments, and for those with renal insufficiency (creatinine clearance >30 mL/min). Febuxostat should be tried before an attempt at allopurinol desensitization. Finally, febuxostat should be used before uricosuric drugs in patients with nephrolithiasis.

### Uricosuric Agents

Administration of a uricosuric agent increases the rate of renal uric acid excretion. In the kidney, there are separate transport systems for the secretion and reabsorption of organic ions including uric acid. Because urate is reabsorbed by a renal tubular brush border anion transporter, the reabsorption of urate can be inhibited when uricosuric agents are present in the lumen and compete with urate for the transporter. This inhibition of reabsorptive anion transporter requires high doses of uricosuric agents. Because the secretory transport system is quantitatively much smaller than that for reabsorption and is located in the basolateral membrane of the tubule, when uricosuric agents are taken in low doses, they actually decrease the renal excretion of uric acid and raise serum urate levels by inhibiting the secretory transport system.

Probenecid and sulfinpyrazone are the most widely used uricosuric agents available in the United States. Benzbromarone is used for this purpose in other countries as well. However, many other agents can reduce serum urate levels by enhancing the renal excretion of uric acid (see Table 95-8).

Probenecid is readily absorbed from the gastrointestinal tract. Its half-life in plasma is dose dependent, varying from

<table>
<thead>
<tr>
<th>Table 95-8</th>
<th>Drugs That Are Uricosuric in Humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetohexamide</td>
<td>Glycerol guaiacolate</td>
</tr>
<tr>
<td>Amflutizole</td>
<td>Glycine</td>
</tr>
<tr>
<td>Ascorbic acid</td>
<td>Glyceropyrrolate</td>
</tr>
<tr>
<td>Azapropazone</td>
<td>Iodopyracet</td>
</tr>
<tr>
<td>Azaauridine</td>
<td>Iopanoic acid</td>
</tr>
<tr>
<td>Benzbroromarone</td>
<td>Losartan</td>
</tr>
<tr>
<td>Calcitonin</td>
<td>Medofenamic acid</td>
</tr>
<tr>
<td>Calcium ipodate</td>
<td>Orotic acid</td>
</tr>
<tr>
<td>Citrate</td>
<td>Outdated tetracyclines</td>
</tr>
<tr>
<td>Dicumarol</td>
<td>Phensulonophtaldehyde</td>
</tr>
<tr>
<td>Diffunisal</td>
<td>Probenecid</td>
</tr>
<tr>
<td>Estrogens</td>
<td>Salicylates</td>
</tr>
<tr>
<td>Fenofibrate</td>
<td>Sulfinpyrazone</td>
</tr>
</tbody>
</table>
6 to 12 hours. This can be prolonged by the concomitant use of allopurinol. Probencid is metabolized in vivo, with less than 5% of the administered dose recovered in the urine. The maintenance dosage of probenecid ranges from 500 mg to 3 g per day and is administered two or three times a day. Acute gouty attacks may accompany the initiation of this medication, and, as with all uricosuric agents, patients using probenecid are at increased risk for developing renal calculi. With long-term use, up to 18% of individuals develop gastrointestinal complaints and 5% develop hypersensitivity and rash. Although serious toxicity is rare, approximately one-third of individuals eventually become intolerant of probenecid and discontinue its use. Probencid alters the metabolism of several other agents by several mechanisms (Table 95-9). Concomitant use of probenecid can increase the potency of some agents by decreasing their renal excretion, delaying their metabolism, or impairing their hepatic uptake. It may decrease the effectiveness of other medications by reducing their volume of distribution.

Sulfinpyrazone is completely absorbed from the gastrointestinal tract and has a half-life of 1 to 3 hours. Most of the drug is excreted in the urine as the parahydroxy metabolite, which is also uricosuric. Sulfinpyrazone is usually given at a dosage of 300 to 400 mg/day divided into three or four doses. The rates of tolerability and types of adverse reactions are similar to those with probenecid. Unfortunately, sulfinpyrazone is no longer readily available.

Benzbromarone is more potent than probenecid and sulfinpyrazone. It is well tolerated and effective in cyclosporine-treated renal transplant patients. It can be used in those with moderate renal dysfunction (creatinine clearance approximately 25 mL/min).

### Table 95-9  
Effects of Probencid on Metabolism of Other Drugs

<table>
<thead>
<tr>
<th>Decreased Renal Excretion</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-Aminohippuric acid</td>
</tr>
<tr>
<td>Phenolsulfonphthalein</td>
</tr>
<tr>
<td>Salicylic acid and its acyl and phenolic glucuronides</td>
</tr>
<tr>
<td>Phlorizin and its glucuronide</td>
</tr>
<tr>
<td>Acetazolamide</td>
</tr>
<tr>
<td>Dapsone and its metabolites</td>
</tr>
<tr>
<td>Sulfinpyrazone and its parahydroxy metabolite</td>
</tr>
<tr>
<td>Indomethacin</td>
</tr>
<tr>
<td>Ampicillin</td>
</tr>
<tr>
<td>Penicillin</td>
</tr>
<tr>
<td>Cephradine</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reduced Volume of Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
</tr>
<tr>
<td>Ancillin</td>
</tr>
<tr>
<td>Nafcillin</td>
</tr>
<tr>
<td>Cephaloridine</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Impairment of Hepatic Uptake</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bromsulfophthalein</td>
</tr>
<tr>
<td>Indocyanine green</td>
</tr>
<tr>
<td>Rifampicin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Delayed Metabolism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heparin</td>
</tr>
</tbody>
</table>

### Uricases

Uricase was lost to man and some nonhuman primates via a missense mutation in the gene encoding the enzyme (see Chapter 94). Pegloticase is a pegylated mammalian (porcine-like) recombinant uricase, recently approved by the FDA at a dosage of 8 mg intravenously every 2 weeks for the treatment of severe tophaceous gout. With intravenous administration of 0.5 mg to 12 mg, the plasma uricase activity increases linearly with doses up to 8 mg, with an enzymatic activity half-life of 6.4 to 13.8 days. With intravenous doses of 4 mg or more, the serum urate falls dramatically in 24 to 72 hours, from a mean of 11.1 mg/dL to 1 mg/dL and stays low for 21 days after infusion. Pegloticase was studied in a phase I, a phase II open-label trial (41 patients), and two replicate phase III randomized control trials (212 patients total). Pegloticase was significantly more effective than placebo at achieving the primary endpoint of a plasma urate concentration lower than 6 mg/dL for 80% of the time in months 3 and 6 of the trial. It also seemed clear that, when tolerated, pegloticase was efficacious at reducing tophi (Figure 95-9). Colchicine or NSAIDs were used for gout prophylaxis, and oral fenofenadine and acetaminophen, as well as hydrocortisone 200 mg intravenously before infusion, were used for infusion reaction prophylaxis. Despite that, gout flares, infusion reactions, and serious adverse events were significantly more frequent in patients given pegloticase than in other patients. The most common reason for withdrawal was infusion reaction.

Important relations were noted among immunogenicity, infusion reactions, and efficacy. High-titer antibodies to pegloticase and/or polyethylene glycol were associated with loss of response and infusion reactions. In fact, 96% of patients with antibodies to poly(ethylene glycol) from the groups assigned every-2-week or every-4-week pegloticase were nonresponders, and 50% and 76%, respectively, had infusion reactions. These antibodies did not neutralize uricase activity in vitro. The development of these anti-pegloticase antibodies was associated with a loss of response in patients coinciding with a rise in plasma urate concentrations above 6 mg/dL. Tellingly, 71% of infusion reactions occurred after this loss of response. In the group assigned pegloticase every 2 weeks, cessation of treatment when concentrations of plasma urate were higher than 6 mg/dL would have avoided 91% of infusion reactions. Because of this observation, the FDA has recommended that clinicians follow serum urates carefully and stop pegloticase when the serum urate rises above 6 mg/dL. An open-label extension study with 151 patients supported the conclusions from the phase III trials. Gout attacks continued to decline in patients, most patients maintained target serum urates, and an additional 20 patients, beyond the 32 patients in the randomized trials, had complete resolution of at least one tophus. Unfortunately, infusion reactions remained common and sometimes serious, frequently leading to discontinuation, particularly in those who had only received placebo in the randomized trials.
All patients in the pegloticase randomized trials had some cardiovascular risk factor. There was a concern for an independent cardiovascular side effect signal in patients treated with pegloticase, but the FDA's independent analysis concluded that the number of events was too small to lend support to a definite cardiovascular signal. The FDA approved pegloticase for use in treatment-failure gout in October 2010. Pegloticase will have a more limited target population than febuxostat. Terkeltaub has proposed guidelines for the use of uricases in gout. They are appropriate for patients with tophaceous gout with a large excess of total body urate and persisting gout attacks, or with damaging arthropathy, who have failed or are intolerant of conventional therapy. Presumably, many patients who fit this description can now be given and will respond to febuxostat, but others will not. The notion of debulking the urate load is a good way to conceptualize the pegloticase strategy and emphasizes it as adjunctive therapy, best followed by other urate-lowering therapy when possible. A clear advantage is that pegloticase can reduce tophi dramatically and more quickly than conventional treatment (see Figure 95-9). Problems with immunogenicity continue, and the continuous use of pegloticase may be restricted by antibody development. At least 25% of patients develop antibodies, with subsequent infusion reactions, restricted efficacy, or drug withdrawal. Anaphylaxis occurs in about 5% of treated patients. Gout flares are sometimes severe and are frequent, occurring despite prophylaxis in more than 80% of treated patients. The required concomitant infusion of corticosteroids with pegloticase may further restrict its use in patients with conditions such as diabetes or glaucoma, and pegloticase should not be given to patients with glucose-6-phosphate dehydrogenase deficiency because it may induce hemolysis.

Despite the potential downsides of pegloticase, previously these often terribly symptomatic patients had no options. Hopefully, monitoring for a rising serum urate concentration as a sign of antibody development and impending infusion reaction will allow for safer administration. Encouraging data recently presented in abstract form suggest that a subset of patients who normalize their serum urate within the first 6 months of pegloticase treatment can be safely maintained on therapy long term, with a median follow-up of 2.5 years. In addition to ongoing clinical improvement, the frequency of infusion reactions was low in this group, even in subjects with breaks in treatment of up to 6 months. Nevertheless, this is a drug that should be administered at centers familiar with its use and with an established capability of dealing with serious infusion reactions including anaphylaxis.
COMPLIANCE WITH TREATMENT

Because the disease processes involved in gout are so well understood, the diagnosis can be definitively established. Once gout is diagnosed, the available therapies are so effective that it should be a readily treated and easily managed disease. However, too many patients including those who are accurately diagnosed do not do well. Failure of anti-hyperuricemic therapy to attain the target urate level is usually due to improper prescribing or poor compliance. Compliance is often a problem when treating chronic asymptomatic conditions, and associated alcoholism can be a factor. Perhaps more important is the fact that patients may need to take up to three different medications on three different schedules to control their symptoms and treat the disease.

It is believed that if patients understand why they are taking medications, they are more likely to be compliant. Toward this end, an analogy has been developed that helps some patients become more compliant. In this analogy, urate crystals are compared with matches. The patient is told that “when the match strikes,” it causes a gout attack. To “put out the fire,” the patient takes an NSAID or colchicine. Although this resolves the attack, “the matches are still there.” To eliminate future attacks, the patient is given prophylactic colchicine, “which makes the matches damp and harder to strike,” and allopurinol (or a uricosuric agent), “which actually removes the matches from the body.”

MANAGEMENT OF GOUT AFTER ORGAN TRANSPLANTATION

The management of patients with gout after organ transplantation requires careful consideration. The use of glucocorticoids, azathioprine, or cyclosporine and the precarious status of renal function in many patients pose complex problems. Colchicine and NSAIDs may be inappropriate for the management of acute gouty arthritis in this setting because of their potential toxicities. Intra-articular glucocorticoid injections may be most helpful, and one may be forced to rely more heavily on pain medications in this setting. Prophylactic colchicine can be used in patients with normal renal function, but treatment must be monitored closely. The combination of colchicine and cyclosporine has induced rhabdomyolysis.

When considering chronic therapy, it is helpful to lower the doses of cyclosporine and eliminate the use of diuretics, if possible. Uricosuric agents can be used safely, but their usefulness declines if renal function is poor. Allopurinol can be used in patients with abnormal renal function, but the dose may need to be reduced. Febuxostat may be safer in this setting but has not been studied. Both allopurinol and febuxostat may have severe interactions with azathioprine. Azathioprine is metabolized by xanthine oxidase, and because these drugs inhibit that enzyme, the breakdown of azathioprine is slowed, increasing the effective dose. If care is not taken, significant bone marrow toxicity can result. Thiopurine S-methyltransferase (TPMT) studies should be obtained before starting azathioprine. When azathioprine and allopurinol are used together, they can be started at 25 and 50 mg/day, respectively. Revised dosing for febuxostat in such a situation has not been determined. Complete blood counts and serum urate level concentrations are then monitored weekly when using allopurinol, and the dose is adjusted to bring the serum urate concentration to less than 6 mg/dL. As an alternative to azathioprine, mycophenolate mofetil has been used effectively with allopurinol in some transplant patients.

Uricase has been used to drastically lower serum urate levels and shrink tophi in a small number of patients with gout after cardiac transplantation. This treatment has been associated with significant allergic reactions including anaphylaxis, bronchospasm, and hemolytic anemia. The newer preparations of uricase such as pegloticase may avoid those complications and prove more effective.

ANCILLARY FACTORS

In addition to anti-inflammatory agents, colchicine prophylaxis, and antihyperuricemic therapy, other factors may be decisive in determining whether the following develop: recurrent attacks, chronic gouty arthritis, kidney stones, or nephropathy. Today, dietary purine restriction solely to control serum urate levels is rarely advised. A totally purine-free diet reduces the urinary excretion of uric acid by only 200 to 400 mg/day and lowers the mean serum urate value by about 1 mg/dL. In addition, the antihyperuricemic agents available today are so effective that this type of dietary manipulation is rarely necessary. Nevertheless, beneficial results have been reported with a diet of moderate calorie and carbohydrate restriction and a proportionally increased intake of protein and unsaturated fat. Some subjects with gout are susceptible to acute attacks after the consumption of alcoholic beverages or rich foods. Others describe idiosyncratic reactions such as acute gout after eating a particular food, but such relationships are rare and questionable. A diet designed to avoid indiscretions known to precipitate acute gouty attacks in a particular individual is recommended.

In addition, diet is important with regard to other medical problems. Many gouty patients are overweight, and restoration of ideal body weight through regulated calorie restriction is recommended. In addition, at least 75% of patients with primary gout have hypertriglyceridemia. The initial step in managing hypertriglyceridemia is reduction to ideal body weight and elimination of alcohol ingestion.

Many patients with gout consume liberal amounts of alcohol. Acute excesses may lead to exacerbations of hyperuricemia secondary to temporary hyperlactacidemia, and chronic ingestion of alcohol may stimulate increased purine production. The added purine load resulting from regular ingestion of beer may also be a contributing factor. Patients should be warned about the deleterious effects of excessive alcohol intake. Compliance with medication is also much worse among patients who consume alcohol.

About one-third of gouty subjects are hypertensive. The complications of hypertension are potentially more serious than those of hyperuricemia, and the clinician should not hesitate to use whatever drugs are necessary to control the hypertension. Many hypertensive gouty patients require a
thiazide diuretic. If this medication is necessary to control hypertension, it should be used, with the recognition that the dosage of concomitant antihypertensive therapy may need to be adjusted to maintain appropriate control of serum urate levels.

**Selected References**

165. Dalbeth N, Kumar S, Stamp L, Gow P: Dose adjustment of allopurinol according to creatinine clearance does not provide adequate control of hyperuricemia in patients with gout, J Rheumatol 33:1646, 2006.


*Full references for this chapter can be found on www.expertconsult.com.*
References


Calcium Crystal Disease: Calcium Pyrophosphate Dihydrate and Basic Calcium Phosphate

ROBERT TERKELTAUB

Microscopic analysis of synovial fluid in the absence of joint infection or other cause of arthritis is the gold standard for diagnosis of CPPD crystal deposition disease, particularly for acute CPPD crystal-associated arthritis (termed pseudogout). Detection of typical CPPD crystals in tissue sections also can be accomplished whether specimens are fixed in formaldehyde or ethanol, unlike the case for monosodium urate crystals (which dissolve in formaldehyde). On occasion, specialized crystal analytic approaches including x-ray energy spectroscopy and powder diffraction analysis or atomic force microscopy may be helpful in establishing or confirming CPPD crystal deposition. In assessing the form of deposited crystals in calcifications, determination of the calcium/phosphate ratios and the spacing of x-ray powder diffraction lines provide the most specific information.

Consensus clinical diagnostic criteria for articular basic calcium phosphate (BCP) crystal deposition disorders are not in place. Importantly, BCP crystals (unlike urate and CPPD) do not demonstrate global birefringence, though the aggregated particles of basic calcium phosphate crystals demonstrate edge birefringence. Hence diagnosis is predicated on detection of (1) radiographic detection of calcifications characteristic of BCP crystals, (2) synovial fluid crystals that stain strongly for the calcium-binding dye Alizarin red S (which only weakly stains CPPD crystals), (3) BCP crystals confirmed by transmission electron microscopy or specialized crystal analytic approaches (including those mentioned earlier for CPPD and discussed later).

Epidemiology

KEY POINTS

Dysregulated chondrocyte differentiation to hypertrophy and inorganic pyrophosphate (Pi) metabolism are central in pathogenesis of calcium pyrophosphate dihydrate (CPPD) crystal deposition disease.

Autosomal dominant familial CPPD crystal deposition disease has been linked in multiple kindreds to certain mutations in ANKH, a gene encoding a PPi transporter.

NLRP3 (cryopyrin) inflammasome activation and consequent caspase-1 activation and interleukin (IL)-1β processing and secretion drive cell responses to CPPD crystals and CPPD crystal-induced inflammation.

Degenerative arthropathy caused by CPPD crystal deposition disease often involves joints uncommonly affected by primary osteoarthritis such as the metacarpophalangeal, wrist, and elbow joints.

Diagnosis of CPPD deposition disease before age 55, particularly if CPPD deposition is polyarticular, should prompt differential diagnostic consideration of a primary metabolic or familial disorder, and hyperparathyroidism should always be considered in CPPD deposition disease presenting in patients older than the age of 55.

High-resolution ultrasound appears particularly helpful in diagnosis of CPPD crystal deposition disease, partly because radiographic chondrocalcinosis is not detectable in all joints affected by the disease.

Basic calcium phosphate (BCP) crystal deposition in articular cartilage is intimately linked with osteoarthritis, particularly with osteoarthritis of increased severity.

BCP crystals (unlike urate and CPPD crystals) do not demonstrate birefringence, and specialized methods are required to conclusively identify BCP crystals in specimens from the joint.

ACR AND EULAR CRITERIA FOR DISEASE

Table 96-1 features proposed diagnostic criteria for calcium pyrophosphate dihydrate (CPPD) deposition disease, adapted by the author from the proposed European League Against Rheumatism (EULAR) criteria and the original criteria proposed in the past by Daniel J McCarty and colleagues. The diagnosis is based on detection of CPPD crystals by one or more methods. These include not only standard clinically applied radiography but also high-resolution ultrasound to detect hyaline articular cartilage or fibrocartilage calcifications characteristic of CPPD crystal deposition (termed chondrocalcinosis). However, identification of CPPD crystals via compensated polarized light microscopic analysis of synovial fluid in the absence of joint infection or other cause of arthritis is the gold standard for diagnosis of CPPD crystal deposition disease, particularly for acute CPPD crystal-associated arthritis (termed pseudogout). Detection of typical CPPD crystals in tissue sections also can be accomplished whether specimens are fixed in formaldehyde or ethanol, unlike the case for monosodium urate crystals (which dissolve in formaldehyde). On occasion, specialized crystal analytic approaches including x-ray energy spectroscopy and powder diffraction analysis or atomic force microscopy may be helpful in establishing or confirming CPPD crystal deposition. In assessing the form of deposited crystals in calcifications, determination of the calcium/phosphate ratios and the spacing of x-ray powder diffraction lines provide the most specific information.

Consensus clinical diagnostic criteria for articular basic calcium phosphate (BCP) crystal deposition disorders are not in place. Importantly, BCP crystals (unlike urate and CPPD) do not demonstrate global birefringence, though the aggregated particles of basic calcium phosphate crystals demonstrate edge birefringence. Hence diagnosis is predicated on detection of (1) radiographic detection of calcifications characteristic of BCP crystals, (2) synovial fluid crystals that stain strongly for the calcium-binding dye Alizarin red S (which only weakly stains CPPD crystals), (3) BCP crystals confirmed by transmission electron microscopy or specialized crystal analytic approaches (including those mentioned earlier for CPPD and discussed later).

Epidemiology

KEY POINTS

The true prevalence of both CPPD crystal deposition disease and pathologic articular BCP crystal deposition is not known due to limits of radiographic detection.

Chondrocalcinosis including asymptomatic disease increases progressively in prevalence with aging.

Idiopathic/sporadic chondrocalcinosis is rare before age 55, particularly in the absence of a history of joint trauma or knee meniscectomy.

In the past, studies of the prevalence of both CPPD crystal deposition disease and various forms of articular BCP crystal deposition disease have been predominantly based on characteristic plain radiographic features of the disease.
in limited numbers of joints. This is an incompletely sensitive and specific approach. Other studies have been based on results of synovial fluid analyses, but definitive studies based on pathologic findings on examination of articular cartilages, or imaging approaches superior to plain radiography, have not been done. As such, the true prevalence of both CPPD crystal deposition disease and pathologic articular BCP crystal deposition is not known.

It is clear that chondrocalcinosis including asymptomatic disease increases progressively in prevalence with aging. Idiopathic/sporadic chondrocalcinosis is rare before age 55, particularly in the absence of a history of joint trauma or knee meniscectomy. Studies of prevalence based on radiographs have estimated higher prevalence of chondrocalcinosis when the hands, wrists, pelvis, and knees have been surveyed. Indeed, most elderly patients with chondrocalcinosis of the knee also have detectable chondrocalcinosis in other joints. Knee meniscal fibrocartilage calcification alone has been detected in 16% of women aged 80 to 89 and in 30% of women older than 89, figures comparable with those obtained in other studies.

In a radiographic survey study of hands, wrists, pelvis, and knees of patients admitted to a geriatrics ward, there was a 44% prevalence of chondrocalcinosis in patients older than 84, a 36% prevalence in the 75- to 84-year-olds, and a prevalence of 15% in 65- to 74-year-olds; studies of U.K. and Italian community cohorts have been limited to analyses of fewer regions (the knee, or knee and pelvis, respectively) and have yielded lower numbers for prevalence.

In a large U.K. community study the age-, sex-, and knee pain–adjusted prevalence of knee chondrocalcinosis was 4.5% for those older than age 40. In some studies, women have appeared somewhat more commonly affected by CPPD crystal deposition disease than men, but in the recent U.K. study, there was no sex predisposition, although strong association between osteoarthritis (OA) and chondrocalcinosis was confirmed. This appeared to be linked more to the presence of osteophytes rather than joint space narrowing with OA. Interestingly, an association between chondrocalcinosis and diuretic use was uncovered, proposed to be due to the capacity of diuretics to induce hypomagnesemia.

CPPD crystal deposition disease is not uniform in epidemiology between populations. In a random sample of Beijing residents older than 60 years of age, radiographic chondrocalcinosis was compared with whites in the American Framingham OA Study. Chinese had a much lower prevalence of knee chondrocalcinosis, and wrist chondrocalcinosis was particularly rare in Chinese elderly. These findings were unexpected because there is an excess of knee OA in Beijing, and chondrocalcinosis and OA are quite commonly associated in the knee joint.

### GENETICS

#### KEY POINTS

The vast majority of CPPD crystal deposition disease is idiopathic/sporadic, but early-onset familial disease also occurs.

**Linkage of familial CPPD crystal deposition disease to the gene ANKH on chromosome 5p (which encodes a transmembrane protein with functions including PP transport) is well established.**

### Table 96-1 Proposed Diagnostic Criteria for Calcium Pyrophosphate Dihydrate (CPPD) Crystal Deposition Disease

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Demonstration of CPPD crystals, obtained by biopsy or aspirated synovial fluid, by definitive means (such as characteristic x-ray diffraction powder pattern)</td>
</tr>
<tr>
<td>II.</td>
<td>A. Identification of monoclinic or triclinic crystals showing a weak positive birefringence (or no birefringence) by compensated polarized light microscopy</td>
</tr>
<tr>
<td>B. Presence of typical calcifications on radiographs (as discussed in text); heavy punctate and linear calcifications in fibrocartilages, articular (hyaline) cartilages, and joint capsules, especially if bilaterally symmetric</td>
<td></td>
</tr>
<tr>
<td>C. Presence of typical findings for CPPD crystal deposition in articular hyaline cartilage or fibrocartilage by high-resolution ultrasound</td>
<td></td>
</tr>
<tr>
<td>III.</td>
<td>A. Acute arthritis, especially of knees, wrists, or other large joints</td>
</tr>
<tr>
<td>B. Chronic arthritis, especially of knee, hip, wrist, carpus, elbow shoulder, and metacarpophalangeal joints, particularly if accompanied by acute exacerbations</td>
<td></td>
</tr>
</tbody>
</table>

**Modified from McCarty DJ: Crystals and arthritis, Dis Month 6:255, 1994.**
indigenous natives of the Chiloe Island region of Chile. This population has a high prevalence of familial CPPD deposition disease. Families affected with diffuse idiopathic skeletal hyperostosis (DISH) and/or chondrocalcinosis have been identified in the Azores Islands, possibly reflecting an unidentified, shared pathogenic mechanism.

PATHOGENESIS

**KEY POINTS**

- The loose avascular connective tissue matrices of articular hyaline cartilage, fibrocartilaginous menisci, and of certain ligaments and tendons are particularly susceptible to pathologic calcification.
- Joint cartilage pathologic calcification reflects complex interplay between organic and inorganic biochemistry of P₄ and PP, metabolism, aging, dysregulated chondrocyte growth factor responsiveness and differentiation, and other factors.

The loose avascular connective tissue matrices of articular hyaline cartilage, fibrocartilaginous menisci, and of certain ligaments and tendons are particularly susceptible to calcification. Calcium-containing crystals deposited in the pericellular matrix of cartilage are often in the form of CPPD (chemical formula, Ca₃P₂O₇·H₂O; calcium-to-phosphate ratio, 1). Crystals of BCP including partially carbonate-substituted HA (Ca₃[PO₄]₁·OH·2H₂O; calcium-to-phosphate ratio, 1.67) also may be deposited pathologically in articular cartilage, particularly in OA. Importantly, physiologic (and noninflammatory) deposition of HA is essential because HA is the principal mineral phase laid down in growth cartilage and in bone.

Inflammatory conditions also may result from deposition of HA, as well as the closely related BCP crystals, octacalcium phosphate (OCP) (Ca₃H₂[PO₄]₂·5H₂O; calcium-to-phosphate ratio, 1.33) and tricalcium phosphate or “whitlockite” (Ca₃[PO₄]₃; calcium-to-phosphate ratio, 1.5) in periarticular structures such as the rotator cuff (calcific tendinitis) and subacromial bursa of the shoulder (see Chapter 46). CPPD and BCP crystal deposition, reviewed here, are by far the most prevalent arthropathies associated with calcium-containing crystals. Articular calcium oxalate crystal deposition is less common.

Articular cartilage, unlike growth plate cartilage, is specialized to avoid the process of matrix calcification. However, the matrix of articular hyaline cartilage, like that of fibrocartilaginous menisci, lends itself well to pathologic calcification, particularly in association with certain changes in extracellular matrix composition and hydration in aging and OA. Joint cartilage calcification reflects complex interplay between organic and inorganic biochemistry, ion transport, aging, genetics, inflammation, oxidative stress, and dysregulated chondrocyte growth factor responsiveness and differentiation. Pathologic cartilage calcification can reflect deficiencies of certain physiologic calcification inhibitors or upregulation of mediators that actively drive certain patterns of tissue injury culminating in calcification within degenerating cartilage.

Alteration of the concentrations of calcium, inorganic phosphate (P₄), PP, and the solubility products of these ions are clearly at work in promoting CPPD and BCP crystal formation. The levels of ambient magnesium and the composition of the chondrocyte extracellular matrix influence the dynamics of CPPD crystal formation and help to determine whether predominantly monoclinic or triclinic CPPD crystals are predominantly formed. Significantly, monoclinic CPPD crystals are more inflammatory than triclinic CPPD crystals. Matrix effects for CPPD and BCP crystals, studied in experimental gel systems, include promotion of CPPD formation by adenosine triphosphate (ATP) and corticosteroids in conjunction with matrix type I collagen and osteopontin, whereas type II collagen and intact proteoglycans appear to suppress ATP-driven CPPD crystal formation in vitro. Experimental systems to analyze CPPD and BCP crystal deposition have commonly employed isolated matrix vesicle cell fragments from chondrocytes that are enriched in promineralizing constituents and provide a nidus for initiation of calcification, specifically with BCP crystals. Matrix vesicles are important in cartilage growth plate calcification, but it is not yet clear whether CPPD and BCP crystal formation in articular cartilages is driven more by matrix vesicle–mediated effects or nucleation of crystals in association with changes in extracellular matrix constituents, or both pathways. However, CPPD crystals are too large (micron size) to form within matrix vesicles. It should be noted that loci of pericellular concentration of P₄, may be necessary to drive CPPD crystal formation at low micromolar P₄ concentrations developing in cartilages with chondrocalcinosis.

Besides physical effects of calcium, P₄, and PP, on crystal nucleation and propagation, these same solutes exert a variety of mineralization-regulating effects on gene expression, differentiation, and viability in chondrocytes, mediated partly by calcium-sensing receptors and sodium-dependent inorganic phosphate co-transport in chondrocytes. Noxious effects of excess PP, on chondrocytes including induction of matrix metalloproteinase-13 (MMP-13) expression and promotion of apoptosis support the clinical terminology pyrophosphate arthropathy to describe chronic cartilage degenerative manifestations of CPPD crystal deposition disease.

**Dysregulated Inorganic Pyrophosphate Metabolism in Pathologic Articular Cartilage Calcification**

PP, is a potent inhibitor of the nucleation and propagation of BCP crystals. Concordantly, maintenance of physiologic extracellular PP, levels by chondrocytes and certain other cells serves to suppress calcification with HA, as illustrated in mouse models of deficient PP, generation and transport, and a variant of human infantile arterial calcification associated with periarticular calcification. The relatively unique capacity of chondrocytes to produce copious amounts of extracellular PP, is double-edged (Figure 96-1), as supersaturation of cartilage extracellular matrix with PP, is a major factor in promoting CPPD crystal deposition. Furthermore, excess PP, generation can promote BCP crystal deposition by providing a source for increased extracellular P, generation via PP, hydrolysis by the ectoenzyme tissue-nonspecific alkaline phosphatase (TNAP) (see Figure 96-1). Depending on cartilage ATP and PP,
concentrations, as well as the level of activity of $P_i$-generating ATPases and TNAP, CPPD and HA crystal formation may be jointly promoted in cartilage, an event that commonly occurs clinically in OA.

**Role of ENPP1 and ANKH in Inorganic Pyrophosphate Metabolism in Chondrocalcinosis**

Sporadic aging-associated CPPD crystal deposition disease is consistently linked with excess chondrocyte $P_i$-generating nucleotide pyrophosphatase phosphodiesterase (NPP) activity and augmented PP, generation by chondrocytes. In this context, the NPP family isoenzymes ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1) (formerly known as NPP1 and plasma cell membrane glycoprotein-1 [PC-1]) and ENPP3 (formerly known as B10) actively generate PP, by hydrolysis of nucleoside triphosphates including ATP. ENPP1 plays a central role in sustaining and augmenting extracellular PP, in chondrocytes and certain other cells (see Figure 96-1). A substantial portion of ATP used by chondrocytes to generate extracellular PP, is provided by the mitochondria.

ENPP1 is one of a family of enzymes that share NPP catalytic activity and modular type II transmembrane ectoenzyme structures. ENPP1 plays by far the greatest role in augmenting extracellular PP, in chondrocytes. Significantly, marked and total ENPP1 deficiency states in vivo and in vitro are associated with up to 50% less plasma and extracellular PP.$^{24,30}$ In contrast, in idiopathic chondrocalcinosis, cartilage NPP activity and PP$_i$ levels may average approximately double those of normal subjects.$^{34}$

Increased ENPP1 expression is associated with both calcification and apoptosis in degenerative human cartilages.$^{31}$ In more advanced osteoarthritics, decreased cartilage ENPP1 has been described and promotes BCP crystal deposition.$^{34}$ Direct upregulation of ENPP1 in chondrocytic cells stimulates calcification, as well as apoptosis.$^{35}$ These effects are not shared by ENPP3, which likely has other intracellular “housekeeping” functions in chondrocytes.$^{31}$ ENPP2, which is also expressed in normal cartilages, functions more actively in physiology as a lysophospholipase D, and ENPP2 only modestly stimulates chondrocytes to calcify in vitro.$^{31}$

ANKH encodes a multiple-pass transmembrane protein that functions in $P_i$-channeling$^{36,39}$ (Figure 96-2) and possibly ATP release$^{40}$ and regulation of $P_i$ metabolism and uptake by the type III sodium-dependent $P_i$-co-transporter Pit-1.$^{41}$ ANKH promotes bidirectional movement of $P_i$ at the plasma membrane in vitro$^{46}$ but the gradient for ANKH-stimulated $P_i$ movement in chondrocytes (which generate abundant $P_i$, both by their robust ENPP1 expression and intense matrix biosynthetic activity) is from the intracellular to the extracellular space.$^{49}$ Indeed, ANKH transport of $P_i$, generated intracellularly by ENPP1$^{42}$ may be the primary means to regulate extracellular $P_i$ levels.$^{49}$ Modeling of the $P_i$-channeling function of ANKH has proposed 10 or 12 membrane-spanning domains in ANKH with an alternating inside-out orientation and with a central channel to accommodate the passage of $P_i$.$^{36,38}$ (see Figure 96-2).

**Figure 96-1** Proposed inorganic pyrophosphate ($P_i$)-dependent mechanisms stimulating calcium pyrophosphate dihydrate (CPPD) and hydroxyapatite (HA) crystal deposition in aging and osteoarthritis (OA). Roles of adenosine triphosphate (ATP) and $P_i$ metabolism and inorganic phosphate ($P_i$) generation in pathologic cartilage calcification. This model accounts for the association of extracellular $P_i$, excess with both CPPD and basic calcium phosphate (HA) crystal deposition in OA and chondrocalcinosis, as well as the paradoxical association of extracellular $P_i$, deficiency (from defective ANKH or ectonucleotide pyrophosphatase/phosphodiesterase 1 [ENPP1] expression) with pathologic calcification of articular cartilage with HA crystals in vivo. Factors driving pathologic calcification are indicated in green and physiologic factors suppressing calcification in red. Excess $P_i$ generation in aging cartilages in idiopathic CPPD deposition disease of aging and in OA cartilages is mediated in part by increased ENPP1. In idiopathic chondrocalcinosis of aging, mean cartilage PP$_i$ concentrations, as well as the level of activity of TNAP, CPPD and HA crystal deposition, may be jointly promoted in cartilage, an event that commonly occurs clinically in OA.

### Model

- **Aging OA**
  - $P_i$-generating ATPases and TNAP
  - CPPD and HA crystal deposition

- **Excess extracellular $P_i$**
  - Matrix saturation
  - CPPD crystals
  - Cartilage matrix calcification

- **Physiologic $P_i$: inhibits HA deposition**
  - $P_i + P_i$$^+$
  - ADP + $P_i$
  - HA crystals

- **$P_i$-mediated $P_i$, channeling**
  - $P_i + P_i$$^+$
  - AMP, adenosine monophosphate; TGF-β, transforming growth factor-beta.
ANKH is clearly implicated in the pathogenesis of familial and idiopathic/sporadic chondrocalcinosis, \(^{10,13}\) and increased ANKH expression in cartilage is a factor in secondary chondrocalcinosis in OA.\(^{28}\) (See Figure 96-1). In this context, expression of wild-type ANKH is highly regulated and ANKH is increased in OA and chondrocalcinotic cartilages.\(^{28}\) Interestingly, hypoxia, via effects of the transcription factor hypoxia inducible factor-1α, suppresses ANKH expression.\(^{32}\) It is conceivable that increased permeability to oxygen of fibrillated and fissured cartilage in OA favors increased ANKH expression. ANKH, in conjunction with signaling via extracellular P, likely derived from PP\(_4\), promotes chondrocyte maturation to the procalcifying hypertrophic differentiation state.\(^{43}\) Figure 96-1 presents a model in which secondary alterations in chondrocyte expression of both wild-type ANKH and ENPP1 drive PP\(_4\) supersaturation in cartilage in idiopathic/sporadic and OA-associated CPPD crystal deposition disease.

Mutations at different locations in ANKH can affect function and the skeleton in a manner including autosomal dominant chondrocalcinosis.\(^{10,37,39}\) and certain other phenotypes. These include murine progressive ankylosis in the ank/lank mouse and human craniometaphyseal dysplasia associated with apparent decrease of the capacity to transport PP\(_4\), within bone and effects on bone resorption and remodeling, putatively mediated in part by direct and indirect effects on osteoclasts.\(^{36,39,44}\) In a consanguineous family with mental retardation, deafness, and ankylosis, with painful small joint soft tissue calcifications, progressive spondylarthropathy, osteopenia, and mild hypophosphatemia, the homozygous ANK missense mutation L244S was detected in all patients.\(^{45}\) The mutant ANK protein was expressed and localized to the plasma membrane, but fibrosis and mineralization of articular soft tissues developed in homozygotes, with heterozygous carriers of the L244S mutation showing mild osteoarthritis without metabolic alterations.\(^{45}\)

Clinical heterogeneity even for chondrocalcinosis associated with ANKH mutations suggests differing functional effects of ANKH mediated by specific regions of the molecule. All the N-terminally clustered ANKH mutations identified to cause familial chondrocalcinosis (see Figure 96-2) appear to increase PP\(_4\), transport.\(^{38}\) However, some ANKH mutations have distinct effects on chondrocyte differentiation.\(^{33}\) The M48T ANKH mutant in the French kindred appears functionally unique by association with increased intracellular PP\(_4\), and also interrupts the interaction of ANKH with the sodium/phosphate co-transporter Pit-1.\(^{47}\) This may be functionally significant because elevated P\(_4\) increases both ANKH and Pit-1 expression and because ANKH and Pit-1 co-localize in chondrocytes.\(^{45}\) In addition, are the P\(_4\) effects on chondrocyte differentiation discussed later. In 5p familial chondrocalcinosis, subtle gain of function of intrinsic ANKH PP\(_4\) channeling activity may lead to chronic, low-grade chondrocyte “P\(_4\)” leakiness, thereby causing matrix supersaturation with PP\(_4\), CPPD crystal deposition, and cartilage degeneration.\(^{30,37,48}\) An alternative mechanism of disrupting PP\(_4\), metabolism may be promoted by the ANKH mutation ΔE590 linked with a case of sporadic chondrocalcinosis because ANKΔE590 appears to indirectly suppress PP\(_4\), catabolism by association with impairing TNAP expression.\(^{49}\)

**Effects of Imbalance of Chondrocyte Growth Factor Responses on Inorganic Pyrophosphate Metabolism in Chondrocalcinosis**

The chondrocyte growth factor transforming growth factor (TGF)-β stimulates ATP release by chondrocytes,\(^ {40}\) as well as ENPP1 expression and ENPP1 subcellular movement to the plasma membrane, which drive elevation of extracellular PP\(_4\).\(^ {32,50}\) Interleukin (IL)-1β suppresses both ENPP1 expression and extracellular PP\(_4\), in chondrocytes and blocks the effects of TGF-β on PP\(_4\).\(^ {32,50}\) The capacity of TGF-β to raise chondrocyte PP\(_4\) rises with aging, as does TGF-β–stimulated NPP activity,\(^ {51}\) whereas growth-promoting effects of TGF-β decrease with aging in articular chondrocytes.\(^ {51}\) The anabolic chondrocyte growth factor insulin-like growth factor-I (IGF-I) normally suppresses extracellular PP\(_4\), (as well as ATP release)\(^ {40}\) in chondrocytes (see Figure 96-1). Moreover, chondrocyte IGF-I resistance is characteristic of OA and aging cartilages\(^ {44}\) (see Figure 96-1). IGF-I induces expression of cartilage intermediate layer protein (CILP)
CPPD Deposition Disease Secondary to Primary Metabolic Disorders: Relationship to Inorganic Pyrophosphate Metabolism and Chondrocyte Differentiation

Hypophosphatasia, hypomagnesemic conditions (including the Gitelman’s variant of Bartter’s syndrome), hemochromatosis, and hyperparathyroidism are the best-characterized primary metabolic disorders linked to secondary CPPD crystal deposition disease. Increased joint fluid PP levels in each of these conditions suggests at least one common thread in the pathogenesis of chondrocalcinosis via cartilage PP excess. Magnesium is a cofactor for pyrophosphatase activity, and iron excess can suppress pyrophosphatase activity. Hypercalcemia may promote CPPD crystal deposition in hyperparathyroidism (and in familial hypercalciuric hypercalcemia) by effects beyond cartilage matrix supersaturation with ionized calcium such as calcium function as a cofactor in ENPP1 catalytic activity, as well as chondrocyte-activating effects mediated by the calcium-sensing receptor. In addition, normal articular chondrocytes express parathyroid hormone/parathyroid-hormone-related protein (PTH/PTHrP) receptors, and functional responses of chondrocytes to PTH can promote proliferation, altered matrix synthesis, and mineralization.

Hypophosphatasia is due to deficient activity of TNAP, consequently with effects including limitation of hydrolysis PP, to generate P., TNPAP is a major physiologic antagonist of ENPP1-mediated elevation of extracellular PP. Conversely, physiologic ENPP1-induced PP generation antagonizes the essential promineralizing effects of TNAP mediated by P generation, and cartilage PP excess presumably drives chondrocalcinosis in hypophosphatasia. Enpp1 knockout mice and mice homozygous for the ENPP1 truncation mutant demonstrate marked articular cartilage calcification with HA and OA, as well as anklyosing spinal ligament hyperostosis and synovial joint ossific fusion; extracellular PP, levels and mineralization disturbances in soft tissues (but not long bones) of Enpp1 knockout and TNAP-deficient mice are mutually corrected by cross-breeding.

Inflammation, Hypertrophic Chondrocyte Differentiation, and Transglutaminase 2 in Joint Cartilage Calcification

Regulated changes in chondrocyte differentiation and viability appear to be a mechanistically unified process that promotes joint cartilage HA and CPPD crystal deposition, as well as OA. Such changes include development of foci of chondrocyte maturation to hypertrophy, with the presence of hypertrophy, as seen in histopathology of the knee cartilage in Figure 96-3, and apoptosis of chondrocytes typically found adjacent to cartilage calcifications. Articular chondrocyte hypertrophy is associated with heightened PP generation; increased production of calcifying, membrane-limited cell fragments known as matrix vesicles, and certain other calcification-promoting changes in differentiation including alteration of extracellular matrix composition, such that osteopontin (which promotes CPPD crystal formation) is increased and normal collagen subtype composition in the matrix is lost. Altered TGF-β signal transduction in aging and OA may be involved in promoting chondrocyte hypertrophy.
P, taken up by Pit-1 sodium-dependent co-transport and calcium sensing can modulate chondrocyte hypertrophic differentiation and apoptosis, as well as PP, modulating responses to TGF-β. Local upregulation of PTHrP expression also may be one of the shared features driving sequential chondrocyte proliferation and altered differentiation in growth plate chondrocytes and articular chondrocytes. Chondrocyte apoptosis also promotes calcification partly through the calcifying potential of apoptotic bodies functioning as “inside-out” matrix vesicles on release from dying chondrocytes. Mitochondrial dysfunction, a central factor in tissue aging and an apparent mediator of OA progression in aging, also can stimulate cartilage matrix degeneration and calcification. Mitochondria are remarkably specialized to regulate calcification, and apoptosis is critically regulated by mitochondrial function. Moreover, chondrocyte ATP depletion is driven via suppression of mitochondrial oxidative phosphorylation by nitric oxide (NO) as OA evolves in aging, thereby promoting increased ATP scavenging by NPP activity and consequent augmentation of extracellular PP.
Inflammation-associated chondrocyte hypertrophy is driven by hypoxia-inducible factor-2α and Indian hedgehog, as well as by multiple cytokines and calgranulins, oxidative stress, Ptx transport, and receptor for advanced glycation end products (RAGE) signaling, and it is modulated by transglutaminase 2 (TG2) release. Chondrocyte hypertrophy and inflammation jointly drive chondrocalcinosis and progression of OA. For example, IL-1β, which is increased in OA cartilage, stimulates articular chondrocytes to calcify the matrix. NO stimulates both apoptosis and calcification in chondrocytes. IL-1β stimulates inducible nitric oxide synthase (iNOS) expression and increased NO generation, as well matrix alterations. IL-1β (as well as TNF, donors of NO, and the potent oxidant peroxynitrite) also induces increased chondrocyte transglutaminase (TG) activity mediated through the TG family enzymes, factor XIIIa and TG2.

TG2 and factor XIIIa, which function in part to cross-link proteins by transamidation, are markers of growth plate chondrocyte hypertrophy. Significantly, there is upregulation of TG2 and factor XIIIa expression in hypertrophic cells in the superficial and deep zones of knee OA articular cartilage and the central (chondrocytic) zone of OA menisci. Moreover, increased factor XIIIa and TG2 activities both directly stimulate calcification by chondrocytes. OA severity–related, donor age–dependent, and marked age–dependent IL-1-induced increases in TG activity occur in chondrocytes from human knee menisci. TG2 is essential for IL-1β to stimulate articular chondrocytes to calcify in vitro. In addition, the closely related inflammatory chemokines CXCL1 and CXCL8, which are both increased in OA cartilage, induce TG2 and chondrocyte hypertrophic differentiation and calcification that requires TG2. Distinct TG2-independent and TG2-dependent mechanisms promote articular chondrocyte hypertrophy and calcification in vitro, and increased TG2 release is sufficient to promote chondrocyte hypertrophy. TG2 also promotes activation of TGF-β.

The multiligand RAGE mediates several chronic degenerative diseases accompanied by low-grade inflammation. RAGE ligands include S100/calgranulins, a class of small, calcium-binding polypeptides, several of which are expressed by chondrocytes. Normal human knee cartilages demonstrate constitutive RAGE and S100A11 expression, and both RAGE and S100A11 expression are increased in OA cartilages. CXCL8 and TNF induce S100A11 release in cultured chondrocytes. Furthermore, S100A11 induces chondrocyte hypertrophy in vitro, and it does so in a manner dependent on S100A11 homodimerization catalyzed by TG2-mediated transamidation and antagonized by the alternative S100A11 receptor CD36. CXCL1-induced and TNF-induced chondrocyte hypertrophy require RAGE signaling.

Special Pathogenic Aspects of Articular and Periarticular Basic Calcium Phosphate Crystal Deposition

CPPD and BCP crystal deposition can develop in different zones of articular cartilage and probably in distinct phases of cartilage degenerative disease such as ongoing loss of viability in hypertrophic chondrocytes. In addition, abundant cartilage NO production may promote mitochondrial dysfunction, chondrocyte extracellular ATP depletion, and lowering of extracellular PPi, consequently favoring HA over CPPD crystal deposition. The observation that OA and HA crystal deposition in articular cartilage (and arteries) are both promoted by extracellular PPi, deficiency states strikingly illustrates the deleterious effects of deprivation of physiologic extracellular PPi levels. Yet it is noted that joint fluid PPi, and NPP activity are elevated in HA-associated shoulder arthropathy (Milwaukee shoulder syndrome [MSS]), consistent with the model in Figure 96-1.

Pathologic BCP crystal deposition may occur in periarticular sites, as well as numerous organs and soft tissues. Significantly, the shoulder is the most common articular region affected by symptomatic BCP crystal deposition, in part reflecting unique shoulder structure-function (see Chapter 46). Degenerative changes promoted by biomechanical stress promote calcific tendinosis in the body of the rotator cuff. Such tendon calcifications can remain asymptomatic and also can eventually resorb, but the degenerative changes can predispose to tendon rupture. Osteopontin, a factor that normally restrains BCP crystal deposition (and is regulated by PPi, and P), can be detected in fibroblast-like cells and multinucleated macrophages surrounding areas of calcification in calcific tendinosis. In this regard, osteopontin promotes oxidative stress, MMP activation, and macrophage recruitment and osteoclast activation. The presence of multinucleated cells with cathepsin K expression and osteoclast-like functions at sites of tendon calcification suggests a mechanism for both resorption of BCP crystal deposits and tendon degeneration.

Crystal-Induced Inflammation

Some of the crystals deposited in cartilage (Supplemental Figure 96-1 on www.expertconsult.com) can subclinically traffic to joint fluid and synovium, and the crystals can directly stimulate chondrocytes, synovial lining cells, and intra-articular leukocytes. Inflammation triggered by CPPD and BCP crystals thereby contributes to cartilage degradation and can potentiate worsening of OA. Many proinflammatory mechanisms active in gout also likely mediate synovitis and cartilage degeneration associated with CPPD and BCP crystal deposition. In this regard, CPPD and BCP crystals activate cells partly via nonspecific activation of signal transduction pathways (e.g., mitogen activated protein kinase activation) and induce cellular release of cyclooxygenase- and lipoxygenase-derived metabolites of arachidonic acid and cytokines including tumor necrosis factor (TNF), IL-1, and CXCL8. Innate immune recognition of extracellular CPPD crystals by Toll-like receptor 2 (TLR2) and CPPD crystal-induced activation of the intracellular NLRP3 (cryopyrin) inflammasome, resulting in caspase-1 activation and IL-1β processing and release, drive cell responses to CPPD crystals in vitro and CPPD crystal-induced inflammation in vivo.

The ingress of neutrophils into the joint is central in triggering acute crystal-induced synovitis, and effects on neutrophil-endothelial interaction likely represent a major locus for prophylactic effects of nanomolar concentrations of colchicine for the acute arthritis of pseudogout. CXCL8...
Supplemental Figure 96-1  A, Chalky calcium pyrophosphate dihydrate (CPPD) crystal deposits observed on synovium, menisci, and hyaline cartilage of a 35-year-old patient with familial calcium pyrophosphate deposition disease. B, Menisci showing black-stained crystalline deposits under von Kossa’s stain. C, Crystalline pyrophosphate deposits in articular cartilage. (Compensated polarized light, ×300.) D, Hypertrophic chondrocytes surrounding CPPD deposits (hematoxylin and eosin, original magnification ×400).
and related chemokines that bind the CXCL8 receptor CXCR2 (including CXCL1) appear to be critical in initiating and perpetuating neutrophil ingress in acute crystal-induced inflammation. Despite the fact that BCP and CPPD crystals share the capacity to activate certain cell signaling pathways and to induce several MMPs, BCP crystals generally trigger much less neutrophil influx into the joint than do CPPD crystals. Concordantly, free intra-articular BCP crystals likely induce less proinflammatory cytokine expression than do CPPD and monosodium urate crystals, though OCP crystals could be more inflammatory than hydroxyapatite crystals.

**CLINICAL FEATURES**

**KEY POINTS**

In the elderly, CPPD deposition can mimic conditions including gout, infectious arthritis, primary osteoarthritis, RA, or polymyalgia rheumatica. It can also present as fever of unknown origin.

Pseudogout is a major cause of acute monoarticular or oligoarticular arthritis in the elderly; attacks typically involve a large joint, most often the knee, and less often the wrist or ankle, and, unlike gout, rarely the first metatarsophalangeal joint.

Chronic degenerative arthropathy in CPPD deposition disease commonly affects certain joints that are typically spared in primary OA (e.g., metacarpophalangeal joints, wrists, elbows, glenohumeral joints).

**Calcium Pyrophosphate Dihydrate Deposition Disease**

Most elderly individuals with CPPD deposition disease in the United States have a primary (idiopathic/sporadic) disorder (Table 96-2). Idiopathic chondrocalcinosis generally appears only after the fifth decade of life. But patients with a history of repetitive joint trauma of knee meniscectomy may present with nonsystemic (monoarticular) chondrocalcinosis before age 55. Familial forms of CPPD crystal deposition disease also have been widely documented, as discussed later. The clinical presentation of familial chondrocalcinosis is often manifested in the third and fourth decades of life, but familial disease can sometimes be detected before age 20 or first present clinically at advanced age. CPPD crystal deposition disease is also a common manifestation of a variety of hereditary and metabolic conditions (including hyperparathyroidism, dialysis-dependent renal failure, and hemochromatosis) in which the CPPD-related arthropathy can present earlier than age 55. For unclear reasons, hemochromatosis can present predominantly with CPPD crystal deposition disease or as OA. The weight of evidence from controlled studies suggests that hypothyroidism (with the probable exception of myxedematous hypothyroidism) is not associated with a significantly increased prevalence of CPPD crystal deposition disease, though both disorders are clearly prevalent in aging. It has been suggested that initiation of thyroxine supplementation therapy may trigger pseudogout.

The clinical manifestations of CPPD deposition disease vary widely (Table 96-3). Quite commonly, the disease can be asymptomatic. Alternatively, it can mimic OA (pseudo-osteoarthritis), gout (pseudogout) (Supplemental Figure 96-2 on www.expertconsult.com), acute-onset or insidious rheumatoid arthritis (RA) (pseudorheumatoid arthritis) (Figure 96-4), or present as “pseudo-neuropathic” arthropathy. Patients with CPPD crystal deposition disease also commonly present with episodes of hemarthrosis, often post-traumatic and in the knee. The contributions of the forms of CPPD deposited (e.g., monoclinc vs. triclinic crystals) and of host factors to these wide differences in clinical manifestations are not clear. Overall, only a small fraction of patients with CPPD deposition disease have prolonged, recurring polyarticular inflammation. Progressive degenerative arthropathy is more common. Though CPPD deposition disease appears to be a common and significant public health problem in the elderly, the disease and health-related quality of life impact and the long-term course of CPPD-

### Table 96-2 Causes of Calcium Pyrophosphate Dihydrate Crystal Deposition Disease

<table>
<thead>
<tr>
<th>High Prevalence</th>
<th>Moderate Prevalence</th>
<th>Low Prevalence (Largely Based on Case Reports)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idiopathic in association with aging (most frequent)</td>
<td>Familial</td>
<td>X-linked hypophosphatemic rickets</td>
</tr>
<tr>
<td>Complication of primary osteoarthritis</td>
<td>Associated with systemic metabolic disease (hyperparathyroidism, dialysis-dependent renal failure, hemochromatosis, hypomagnesemia)</td>
<td>Familial hypocalciuric hypercalcemia</td>
</tr>
<tr>
<td>Long-term consequence of mechanical joint trauma or knee meniscectomy</td>
<td>Ochronosis</td>
<td>Ochronosis</td>
</tr>
<tr>
<td></td>
<td>Gout</td>
<td>Gout</td>
</tr>
<tr>
<td></td>
<td>Articular amyloidosis</td>
<td>Articular amyloidosis</td>
</tr>
<tr>
<td></td>
<td>Myxedematous hypothyroidism</td>
<td>Myxedematous hypothyroidism</td>
</tr>
<tr>
<td></td>
<td>Osteochondrodysplasias and spondyloepiphyseal dysplasias</td>
<td>Osteochondrodysplasias and spondyloepiphyseal dysplasias</td>
</tr>
<tr>
<td></td>
<td>Neuropathic joints</td>
<td>Neuropathic joints</td>
</tr>
<tr>
<td></td>
<td>Wilson’s disease</td>
<td>Wilson’s disease</td>
</tr>
</tbody>
</table>

### Table 96-3 Common Clinical Presentations of Calcium Pyrophosphate Dihydrate (CPPD) Crystal Deposition Disease

| Asymptomatic or incidental finding (e.g., asymptomatic knee fibrocartilage chondrocalcinosis in the elderly) | Recurrent acute inflammatory monoarticular arthritis (pseudogout) (e.g., wrist, knee, including provocation by trauma, concurrent medical or surgical illness, or intra-articular hyaluronan) | Pseudoseptic arthritis |
| Recurrent acute hemarthrosis | Chronic degenerative arthritis (pseudo-osteoarthritis or pseudo-neuropathic arthritis) | Recurrent acute inflammatory monoarticular arthritis (pseudogout) (e.g., wrist, knee, including provocation by trauma, concurrent medical or surgical illness, or intra-articular hyaluronan) |
| Chronic symmetric inflammatory polyarthritis (pseudorheumatoid arthritis) | Systemic illness (pseudo-polymyalgia rheumatica, fever of unknown origin) | Destructive arthritis in dialysis-dependent renal failure |
| Central nervous system disease complicating ligamentum flavum or transverse ligament of atlas involvement (cervical canal stenosis, cervical myelopathy, meningismus, foramen magnum syndrome, odontoid fracture) | Carpal tunnel syndrome | Tumoral and pseudotophaceous CPPD crystal deposits |
Supplemental Figure 96-2  A, Acute pseudogout. The patients in A, C, and D are siblings. B, Pseudo-osteoarthritis. C, Pseudo-rheumatoid arthritis with boutonnière deformity. D, Pseudo-rheumatoid arthritis showing ulnar deviation, interosseous muscle atrophy, and metacarpophalangeal and wrist joint involvement.
associated degenerative arthropathy in an unsel ected population have not been adequately evaluated.

**Acute Synovitis**

Pseudogout is a major cause of acute monoarticular or oligoarticular arthritis in the elderly. The attacks typically involve a large joint, most often the knee and less often the wrist or ankle, and, unlike gout, rarely the first metatarso phalangeal joint. Acute attacks of inflammatory pseudogout in patients with CPPD deposition disease typically have a sudden onset and can be excruciatingly painful, with pronounced periarticular erythema, warmth, and swelling, comparable to gout. In addition, arthritis in some attacks of pseudogout can be migratory or can be additive, polyarticular, and bilateral. Polyarticular pseudogout is particularly common in association with familial chondrocalcinosis and hyperparathyroidism.

Pseudogout can be provoked by minor trauma or intercurrent medical or surgical conditions including pneumonia, myocardial infarction, cerebrovascular accident, and pregnancy. Parathyroid surgery for hyperparathyroidism frequently triggers pseudogout attacks. In addition, pseudogout of the knee can be precipitated by arthroscopy or by intra-articular administration of hyaluronic acid that could reflect proinflammatory mechanisms triggered through the hyaluronic receptor CD44. Parenteral administration of granulocyte colony-stimulating factor (G-CSF) and of bisphosphonates also can trigger pseudogout, the former likely by ignition of smoldering subclinical intra-articular inflammation, the latter theoretically via pyrophosphate inhibition because bisphosphonates are nonhydrolyzable analogues of PP.

Acute and subacute pseudogout can be associated with fever, chills, elevated erythrocyte sedimentation rate, and systemic leukocytosis, particularly with polyarticular involvement and in the elderly. Leukocyte counts in the synovial fluid are substantially elevated, and intraleukocytic CPPD crystals are most often (but not universally) detectable by compensated polarized light microscopy in pseudogout. The attacks typically last for 7 to 10 days but also can be clustered and last for weeks to months. Occasionally the leukocyte count in pseudogout can exceed 50,000 per mm³ (pseudoseptic arthritis).

**Chronic Degenerative and Inflammatory Arthropathies**

Acute pseudogout attacks may be interspersed with chronic arthropathy in CPPD crystal deposition disease, though it has been suggested that acute flares of pseudogout may become less common in those with established chronic degenerative CPPD deposition arthropathy. Chronic degenerative arthropathy in CPPD deposition disease commonly affects certain joints that are typically spared in primary OA (e.g., metacarpophalangeal joints, wrists, elbows, glenohumeral joints). The development of cartilage degenerative changes in CPPD deposition disease in both typical and atypical joints for primary OA suggests one or more systemic abnormalities.

Degenerative cartilage disease associated with sporadic CPPD crystal deposition disease may present as destructive arthropathy of the knees, hips, and/or shoulders, particularly in elderly females (Figures 96-5 and 96-6). The CPPD crystal arthropathy–associated degenerative disease can be less or more destructive than that observed in primary OA. For example, patients with primary OA and CPPD crystals have been reported to require knee replacement surgery more often than with primary OA without crystals. In another study, 60% of patients undergoing joint replacement had CPPD or BCP crystals (and commonly both) in their knee synovial fluids, and higher mean radiographic scores correlated with the presence of calcium-containing crystals. However, prospective analysis of CPPD deposition disease that principally involved the knee has suggested that radiographic worsening of degenerative changes may be slow. The disease also may not appear to be clinically progressive in the involved knee after substantial periods of follow-up in a subset of patients, though clinical involvement may spread to other joints in the same time frame. Most patients develop changes in radiographic extent of chondrocalcinosis over time. But there is no clear correlation between the extent of calcification and progression of CPPD deposition arthropathy. There may be a relatively
Figure 96-5  Radiographic features of calcium pyrophosphate dihydrate arthropathy. A, Destructive shoulder arthropathy. B, Metacarpophalangeal joint arthropathy. C, Knee degenerative joint disease with large subchondral bone cyst. D, Wraparound patella (same patient shown in A).

Figure 96-6  Chondrocalcinosis of the most commonly affected joints in calcium pyrophosphate dihydrate deposition disease. A, Linear calcifications observed in knee menisci and fibrocartilage. B, Lateral view showing calcification of the articular cartilage as a line parallel to the femoral condyles. C, Calcification of intercarpal joints and triangular ligament. D, Calcification of the symphysis pubis fibrocartilage associated with subchondral bone erosions and subchondral increased bone density.
good prognosis for initial presentation of CPPD deposition disease in the knee as acute pseudogout attacks alone. Overrhematoid involvement in a small subset of patients with CPPD deposition disease presents as a chronic, bilateral, symmetric deforming inflammatory polyarthopathy (see Figure 96-4). Many of these patients have bilateral wrist and metacarpophalangeal joint involvement. Wrist tenosynovitis and carpal tunnel syndrome, cubital tunnel syndrome, and tendon rupture may develop. Ingestion of CPPD crystals by synovial lining cells and lysosomal catabolism of such ingested crystals stimulates synovial proliferation, in part via solubilization of the crystalline calcium. Such effects may contribute to regional synovial and periarticular tenosynovial proliferation promoted by CPPD crystal deposition.84

Other Clinical Forms of Calcium Pyrophosphate Dihydrate Crystal Deposition

Concentrated (tumoral or pseudophoanaceous) CPPD crystal deposition can occur in periarticular structures including tendons, ligaments, bursae, and occasionally in bone.10,11 CPPD deposits in tendons (e.g., Achilles, triceps, and obturator tendons) are usually fine and linear on radiographs. Pseudophoanaceous deposits of CPPD crystals have been detected in the temporal bone, around the knee and hip, and in the acromioclavicular, temporomandibular, elbow, and small hand joints.11 Peripheral tumoral CPPD crystal deposits may sometimes present with acute arthritic attacks. Rarely, tumoral CPPD deposits around the knee can mimic osteonecrosis.14 Tumoral CPPD crystal deposition is typically associated with tissue chondroid metaplasia and behaves like a benign but locally aggressive chondroid tumor, with some of the connective tissue invasion and destruction likely mediated by CPPD crystal-induced cell activation.

Axial skeletal CPPD crystal deposition occasionally involves the intervertebral disks, sacroiliac joints, and lumbar facet joints, and radiographic findings such as linear calcification and spinal ankylosis may appear.115 Meningitis and clinical manifestations resembling herniated intervertebral disk, ankylosing spondylitis, and acute pseudogout of lumbar facet joints have been observed.115,116 In addition, CPPD deposits within the ligamentum flavum or the transverse ligament of atlas can be sizeable and can progress to cause cervical canal stenosis, cervical myelopathy, and foramen magnum syndrome.117-119 Odontoid fracture due to the calcification of the atlantoaxial joint may occur in CPPD deposition disease.117,118 Thus CPPD deposition disease can factor in the differential diagnosis of patients with neurologic disturbances and painful cervical mass, especially in the elderly.

Familial Chondrocalcinosis

Familial CPPD deposition disease has been described in numerous countries and ethnic groups including kindreds from Czechoslovakia, Holland, France, England, Germany, Sweden, Israel, the United States, Canada, and Japan and may be most prevalent in Chile and Spain. In one English kindred with CPPD disease linked to ANKH mutation on chromosome 5p, recurrent childhood seizures were strongly associated with later development of CPPD deposition disease.11 For linkages to ANKH on chromosome 5p, some families manifest early-onset polyarthritis, which can include ankylosing intervertebral and sacroiliac joint disease. In others a late-onset chondrocalcinosis occurs, and the disease can be oligoarticular, mild in intensity and destructiveness, and nearly indistinguishable from idiopathic CPPD deposition disease.12-14 Kindreds from Argentina and the Alsace region of France linked to 5p did share similar phenotypic features of chondrocalcinosis including early age at onset (third decade of life), common but not universal premature OA, some cases of pseudorheumatoid arthritis peripheral joint disease, and radiographic evidence of fibrocartilage and hyaline cartilage calcifications typical of CPPD deposition.12 The most commonly affected joints in these kindreds were the knees and wrists, with involvement of the pubic symphysis and intervertebral disks also described.12

CLINICAL FEATURES OF ARTICULAR BASIC CALCIUM PHOSPHATE CRYSTAL DISEASE

KEY POINTS

Unlike urate and CPPD crystal deposition, acute synovitis due to HA crystal deposition is unusual. Acute inflammatory syndromes including subacromial bursitis and a form of pseudopodagra described in young women may occur in association with periarticular HA crystal deposition in bursae, tendons, ligaments, and soft tissues.

Patients with advanced chronic renal failure, particularly on dialysis, may develop symptomatic articular and periarticular BCP crystal deposition, which may be destructive and involve the axial skeleton. They may resemble or be associated with CPPD deposition disease.

Clinical Features of Pathologic Basic Calcium Phosphate Crystal Deposition in Joint Tissues

Unlike the case for urate and CPPD crystal deposition, acute synovitis due to HA crystal deposition is unusual. But acute inflammatory syndromes including subacromial bursitis and a form of pseudopodagra described in young women may occur in association with periarticular HA crystal deposition in bursae, tendons, ligaments, and soft tissues. Patients with advanced chronic renal failure, particularly on dialysis, may develop symptomatic articular and periarticular BCP crystal deposition (Figure 96-7), which may be destructive and involve the axial skeleton. In some cases of dialysis-dependent renal failure, destructive arthropathy associated with BCP crystal deposition may resemble or be associated with CPPD deposition disease, and monosodium urate crystal deposits in the joint also may occur in the setting. Hyperparathyroidism can promote BCP-associated arthropathy12 and periarticular disease including calcific bursitis (see Figure 96-7). Clinically significant periarticular HA crystal deposition also may occur in certain post-traumatic conditions and the systemic autoimmune diseases scleroderma and dermatomyositis. BCP crystal deposition has a particular predilection for the shoulder (see Chapter 46), where it may manifest as...
calcific tendinitis of the rotator cuff or as a destructive process associated with rotator cuff tear, which is most prevalent in the elderly and more common in females. Abundant intra-articular BCP crystalline material is typically present in the distinctive noninflammatory syndrome of rotator cuff tear and marked cartilage degeneration, an entity termed MSS, cuff tear arthropathy, or apatite-associated destructive arthritis. Mechanical instability of the shoulder due to rotator cuff tear may be the driving force in many of these patients, with consequent release of BCP crystals from bone fragments into the joint space promoting secondary synovitis and connective tissue destruction. The process may be bilateral but is generally worse on the side of the dominant hand. Substantial glenohumeral joint effusions are typically seen, and synovial fluid is often blood stained but contains, at most, relatively low numbers of mononuclear leukocytes. Joints other than the shoulder such as the knee and hip can be affected by a condition similar to MSS, sometimes in the same individual with shoulder involvement. In contrast to primary OA, lateral tibiofemoral
compartment involvement is common in BCP-associated destructive knee arthropathy. Concurrent CPPD deposition, biomechanical abnormalities, chronic renal failure, and neuropathic factors appear to be predisposing factors. A kindred with familial OA and apparent Milwaukee shoulder-knee syndrome (MSKS) had an unusual type of degenerative joint disease with both intra-articular and periarticular calcifications.

Aging itself is a factor in articular cartilage BCP crystal deposition. Multiple studies of synovial fluids and cartilage specimens from OA including recent work that has taken advantage of high-resolution means for BCP crystal detection have suggested that deposition of intra-articular BCP crystalline material including HA in the pericellular matrix of chondrocytes and the capacity of joint cartilages to form such deposits are intimately linked with OA, as well as increased chondrocyte hypertrophy and severity of OA. One limitation of some of these studies is that cartilages were fixed before analysis, and calcifications can develop as a fixation artifact. However, unequivocally, cartilage and synovial fluid BCP crystals, frequently in conjunction with CPPD crystals, are commonly detectable in advanced disease of the knee at the time of total joint replacement for OA. Traffic of crystals from articular cartilage to synovium can promote calcific synovial crystal deposits at or just beneath the synovial surface, and synovium-derived rice bodies can give rise to BCP crystal deposits released into the joint space.

Collectively, the abundance of HA and CPPD crystals in OA joints is likely significant under many clinical circumstances because HA and CPPD crystal-induced synovial proliferation, cytotoxic effects on chondrocytes, and synovial and chondrocyte MMP-13 expression have the potential to promote OA progression. Better surveys of joint tissues with advanced methods for detection of BCP crystals, in particular, will be necessary to improve understanding of the clinical impact of these calcific crystals on OA.

**DIAGNOSIS AND DIAGNOSTIC TESTS**

**KEY POINTS**

- Presence of radiographic evidence for chondrocalcinosis is a common finding in the aged and does not necessarily indicate that the patient’s symptomatic articular problem is due to CPPD deposition disease, which is often asymptomatic.

- The use of compensated polarized light microscopy is essential to confirm the presence of positively birefringent CPPD crystals, though it should be noted that some CPPD crystals are nonbirefringent.

- Patients with arthritis in whom CPPD deposition disease is part of the differential diagnosis can be screened by plain radiographs, but high-resolution ultrasound of the affected joint is a useful and sensitive alternative approach.

**Differential Diagnosis**

CPPD deposition disease can imitate a number of other conditions and vice versa (see Table 96-3), which mandates attention to fulfillment of diagnostic criteria for CPPD deposition disease (see Table 96-1) and necessitates careful adherence to a diagnostic algorithm (Figure 96-8). Conversely, it is important to note that presence of radiographic evidence for chondrocalcinosis is a common finding in the aged and does not necessarily indicate that the patient’s symptomatic articular problem is due to CPPD deposition disease, which is often asymptomatic. The demonstrable presence of CPPD crystals in synovial fluid or in tissues using compensated polarized light microscopy (as discussed earlier for gout vs. pseudogout) is definite evidence for CPPD deposition disease. Though weakly birefringent relative to urate crystals and often rhomboid in shape, CPPD crystals can be rod shaped and intracellular, thereby resembling urate crystals. Thus the use of compensated polarized light microscopy is essential to confirm the presence of positively birefringent CPPD crystals, though it should be noted that some CPPD crystals are nonbirefringent. The appearance and number of CPPD crystals can change with storage. Therefore clinicians should examine relatively fresh specimens collected in vials free of calcium-chelating anticoagulants such as EDTA. Cytocentrifugation increases sensitivity of detection of rare CPPD crystals, as observed in one study that highlighted approximately 75% to 78% of synovial fluid samples (of subjects with gout and OA) having both urate and CPPD crystals.

The ability of pseudogout to mimic septic arthritis (pseudoseptic arthritis) and vice versa underscores the diagnostic importance of arthrocentesis with appropriate synovial fluid crystal analysis and, in many instances, concomitant exclusion of joint infection. Significantly, crystal deposits can be “enzymatically strip-mined” by inflammation associated with joint sepsis. Hence CPPD (as well as other crystals) may be observed free in the joint fluid and within synovial fluid leukocytes in an infected joint.

Diagnosis of CPPD deposition disease before age 55, particularly if CPPD deposition is widespread, should prompt differential diagnostic consideration of a primary metabolic or familial disorder (see Table 96-2). In the elderly, presentation of CPPD deposition as diffuse pain and fever of unknown origin can mimic infection, polymyalgia rheumatica, and RA. A false-positive rheumatoid factor (RF) test is common in the elderly (≥30% positivity). Thus patients with pseudorheumatoid CPPD deposition disease are often RF seropositive.

**Differential Diagnostic Considerations for Basic Calcium Phosphate Crystal Deposition**

BCP crystals may be detected as nonbirefringent globular clumps within leukocytes in some synovial and bursal fluids (see Figure 96-7), and BCP crystal clumps stain with the calcium-binding dye Alizarin red S under light microscopy. CPPD crystals can also be detected using Alizarin red S but stain more weakly than BCP crystals. Relative paucity of osteophytes (so-called atrophic degenerative arthropitis) on plain radiographs and the sizeable glenohumeral joint effusions with abundant synovial fluid BCP crystalline material associated with MSS help distinguish MSS from primary OA of the glenohumeral joint. However, destructive, neuropathic shoulder arthropathy due to syringomyelia or alcoholism sometimes merits consideration in the
Acute or chronic arthritis

Suspect CPPD crystal deposition disease:
- Elderly?
- History of OA or trauma
- Typical pseudogout or pseudo-OA
- Hemarthrosis
- Specific metabolic disorder
- Family history of chondrocalcinosis

• Aspirate affected joint
• Synovial fluid analysis by compensated polarized light microscopy

• Plain radiographs of affected joint
Or: High-resolution ultrasound of affected joint

• Crystal analysis positive for CPPD, or x-ray evidence for CPPD deposition
• Gout excluded
• Joint cultures negative

Crystal analysis and radiographs or ultrasound negative: assess for alternative diagnosis

Treat inflammatory arthritis:
- NSAIDs
- Systemic or intra-articular steroids (or ACTH as an alternative)
- Prophylactic colchicine for recurrent pseudogout
- For refractory chronic inflammatory arthritis, experimental use of:
  - Methotrexate
  - Hydroxychloroquine
  - IL-1 inhibition

Evaluate for primary metabolic disorder:
- Serum Ca²⁺, PO₄⁻³, alkaline phosphatase
- Serum Mg²⁺
- Serum Iron, TIBC, ferritin
- Renal function
- TSH

Another diagnosis established

Figure 96-8  Algorithm for diagnosis, evaluation, and treatment of calcium pyrophosphate dihydrate (CPPD) deposition disease. The algorithm is discussed in detail in the text. Treatment options are in line with those recently advanced by the European League Against Rheumatism. ACTH, adrenocorticotropic hormone; IL-1, interleukin-1; NSAIDs, nonsteroidal anti-inflammatory drugs; OA, osteoarthritis; TIBC, total iron binding capacity; TSH, thyroid-stimulating hormone. (From Guerne P-A, Terkeltaub R: Calcium pyrophosphate dihydrate crystal deposition: epidemiology, clinical features, diagnosis, and treatment. In Terkeltaub R, editor: Gout and other crystal arthropathies, Philadelphia, 2011, Elsevier.)

differential diagnosis of MSS (see Chapter 47). Oxalate crystal deposition arthropathy can be a major differential diagnostic consideration with BCP-associated arthritis and periarticular calcifications in dialysis-dependent renal failure.

Chronic CPPD Deposition Arthropathy, BCP Crystal-Associated Arthritis, and Use of Plain Radiographs in Diagnosis

Chronic arthritis in CPPD deposition disease has several clinical and plain radiographic features helpful in differentiating it from OA. These include involvement at sites uncommon for primary OA such as the wrist, metacarpophalangeal joints, elbow, or shoulder, as well as radiographic heavy punctate and linear calcifications in fibrocartilages, articular (hyaline) cartilages, and joint capsules, especially if bilaterally symmetric (see Figures 96-5 and 96-6). It should be noted that faint or atypical calcifications may be due to BCP-related vascular calcifications. Deposition of nonpathologic dicalcium phosphate dihydrate (DCPD) (CaHPO₄ · 2H₂O, calcium-to-phosphate ratio, 1) (“brushite”) crystals has been thought to cause some atypical calcifications, but brushite crystals can arise as an artifact of acid preparation of calcified tissue for pathologic analyses.

CPPD crystal deposits often appear as broad linear streaks or linear “chunks” in articular hyaline and fibrocartilages on plain radiographs, whereas BCP crystal deposits in articular cartilage require high-resolution radiography for detection. Only a pattern of “atrophic” degenerative arthritis without osteophytes and variable subchondral bone thickening may be seen on plain radiographs of BCP arthritis in the shoulder and other large joints.

Patients with arthritis in whom CPPD deposition disease is part of the differential diagnosis can be screened radiologically by obtaining an anteroposterior (AP) view of each knee, an AP view of the pelvis (to detect symphysis pubis involvement, which is quite common), and posteroanterior (PA) views of both hands that include visualization of both wrists (see Figures 96-5 and 96-6). Calcific deposits may or may not be detectable by x-ray screening of these areas in...
CPPD deposition disease. In this circumstance, radiographic evidence other than chondrocalcinosis may point to the correct diagnosis.\textsuperscript{142} For example, radiographic findings suggestive of CPPD deposition disease, as opposed to primary OA, include radiocarpal or marked patellofemoral joint space narrowing, especially if isolated (such as the patella “wrapped around” the femur), as well as scaphoid-lunate widening and femoral cortical erosion superior to the patella. Severe progressive degeneration in the knee with subchondral bony collapse (microfractures) and fragmentation with formation of intra-articular radiodense bodies is a feature of CPPD presenting as a “pseudoneuropathic” joint. CPPD deposition disease involving the metacarpophalangeal joints can be distinguished radiographically from RA by metacarpal squaring associated with “beaklike” osteophytes and subchondral cyst formation. Tendon calcifications (e.g., Achilles, triceps, and obturator tendons) are a valuable differential diagnostic feature of CPPD deposition. Osteophyte formation is more variable with CPPD deposition than with OA. Clearly, x-ray findings may not correlate with pathologic and clinical manifestations in CPPD disease. For example, the correlation between radiographic and pathologic findings was only 39.2\% in a study of patients via knee arthroscopy.\textsuperscript{143} 

High-Resolution Ultrasound and Advanced Imaging for Diagnosis of CPPD and BCP Crystal Deposition Diseases

Ultrasound can clearly detect aggregated BCP-related calcifications outside of joint cartilages including in the rotator cuff of the shoulder. In addition, in small, preliminary studies, high-resolution ultrasound (e.g., in the 6 to 13 MHz range using the current generation of equipment) detection of CPPD crystal deposits in joints has correlated well with positive results for synovial fluid analyses and has detected CPPD in some patients in whom plain radiographs were negative in the affected joint.\textsuperscript{144-149} Preliminary criteria proposed for CPPD calcifications by ultrasound are summarized in Table 96-4, and Figure 96-9 (and Supplemental Figure 96-3 on www.expertconsult.com) illustrates characteristic ultrasound features of CPPD deposits. Further validation is necessary for reliance on ultrasound without radiography in CPPD diagnosis.

The ultrasound approach is likely most specific for detection of CPPD crystal deposition in fibrocartilages (e.g., triangular fibrocartilage of the wrist and the midzones of articular hyaline cartilages (see Figure 96-9). Ultrasound, without use of synovial fluid analysis in diagnosis of acute pseudogout, runs the risk of missing other conditions such as infectious arthritis. Gout can generally be differentiated, particularly because redundant hyperechoic contouring of the cartilage surface is seen more in gout, whereas CPPD crystal deposition disease is typically visualized within the cartilage.\textsuperscript{145} Because enthesopathies other than CPPD disease can also give rise to calcification of tendons and plantar fascia, the diagnostic value for CPPD disease of ultrasound detection of calcification in plantar fascia and Achilles tendon\textsuperscript{150} is not yet clear.

Limitations of ultrasound include difficulty in visualizing crystal deposits in deep recesses of the joint space, the need for a high-resolution machine of the current generation, dependence on the skill of the ultrasonographer, and aforementioned issues with specificity of findings.

Dual-energy computed tomography has not been specifically studied for CPPD detection but is useful for specific discrimination of urate from BCP deposits.\textsuperscript{151} Magnetic resonance imaging (MRI) is not yet a reliable approach for detecting CPPD crystal deposition disease due to a lack of mobile protons in CPPD crystals, and nonenhanced MRI is less sensitive in detecting knee meniscal fibrocartilage calcification than hyaline cartilage calcification.\textsuperscript{152}

Laboratory Diagnostic Tests for CPPD and BCP Crystal Deposition Disease

Conventional radiography or ultrasound is usually the first method to evaluate patients with suspected chondrocalcinosis, but thorough laboratory evaluation of the newly diagnosed CPPD disease patient routinely includes serum levels of calcium, phosphorus, magnesium, alkaline phosphatase, ferritin, iron and total iron-binding capacity, and thyroid-stimulating hormone (TSH) (see Figure 96-8).

Specialized techniques beyond Alizarin red S staining such as x-ray diffraction, Raman spectroscopy, Fourier transform infrared spectroscopy, atomic force microscopy, or transmission electron microscopy showing electron-dense clumps of needle-like crystals may be necessary to confirm BCP crystal deposition (see Figure 96-7).\textsuperscript{129} Cytocentrifugation can increase sensitivity of CPPD detection in synovial fluid.\textsuperscript{138} Under conditions where synovial fluid specimens are not fresh (or stored at 4° C for analysis after significant delay\textsuperscript{155}), Gram stain and Diff Quik staining methods for crystal analysis in synovial fluids have been suggested to provide information beyond that from compensated polarized light microscopy.\textsuperscript{134,155} Demonstration of CPPD crystals in articular tissues (see Figure 96-3) can be difficult in specimens stained with hematoxylin and eosin because the acidity of hematoxylin solutions promotes decalcification. However, the decalcifying effect of hematoxylin can be diminished by limiting the staining period with Mayer’s hematoxylin to 3 minutes.\textsuperscript{156}

Table 96-4 Preliminary Criteria for Calcium Pyrophosphate Dihydrate (CPPD) Crystal Deposition Diagnosis by High-Resolution Ultrasound

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. All CPPD deposits are hyperechoic and present as one of the following patterns:</td>
<td></td>
</tr>
<tr>
<td>A “punctate” pattern, composed of several thin hyperechoic spots, more common in fibrous cartilage and in tendons</td>
<td></td>
</tr>
<tr>
<td>Homogeneous hyperechoic nodular or oval deposits localized in bursae and articular recesses (frequently mobile)</td>
<td></td>
</tr>
<tr>
<td>Bursae and articular recesses (frequently mobile)</td>
<td></td>
</tr>
<tr>
<td>2. CPPD crystal deposit calcifications always have a sparkling appearance and create posterior shadowing only when they reach dimensions of greater than 10 mm. In contrast, calcifications that present a hypoechoic appearance with posterior shadowing even at an early stage (2-3 mm in diameter) are considered as crystalline deposits of another nature, most commonly due to basic calcium phosphate crystal deposition disease.</td>
<td></td>
</tr>
</tbody>
</table>

Supplemental Figure 96-3  Tumoral calcium pyrophosphate dihydrate (CPPD) crystal deposition in wrist. A, Radiograph of wrist showing nodular radiodensity. B, CPPD pseudotumour. Nodular mass removed from wrist joint, 4 × 2 × 2 cm. C, Photomicroscopy demonstrating crystal deposits bounded by connective tissue cell showing cartilaginous metaplasia (hematoxylin and eosin, original magnification ×100). D, Compensated polarized light microscopy showing CPPD crystal aggregates. Individual CPPD crystals have rod and rhomboidal shapes and are positively birefringent. (Courtesy Dr. Ken Pritzker, Mount Sinai Hospital Pathology Department, University of Toronto, Ontario, Canada.)
Figure 96-9  Calcium pyrophosphate dihydrate (CPPD) deposition disease detected by high-resolution ultrasound (US) versus plain radiography.  
A and B, Triangular fibrocartilage of the wrist. Longitudinal (A) lateral scan and (B) corresponding wrist radiograph showing hyperechoic rounded deposits within the substance of the fibrocartilage.  
C and D, Meniscal calcification of the knee as it appears on US (C) and corresponding radiograph (D). Arrowhead, fibrocartilage calcification; et, extensor carpi ulnaris tendon; f, femur; t, tibia; tr, triquetrum bone; u, ulna.  
A was obtained with a Diasus (Dynamic Imaging, Livingstone, United Kingdom) using an 8- to 16-MHz linear probe.  
C was obtained using a Logiq 9 (General Electric Medical Systems, Milwaukee, Wisc) using a 4D16L probe.  
E, Ultrasonographic manifestations of hyaline cartilage CPPD crystal deposition disease as hyperechoic linear band parallel to the femoral condyle in a posterior location (arrowheads).  
TREATMENT

**KEY POINTS**

CPPD deposition disease treatment involves alleviation and prophylaxis of acute arthritic attacks, but therapy to lessen chronic and anatomically progressive sequelae of crystal deposition is not well developed for CPPD disease.

The approach to pseudogout treatment is similar to that for acute gout.

**Calcium Pyrophosphate Dihydrate Deposition Disease**

A treatment algorithm is presented at the bottom of Figure 96-8 that is consonant with the preliminary EULAR guidelines. As in gout (see Chapter 95), therapeutic approaches to patients with CPPD deposition disease involve treatment and prophylaxis of acute arthritic attacks, but therapy to lessen chronic and anatomically progressive sequelae of crystal deposition is not well developed for CPPD (Table 96-5). Reduced meniscal calcification was reported over a 10-year period in association with administration of oral magnesium to a patient with secondary CPPD deposition disease caused by hypomagnesemia. However, there is no specific treatment validated to prevent or lessen crystal deposition of idiopathic CPPD deposition disease. Metabolic disorders that secondarily cause CPPD crystal deposition obviously require treatment. However, the potential benefits for prevention of chondrocalcinotic cartilage degeneration in the appropriate treatment of hemochromatosis and hyperparathyroidism are unclear because the ability to detect chondrocalcinosis radiologically is usually indicative of advanced crystal deposition disease.

Episodes of pseudogout generally respond to nonsteroidal anti-inflammatory drugs (NSAIDs) (including cyclooxygenase-2 inhibitors) and/or intra-articular steroids, though sometimes more slowly than in gout. Systemic glucocorticosteroids or adrenocorticotropic hormone (ACTH), generally given as described for acute gout (see Chapter 95), appear effective in most cases of acute pseudogout. The response to colchicine bolus is less consistent than that usually seen in acute gout. Intravenous colchicine is not recommended as treatment for pseudogout and was withdrawn from active marketing in 2008 by the U.S. Food and Drug Administration. However, pseudogout episodes can be diminished in frequency by low-dose daily colchicine prophylaxis, as for gouty arthritis. The self-limited nature of most acute pseudogout attacks in the knee can sometimes be accelerated by simple arthrocentesis and thorough drainage of the joint effusion, but there are currently no data for measures such as tidal irrigation in CPPD deposition disease, unlike the case for MSS.

Hydroxychloroquine has been suggested to be of some benefit to patients with refractory chronic polyarticular CPPD deposition disease and to reduce flares of pseudogout, and it has theoretic benefits by potentially stabilizing phagolysosomes to suppress NLRP3 inflammasome activation in response to CPPD crystal uptake. Methotrexate was promising to patients with refractory chronic polyarticular CPPD deposition disease and to reduce flares of pseudogout, though in an exploratory study limited to five consecutive patients, and with patients prior to methotrexate as their own controls, IL-1 antagonism (e.g., with off-label use of anakinra) has had anecdotal success, whereas breakthrough pseudogout has been reported during TNF antagonist therapy. Collectively, at this time, there is insufficient evidence basis for hydroxychloroquine, methotrexate, and IL-1 antagonism as standard therapies for refractory inflammation in CPPD crystal deposition disease.

Effective cartilage-preserving therapy is still lacking in idiopathic chronic progressive CPPD deposition disease. Only limited evidence suggests that OA patients with cartilage calcification respond distinctly to arthroscopic irrigation and daily low-dose colchicines, but further substantiation is necessary. There is currently no evidence to support arthroscopic debridement as a treatment modality for CPPD deposition disease. There is insufficient evidence for beneficial effects of intra-articular hyaluronan therapy in CPPD deposition disease of the knee, and risks of precipitating pseudogout appear significant with this treatment modality, as cited earlier.

**Basic Calcium Phosphate Crystal Arthropathies**

NSAIDs and local glucocorticoid injection (Table 96-6) are effective treatment options for BCP crystal-associated calcific tendinits and subacromial bursitis (see Chapter 46). BCP crystal–associated inflammation of the rotator cuff and subacromial bursa of the shoulder can be successfully treated using needle aspiration, irrigation, and steroid injections. Ultrasound-guided techniques, which promote resorption of

---

**Table 96-5  Therapeutics for Calcium Pyrophosphate Dihydrate (CPPD) Crystal Deposition Disease**

<table>
<thead>
<tr>
<th>Proven Benefits</th>
<th>Possible Benefits Already Observed Clinically</th>
<th>Theoretic Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSAIDs or COX-2 inhibitors</td>
<td>Methotrexate for refractory chronic inflammation and recurrent pseudogout</td>
<td>Phosphocitrate</td>
</tr>
<tr>
<td>Intra-articular corticosteroids</td>
<td>Oral magnesium (for patients with hypomagnesemia)</td>
<td>Caspase-1 or IL-1 antagonism for CPPD crystal-induced inflammation</td>
</tr>
<tr>
<td>Systemic corticosteroids</td>
<td></td>
<td>Hydroxychloroquine for refractory chronic inflammation</td>
</tr>
<tr>
<td>ACTH</td>
<td></td>
<td>TLR2 antagonism for CPPD-associated degenerative arthropathy</td>
</tr>
<tr>
<td>Prophylactic low-dose colchicine</td>
<td></td>
<td>Oral calcium supplementation to suppress PTH levels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANKH anion channel blockade (probenecid)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPP1 inhibition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TG2 inhibition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Polyphosphates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Promotion of crystal dissolution by alkaline phosphatase or polyamines</td>
</tr>
</tbody>
</table>

ACTH, adrenocorticotropic hormone; COX-2, cyclooxygenase-2; IL-1, interleukin-1; NPP1, nucleotide pyrophosphate phosphodiesterase 1; NSAIDs, nonsteroidal anti-inflammatory drugs; PTH, parathyroid hormone; TG2, transglutaminase 2; TLR2, Toll-like receptor 2.
rotator cuff and bursal calcifications, can enhance the success of such approaches.\textsuperscript{170,171} Tidal irrigation may be of benefit for symptoms and function in MSS.\textsuperscript{152,160}

**Future Directions in Treatment**

A potential factor that helps suppress the prevalence of chondrocalcinosis in China is high oral calcium intake, which can limit PTH production by the parathyroid. Specifically, calcium levels in tap water in Beijing were 12- to 20-fold higher than that in Framingham, whereas no difference was found in levels of magnesium in the aforementioned study of China versus USA chondrocalcinosis prevalence by Zhang and colleagues.\textsuperscript{9} Deficient calcium intake in aging is a major public health problem in Western countries. It is possible that chondrocalcinosis is more of an environmentally mediated finding than previously recognized, via subclinical variability in calcium intake and parathyroid function. Given the current lack of effective, rational therapies to prevent or lessen idiopathic CPPD crystal deposition, further study of the potential prophylactic and therapeutic benefits of dietary calcium supplementation on chondrocalcinosis is warranted.

The potential to develop therapies for both CPPD and BCP crystal-associated arthropathies based on new molecular targets has been elevated by identification of ANKH, ENPP1, and TG2 as specific molecular mediators of cartilage calcification. Intriguingly, the anion transport inhibitor probenecid suppresses ANKH-induced and TGF-B-induced increases in extracellular PP\textsubscript{i} in vitro.\textsuperscript{36,18,172} Prevention of CPPD deposition by polyphosphates or modulation of CPPD dissolution by depot alkaline phosphatase and by pyrophosphatase activation-promoting polyamines could provide alternative therapeutic approaches.\textsuperscript{173-175} However, incomplete CPPD crystal dissolution by intra-articular lavage of patients with chondrocalcinosis of the knees with disodium EDTA and magnesium ions in the past was a therapeutic failure in that insignificant amounts of CPPD were removed and all subjects developed postlavage attacks of pseudogout mediated by crystal shedding.\textsuperscript{176}

The PP\textsubscript{i} analogue phosphocitrate, a natural compound in mammalian mitochondria and in the urinary tract, is a potent inhibitor of HA crystal formation.\textsuperscript{177} Phosphocitrate inhibits nitric oxide–induced calcification of cartilage and also inhibits both HA and CPPD crystal-associated cell stimulation including induction of MMP-3 in fibroblasts.\textsuperscript{178} Systemic phosphocitrate treatment suppresses ankylosing ossification in murine progressive ankylosis of ank/ank mice.\textsuperscript{179} Moreover, an analogue of phosphocitrate (CaNaPC) decreased both the abundant meniscal cartilage HA deposition and the continuous progression of OA in the Hartley guinea pig model of spontaneous knee OA.\textsuperscript{180} The CaNaPC treatment did not exert therapeutic effects in a rabbit knee hemimenisectomy model of OA in which there was an absence of intra-articular calcification.\textsuperscript{180} Such results suggest that phosphocitrate acts on calcification-mediated mechanisms of dysregulation of joint biomechanics and cartilage degeneration without exerting nonspecific chondroprotective effects. Further clinical development of phosphocitrate would be of interest but has been slowed in part by low bioavailability unless given parenterally.\textsuperscript{177}

The use of bisphosphonates as PP\textsubscript{i} analogues can be beneficial in some cases of soft tissue calcification with HA. Lastly, the identified roles of TLR2 in chondrocyte responsiveness to CPPD crystals,\textsuperscript{92} of TLR4 in hydroxyapatite crystal–induced inflammatory responses,\textsuperscript{181} and of NLRP3 inflammasome-mediated caspase-1 activation and IL-1\textbeta processing in CPPD crystal–induced inflammation\textsuperscript{93} suggest certain mediators of innate immunity including TLR2, TLR4, caspase-1, and IL-1\textbeta to be potential therapeutic targets for human forms of CPPD and hydroxyapatite crystal-driven inflammation and connective tissue destruction.

**OUTCOME**

**KEY POINTS**

It is not clear whether the presence of CPPD crystals in primary knee OA is a predictive factor for more frequent knee replacement surgery, despite the fact that CPPD crystals are frequently found in OA knee tissues at the time of total joint arthroplasty.

There is no clear correlation between the extent of calcification and progression of primary CPPD deposition arthropathy.

The presence of CPPD crystals in primary knee OA had been proposed to be a predictive factor for more frequent knee replacement surgery.\textsuperscript{110} Moreover, mean radiographic scores directly correlated with the presence of calcium-containing crystals in OA in patients at the time of total joint arthroplasty.\textsuperscript{111} However, degenerative cartilage disease associated with sporadic CPPD crystal deposition disease may be less destructive than that observed in primary OA. For example, prospective analysis of CPPD deposition disease of the knee suggested that radiographic worsening of degenerative arthritis was slow to progress.\textsuperscript{172} Typically, changes in radiographic extent of chondrocalcinosis are observed over time,\textsuperscript{172} but there is no clear correlation between the extent of calcification and progression of CPPD deposition arthropathy.

In the Boston OA Knee Study (BOKS) and in the Health, Aging and Body Composition (Health ABC) Study\textsuperscript{182} the relationship between chondrocalcinosis and the progression of knee OA was prospectively evaluated.
longitudinally using MRI. In BOKS, knees with chondrocalcinosis had a decreased risk of cartilage loss compared with knees without chondrocalcinosis and there was no difference in risk in Health ABC. Stratification by the presence of intact or damaged knee menisci produced comparable results within each cohort. In a Thai study, CPPD crystal deposition disease was identified radiographically and/or by synovial fluid analysis in 52.9% of 102 patients undergoing total knee arthroplasty. Patients with and without chondrocalcinosis did not differ in difficulties in performing daily activities or treatment, and those with chondrocalcinosis did not undergo knee arthroplasty at an earlier age than those without chondrocalcinosis.

In the setting of OA, the processes leading to matrix calcification had been regarded to reflect passive secondary consequences of advanced cartilage pathology. Moreover, joint inflammation induced by deposited crystals was thought to be the primary determinant of the clinical impact of chondrocalcinosis on the progression of OA. The aforementioned studies and advances in understanding the pathogenesis of OA and chondrocalcinosis paint a different picture of the impact of chondrocalcinosis in OA. In essence, the dysregulated cartilage matrix repair that generates cartilage calcification may be as effective (or in some cases more effective) at slowing cartilage tissue failure than other phenotypes of cartilage repair in OA.

Websites
Wellcome Trust Centre for Human Genetics: www.well.ox.ac.uk
Teaching resource page for radiographic images of BCP and CPPD arthropathies:
www.orthopaedicweblinks.com/Teaching_Resources/Radiology/more3.html

Selected References

References


94. Deleted in proofs.


The familial autoinflammatory syndromes, often referred to as hereditary periodic fever syndromes, comprise rare hereditary disorders with a common phenotype of lifelong, recurrent inflammatory episodes, characterized by inflammatory symptoms such as fever, abdominal pain, diarrhea, rash, or arthralgia. Between the fever episodes, patients with most of these syndromes generally feel healthy and function normally. Routine laboratory investigations during a fever attack invariably reveal a severe acute-phase response with a high erythrocyte sedimentation rate, leukocytosis, and high concentrations of acute-phase proteins such as C-reactive protein (CRP) and serum amyloid A (SAA). The inflammatory episodes occur without an obvious trigger, although some patients note a relationship to physical stimuli (e.g., exposure to cold), emotional stress, or the menstrual cycle. The episodes resolve spontaneously in days or weeks. Patients go undiagnosed for years, generating a high level of discouragement and frustration for patients and physicians when no diagnosis is made.

A significant proportion of patients with a periodic fever phenotype still do not fit into this genetically based classification, probably representing additional (genetic) defects that can lead to autoinflammatory disease. This chapter describes the seven best characterized familial autoinflammatory syndromes at this time in detail.

**DIFFERENTIAL DIAGNOSIS**

When a patient has had recurrent fever episodes for more than 2 years, it is increasingly unlikely that these are caused by an infection or a malignant disorder. The differential diagnosis at that time may include numerous inflammatory disorders such as juvenile rheumatoid arthritis, adult-onset Still’s disease, inflammatory bowel disease, Schnitzler syndrome, and Behçet’s disease, in addition to the hereditary periodic fever syndromes (Table 97-2). Because the hereditary syndromes are rare (except for familial Mediterranean fever [FMF] in individuals with a distinct ethnic background), the more common diagnoses should usually be excluded first.

The mainstay of the diagnosis of hereditary autoinflammatory syndromes is clinical assessment, with a detailed medical and family history, and preferably at least one observation of the patient during a fever episode because physical examination of the patient in a period of remission is seldom abnormal. Another helpful clue, although not pathognomonic, is often gained from knowing the patient’s ethnic origin. This clinical assessment often yields enough information to build a differential diagnosis of the specific familial autoinflammatory syndromes (see Table 97-1) to determine the direction of genetic testing (Figure 97-2).

When the specific diagnosis remains elusive, a trial treatment with an interleukin-1 (IL-1) inhibitor (e.g., anakinra) can give a diagnostic clue. A significant proportion of patients that suffer from an autoinflammatory disease “not otherwise specified” respond well to treatment with anakinra; we even use the (temporary) label of “anakinra-responsive disease” for this category in daily practice.
### Table 97-1  Differential Diagnosis of Familial Autoinflammatory Syndromes

<table>
<thead>
<tr>
<th>Mode of Inheritance</th>
<th>Mevalonate Kinase (MKD) Deficiency</th>
<th>Tumor Necrosis Factor Receptor–Associated Periodic Syndrome (TRAPS)</th>
<th>Cryopyrin-Associated Periodic Syndrome (CAPS)</th>
<th>Chronic Infantile Neurologic Cutaneous and Articular Syndrome (CINCA)</th>
<th>Deficiency of IL-1 Receptor Antagonist (DIRA)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Familial Mediterranean Fever (FMF)</strong>, Mevalonic Aciduria</td>
<td>Autosomal recessive</td>
<td>Autosomal dominant</td>
<td>Mevalonic kinase</td>
<td>Autosomal dominant</td>
<td>Birth, &lt;4 wk</td>
</tr>
<tr>
<td><strong>Hyper–immunoglobulin D Syndrome (HIDS)</strong></td>
<td>Autosomal recessive</td>
<td>Autosomal recessive</td>
<td>Cryopyrin</td>
<td>Autosomal dominant</td>
<td>Autosomal recessive</td>
</tr>
<tr>
<td><strong>Familial Cold Autoinflammatory Syndrome (FCAS)</strong></td>
<td>Autosomal recessive</td>
<td>Autosomal dominant</td>
<td>Muckle-Wells Syndrome (MWS)</td>
<td>Autosomal dominant</td>
<td>Autosomal recessive</td>
</tr>
<tr>
<td><strong>Muckle-Wells Syndrome (MWS)</strong></td>
<td></td>
<td></td>
<td>Cold-induced urticaria-like lesions</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chronic Infantile Neurologic Cutaneous and Articular Syndrome (CINCA)</strong></td>
<td></td>
<td></td>
<td>Urticaria-like rash</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Deficiency of IL-1 Receptor Antagonist (DIRA)</strong></td>
<td></td>
<td></td>
<td>Urticaria-like lesions</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Erysipelas-like erythema</strong></td>
<td>Erysipelas-like erythema</td>
<td></td>
<td>Generalized pustulosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Maculopapular rash</strong></td>
<td>Maculopapular rash</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Morbilliform rash</strong></td>
<td>Morbilliform rash</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Migratory rash, overlying area of myalgia</strong></td>
<td>Migratory rash, overlying area of myalgia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cold-induced urticaria-like lesions</strong></td>
<td>Cold-induced urticaria-like lesions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Urticaria-like rash</strong></td>
<td>Urticaria-like rash</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Urticaria-like lesions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Maritomyalgia</strong></td>
<td>Severe myalgia common; occasional frank monarthritis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Epiphyseal bone formation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Splenomegaly, pain may occur</strong></td>
<td>Splenomegaly, pain may occur</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Splenomegaly, pain may occur</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sterile peritonitis common</strong></td>
<td>Sterile peritonitis common</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Splenic enlargement, severe pain common</strong></td>
<td>Splenic enlargement, severe pain common</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Severe pain common</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>None</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>May occur</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hepatosplenomegaly</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Uncommon</strong></td>
<td>Uncommon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Uncommon</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Conjunctivitis and periordial edema common</strong></td>
<td>Conjunctivitis and periordial edema common</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Conjunctivitis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Conjunctivitis; sometimes optic nerve elevation</strong></td>
<td>Conjunctivitis; sometimes optic nerve elevation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Papilledema with possible loss of vision, uveitis</strong></td>
<td>Papilledema with possible loss of vision, uveitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Erysipelas-like erythema</strong></td>
<td>Erysipelas-like erythema</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prominent cervical lymphadenopathy</strong></td>
<td>Prominent cervical lymphadenopathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dysmorphic features, neurologic symptoms</strong></td>
<td>Dysmorphic features, neurologic symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Migratory nature of myalgia and rash, periordial edema</strong></td>
<td>Migratory nature of myalgia and rash, periordial edema</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cold-induced urticaria-like lesions</strong></td>
<td>Cold-induced urticaria-like lesions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sensorineural hearing loss</strong></td>
<td>Sensorineural hearing loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chronic aseptic meningitis, sensorineural hearing loss, arthropathy</strong></td>
<td>Chronic aseptic meningitis, sensorineural hearing loss, arthropathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gene Involved</strong></td>
<td>MEFV</td>
<td>MKV</td>
<td>CIAS1 = NLRP3</td>
<td>CIAS1 = NLRP3</td>
<td>IL-1RN</td>
</tr>
<tr>
<td><strong>Protein Involved</strong></td>
<td>Pyrin (marenostrin)</td>
<td>Mevalonate kinase</td>
<td>Type 1 tumor necrosis factor receptor</td>
<td>Cryopyrin</td>
<td>IL-1RA</td>
</tr>
<tr>
<td><strong>Age at Onset (yr)</strong></td>
<td>&lt;20</td>
<td>4-6</td>
<td>&lt;20</td>
<td>1-2</td>
<td>?</td>
</tr>
<tr>
<td>*<em>Duration of attack (days)</em></td>
<td>&lt;2</td>
<td>4-5</td>
<td>&gt;14</td>
<td>&lt;2</td>
<td>Continuous</td>
</tr>
<tr>
<td><strong>Musculoskeletal Involvement</strong></td>
<td>Monoarthritis common</td>
<td>Arthralgia, occasional oligoarthritis</td>
<td>Arthralgia common</td>
<td>Arthralgia common; occasional mild myalgia</td>
<td></td>
</tr>
<tr>
<td><strong>Abdominal Involvement</strong></td>
<td>Sterile peritonitis common</td>
<td>Splenomegaly, severe pain common</td>
<td>Splenomegaly, pain may occur</td>
<td>None</td>
<td>May occur</td>
</tr>
<tr>
<td><strong>Eye Involvement</strong></td>
<td>Uncommon</td>
<td>Uncommon</td>
<td>Conjunctivitis and periordial edema common</td>
<td>Conjunctivitis</td>
<td>Papilledema with possible loss of vision, uveitis</td>
</tr>
<tr>
<td><strong>Distinguishing Clinical Symptoms</strong></td>
<td>Erysipelas-like erythema</td>
<td>Prominent cervical lymphadenopathy</td>
<td>Dysmorphic features, neurologic symptoms</td>
<td>Migratory nature of myalgia and rash, periordial edema</td>
<td></td>
</tr>
<tr>
<td><strong>Duration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Duration may vary; this is a typical duration.

Note: For details on Blau syndrome, DIRA, and pyogenic sterile arthritis, pyoderma gangrenosum, and acne (PAPA) syndrome, see text.

FAMILIAL MEDITERRANEAN FEVER

Epidemiology

FMF (Mendelian inheritance in men [MIM] 249100) is the most prevalent disorder among the hereditary autoinflammatory syndromes, with more than 10,000 patients affected worldwide. It occurs primarily in people originating from the Mediterranean basin including Armenians, Sephardic Jews, Arabs, and Turks. FMF is an autosomal recessively inherited disorder. Most families reported with an apparent autosomal dominant inheritance pattern of FMF represent examples of pseudodominant inheritance owing to consanguinity combined with the high carrier frequency of FMF mutations in certain populations; however, at least three families studied do seem to show a true dominant inheritance, even after extensive genetic analysis.
**Table 97-2  Differential Diagnosis of Periodic Fever**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hereditary (see Table 97-1)</td>
<td></td>
</tr>
<tr>
<td>2. Nonhereditary</td>
<td></td>
</tr>
<tr>
<td>a. Infectious</td>
<td></td>
</tr>
<tr>
<td>i. Hidden infectious focus (e.g., aortoenteric fistula, Caroli’s disease)</td>
<td></td>
</tr>
<tr>
<td>ii. Recurrent reinfection (e.g., chronic meningococccemia, host defense defect)</td>
<td></td>
</tr>
<tr>
<td>iii. Specific infection (e.g., Whipple’s disease, malaria)</td>
<td></td>
</tr>
<tr>
<td>b. Noninfectious inflammatory disorder, e.g.:</td>
<td></td>
</tr>
<tr>
<td>i. Adult-onset Still’s disease</td>
<td></td>
</tr>
<tr>
<td>ii. Juvenile chronic rheumatoid arthritis</td>
<td></td>
</tr>
<tr>
<td>iii. Periodic fever, aphthous stomatitis, pharyngitis, and adenitis</td>
<td></td>
</tr>
<tr>
<td>iv. Schnitzler syndrome</td>
<td></td>
</tr>
<tr>
<td>v. Behçet’s syndrome</td>
<td></td>
</tr>
<tr>
<td>vi. Crohn’s disease</td>
<td></td>
</tr>
<tr>
<td>vii. Sarcoidosis</td>
<td></td>
</tr>
<tr>
<td>viii. Extrinsic alveolitis</td>
<td></td>
</tr>
<tr>
<td>ix. Humidifier lung, polymer fume fever</td>
<td></td>
</tr>
<tr>
<td>c. Neoplastic</td>
<td></td>
</tr>
<tr>
<td>i. Lymphoma (e.g., Hodgkin’s disease, angioimmunoblastic lymphoma)</td>
<td></td>
</tr>
<tr>
<td>ii. Solid tumor (e.g., pheochromocytoma, myxoma, colon carcinoma)</td>
<td></td>
</tr>
<tr>
<td>d. Vascular (e.g., recurrent pulmonary embolism)</td>
<td></td>
</tr>
<tr>
<td>e. Hypothalamic</td>
<td></td>
</tr>
<tr>
<td>f. Psychogenic periodic fever</td>
<td></td>
</tr>
<tr>
<td>g. Factitious or fraudulent</td>
<td></td>
</tr>
</tbody>
</table>

**Etiology**

In 1997 two groups independently traced the genetic background of FMF to a hitherto unknown gene on the short arm of chromosome 16, dubbed the MEderitean FeVer (MEFV) gene. At least 80 disease-linked mutations in the MEFV gene have been described so far, most of which are clustered in the tenth exon of this gene (for details see the online mutation database at [http://fmf.igh.cnrs.fr/](http://fmf.igh.cnrs.fr/)). Most are missense mutations that produce a single amino acid change in the protein (Figure 97-3).

There are six common mutations, accounting for almost 99% of all FMF chromosomes: M694V (occurring in 20% to 65% of cases, depending on the population examined), V726A (in 7% to 35%), M680I, M694I, V694I, and E148Q. For the first three mutations mentioned here, a founder effect has been established, pointing to common ancestors at least 2500 years ago. The high frequency of the mutated MEFV gene in more than one Middle Eastern population has led to the hypothesis that heterozygous carriers have an as-yet-unknown advantage, possibly a heightened (inflammatory) resistance to an as-yet-unidentified endemic pathogen of the Mediterranean basin. In about 30% of patients, only one or no mutations in the MEFV gene can be detected; the etiology in these patients still needs to be determined.

**Pathogenesis**

The MEFV gene encodes for a protein of 781 amino acids, known as pyrin or marenostrin. Pyrin is expressed as a cytoplasmic protein in mature monocytes in association with microtubules but is predominantly found in the nucleus in granulocytes, dendritic cells, and synovial fibroblasts. The expression of pyrin is induced by inflammatory mediators such as interferon-α and tumor necrosis factor (TNF). The pyrin domain is shared by many proteins involved in apoptosis and inflammation and is a member of the death-domain superfamily that includes death domains, death-effector domains, and caspase-recruitment domains. Pyrin binds specifically to other proteins that contain a pyrin domain, which include the adapter protein “apoptosis-associated specklike-like protein with a CARD” (ASC).

The proinflammatory cytokine interleukin (IL)-1β is central in the pathogenesis of FMF. This cytokine is expressed as an inactive precursor, which is cleaved by caspase-1 to yield the active IL-1β. Caspase-1 itself first needs to be activated through the interaction with a protein complex termed an inflammasome. Several inflammasomes have been described so far. The major inflammasome complex involved in the activation of caspase-1 and IL-1β is the cryopyrin or NLRP3 inflammasome. Two hypotheses have been proposed regarding the effect of pyrin on IL-1β processing. The “sequestration hypothesis” holds that pyrin has an inhibitory effect on caspase-1–mediated activation of IL-1β, through its prevention of the formation of the cryopyrin inflammasome by competitive binding of the adapter protein ASC and procaspase-1 and binding caspase-1.

Under this hypothesis, FMF mutations are thought to interfere with the inhibiting interactions of pyrin, resulting in decreased regulation of IL-1β activation. The second hypothesis, proposed by Yu and colleagues, suggests that pyrin can form its own specific inflammasome for activation of IL-1β, although not all the components of this proposed inflammasome have been specified so far. The FMF mutations would increase the sensitivity of this putative pyrin inflammasome. Apart from its role in regulation of IL-1β, there is also conflicting evidence for the effect of pyrin on regulation of nuclear factor κB (NFκB) or apoptosis, varying from inhibition to stimulation.

**Clinical Features**

In approximately 90% of FMF patients, symptoms start before age 20 years. The inflammatory attacks of FMF usually last 1 to 3 days. The frequency can vary widely; 2 to 4 weeks is the most common interval (see Figure 97-1). Symptoms of serositis (i.e., peritonitis, pleuritis, synovitis) are the main feature of FMF attacks, usually accompanied by fever. Abdominal pain of 1 or 2 days’ duration occurs in 95% of patients, varying in severity from severe peritonitis resembling an acute abdomen to only mild abdominal pain without overt peritonitis. Arthritis (rarely destructive) is often confined to one large joint such as the knee, ankle, or wrist and may be the only symptom. Chest pain resulting from pleuritis is usually unilateral and associated with a friction rub or transient pleural effusion. Skin involvement occurs in approximately 30% of patients, most often as erysipelas-like skin lesions on the shins or feet. Other, more uncommon, symptoms are pericarditis, occurring in less than 1%; acute scrotal swelling and tenderness; aseptic meningitis; and severe protracted myalgia, especially of the legs.
Figure 97-2  Differential diagnosis of the familial autoinflammatory syndromes. First exclude other, more common causes of fever and inflammation in the patient. When a familial autoinflammatory syndrome seems likely, check the clinical characteristics found in the patient on the right and left of the diagram, and assign one point to each syndrome that is linked to these characteristics by a line (one characteristic could lead to or point to more than one syndrome). The final combined score assigns a rank for the likelihood of the disorders in this patient and offers help in deciding on the correct subsequent diagnostic tests. This algorithm is not evidence based but is solely derived from expert opinion. CAPS, cryopyrin-associated periodic syndrome; EOS, early-onset sarcoidosis; FMF, familial Mediterranean fever; HIDS, hyper-IgD syndrome; PAPA, pyogenic sterile arthritis, pyoderma gangrenosum, and acne; TRAPS, tumor necrosis factor receptor–associated periodic syndrome.

Figure 97-3  Schematic representation of pyrin (marenostrin) protein, with four conserved domains including a pyrin domain, a B-box (B), coiled-coil domain (CC), and a B30.2 domain. Indicated are mutations as found in familial Mediterranean fever, with the five most common missense mutations in bold type. For complete listing of all currently known mutations, see INFEVERS website: http://fmf.igh.cnrs.fr/infevers/.
FMF is still primarily a clinical diagnosis. There is a set of validated diagnostic criteria with a reported sensitivity and specificity of 96% to 99% (Table 97-3). These criteria were validated in a population with a high prevalence of FMF and low prevalence of the other autoinflammatory disorders, however, and the ethnic origin of the patient needs to be taken into account. In molecular diagnostic testing, genetic laboratories usually screen for the five most common mutations and rare mutations are missed. MEFV mutations occur on both alleles in only 70% of typical cases, whereas in the remaining 30%, only one or no mutation can be detected, even after sequencing. There is also evidence of reduced penetrance. Despite these limitations, molecular testing can be used as a confirmatory test. Whether or not the results are positive, treatment with colchicine is warranted in symptomatic cases of fitting ethnic origin fulfilling the diagnostic criteria. No specific biologic marker is available to distinguish an inflammatory FMF attack from an infectious fever or appendicitis. During an inflammatory attack, there is an acute-phase response, which includes elevation of SAA, CRP, and plasma fibrinogen and polymorphonuclear leukocytosis. Proteinuria in patients with FMF is highly suggestive of renal amyloidosis.

### Treatment

Colchicine is the first-line treatment for patients with FMF. Colchicine prevents inflammatory attacks completely in 60% to 75% of patients, and it significantly reduces the number of attacks in an additional 20% to 30%. The average dose in adults is 1 mg daily, but this may be increased to 3 mg in cases in which no response is seen at the lower dose. This regimen is usually well tolerated; gastrointestinal side effects including diarrhea and abdominal pain generally resolve with dose reduction. More serious side effects such as myopathy, neuropathy, and leukopenia are rare and occur primarily in patients with renal or liver impairment. During a fever attack, oral or intramuscular nonsteroidal anti-inflammatory drugs (NSAIDs) can be used for pain relief. Glucocorticoids have limited efficacy.

Compliance with colchicine use is important because colchicine has been shown to prevent the occurrence of amyloidosis. Since the introduction of colchicine therapy, the incidence of amyloidosis in FMF has decreased dramatically, whereas in areas with a high prevalence of FMF where colchicine is not routinely available, such as Armenia, amyloidosis is still common. Colchicine’s principal effect at the cellular level is to depolymerize microtubules by interacting with tubulin, inhibiting motility and exostosis of intracellular granules. It has a powerful antimitotic effect, causing metaphase arrest. It has been speculated, in cases of infertility in patients treated with colchicine, that this medication causes azoospermia. Colchicine does not have a significant adverse effect on sperm production or function, however.

Unfounded fear of teratogenic effects of colchicine often wrongly leads to cessation of this drug in young women who want to get pregnant, with a subsequent increased frequency and severity of attacks, which enhances problems with fertility and pregnancy. Colchicine has proved to be safe, even in early pregnancy, and treatment should not be interrupted for this reason. It can also be used while breastfeeding.

In about 5% to 10% of patients, FMF is refractory to colchicine use. Lidar and colleagues used parenteral colchicine in such refractory cases, but this can be toxic. More recently, the IL-1β inhibitor anakinra has been shown to be effective in several case reports and series.

### Outcome

Recurrent attacks of peritonitis may lead to intra-abdominal or pelvic adhesions, resulting in complications such as small bowel obstruction. Another serious long-term complication of FMF is amyloid A (AA) amyloidosis. This amyloidosis is primarily found in the kidneys, resulting in renal failure, but can also occur in the gastrointestinal tract, liver, and spleen, and eventually in the heart, testes, and thyroid. The prevalence of amyloidosis varies, especially depending on the ethnic origin, but it is high in untreated patients. It is common among Sephardic Jews but rare in Ashkenazi Jews.
For a variety of reasons including peritoneal adhesions and ovulatory dysfunction, subfertility in women is common. In men, subfertility secondary to azoosperma (sometimes secondary to testicular amyloidosis) or impairment of sperm penetration has been found.

**HYPER–IMMUNOGLOBULIN D SYNDROME (MEVALONATE KINASE DEFICIENCY)**

**Epidemiology**

Hyper-IgD syndrome (HIDS) (MIM 260920), also known as mevalonate kinase deficiency, is an autosomal recessively inherited disorder, but it is far less prevalent than FMF. The International Hyper-IgD Syndrome Registry, based in Nijmegen, the Netherlands, in which clinical information is collected from physicians worldwide, currently holds data on approximately 220 patients. Approximately 75% of these patients are from Western Europe, and 50% are from the Netherlands and France. Most HIDS patients are of Caucasian origin. These observations can be explained partly by a founder effect. In the Netherlands, the carrier frequency of the most common mevalonate kinase mutation is 1 : 153. Men and women are affected in equal numbers.

**Etiology**

HIDS is caused by mutations in the gene encoding for the enzyme mevalonate kinase, located on the long arm of chromosome 12 (for details, see the online mutation database available at [http://fmf.igh.cnrs.fr/infevers/](http://fmf.igh.cnrs.fr/infevers/)). Patients with classic HIDS are most often compound heterozygotes for two missense mutations (Figure 97-5). Two mutations (V377I and I268T) account for more than 85% of the patients described to date.

The term variant type HIDS, which we proposed for patients with an autoinflammatory disease and high immunoglobulin IgD without mevalonate kinase deficiency, has been largely abandoned. Many patients with fever syndromes have a raised IgD and this subclass is heterogeneous. It seems more useful to designate these patients as “autoinflammatory disease not otherwise specified” to indicate that a more specific diagnosis may still be found in the future.

**Pathogenesis**

Mevalonate kinase is part of the isoprenoid pathway; it is the step after 3-hydroxy-3-methylglutaryl–coenzyme A (HMG-CoA) reductase, phosphorylating mevalonic acid. The isoprenoid pathway has many diverse end products that include cholesterol, dolichol, and ubiquinone, and it leads to isoprenylation of proteins, with a post-translational modification directing these proteins such as Rho and Ras to the cell membrane.

HIDS mutations lead to a constantly diminished activity of mevalonate kinase to about 5% to 15% of normal levels, and these levels decrease further during a fever attack. Because of this reduced enzyme activity, the substrate mevalonic acid accumulates in serum and urine. Higher levels are found during the episodes of fever. There does not seem to be a dramatic shortage of any specific end product; concentrations of cholesterol, ubiquinone, and dolichol in patients are normal to slightly decreased.

Another syndrome was already linked to mutations in the mevalonate kinase gene before the discovery of HIDS—classic mevalonic aciduria. Patients with mevalonic aciduria carry specific mutations that cause a more severe reduction of mevalonate kinase enzyme activity, reducing it to undetectable levels. These patients constantly produce large amounts of mevalonic acid and often have more than 1000 times as much mevalonic acid in their urine than do HIDS patients. Patients with mevalonic aciduria also have a more severe phenotype. Classic mevalonic aciduria and HIDS seem to be two extremes of a continuous spectrum of disease related to mevalonate kinase deficiency.

The pathogenetic link between mevalonate kinase deficiency and inflammation is still unclear, but there is increasing evidence for a connection between the isoprenoid pathway and inflammation. Inhibition of the isoprenoid pathway by statins, the inhibitors of HMG-CoA reductase (the enzymatic step before that of mevalonate kinase), can have anti-inflammatory effects, ranging from increased apoptosis of inflammatory cells to reduction of expression of cytokines. In other settings, statins seem to be proinflammatory, most notably in a study in which stimulation with Mycobacterium tuberculosis or mitogens in combination with statins increased caspase-1 activation and IL-1β secretion by monocytes, through a decrease of geraniolgeraniol.

The ex vivo production of IL-1β is increased in HIDS.

*Figure 97-5* Mevalonate kinase, with four conserved domains represented by colored boxes. Indicated are missense mutations, nonsense mutations, and two deletions, which have been identified in mevalonate kinase deficiency. In bold are mutations found in mevalonic aciduria patients; in bold and italic are mutations found in classic hyper-IgD syndrome and mevalonic aciduria. For complete listing of all currently known mutations, see INFEVERS website: [http://fmf.igh.cnrs.fr/infevers/](http://fmf.igh.cnrs.fr/infevers/).
Clinical Features

Ninety percent of patients with HIDS experience their first fever episode in the first year of life, and these episodes become most frequent in childhood and adolescence. The high fevers may lead to seizures, especially in young children. Vaccination, minor trauma, surgery, and physical or emotional stress are factors that provoke a fever episode, although often a triggering factor is not obvious. The fevers often begin with cold chills and a sharp increase in body temperature. They are almost always accompanied by cervical lymphadenopathy and abdominal pain with vomiting and diarrhea. Other frequent symptoms are headache, myalgia, and arthralgia. Apart from the lymphadenopathy, physical signs frequently consist of splenomegaly and a skin rash with erythematous macules and papules (Figure 97-6) or petechiae (Figure 97-7). Sometimes there are also signs of frank arthritis (principally large joints) and hepatomegaly. About 40% of patients report painful aphthous ulcers in the mouth, vagina, or scrotum (Figure 97-8). The fever disappears spontaneously after 3 to 5 days, although it may take longer before the symptoms in joints or skin disappear completely. These inflammatory attacks occur, on average, once every 4 to 6 weeks, although this may vary from patient to patient or in an individual patient.

Patients with mevalonic aciduria, the metabolic disorder that is also caused by mevalonate kinase gene mutations, experience similar but more severe inflammatory episodes as HIDS patients. A decrease in apoptosis would result in increased survival of lymphocytes and may delay the resolution of the inflammatory response. An ordinarily innocuous stimulus in HIDS patients would more easily lead to a full-blown fever episode.

The cause of the characteristic high serum concentrations of IgD in this syndrome, which have led to its name, is still unexplained.
Table 97-4 Diagnostic Indicators of Hyper–Immunoglobulin D Syndrome

<table>
<thead>
<tr>
<th>At Time of Attacks</th>
<th>Specific Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevated erythrocyte sedimentation rate and leukocytosis</td>
<td>Mutations in mevalonate kinase gene</td>
</tr>
<tr>
<td>Abrupt onset of fever (≥38.5°C)</td>
<td>Decreased mevalonate kinase enzyme activity</td>
</tr>
<tr>
<td>Recurrent attacks</td>
<td></td>
</tr>
<tr>
<td>Lymphadenopathy (especially cervical)</td>
<td>*Extremely high serum concentrations of IgD are characteristic but not obligatory.</td>
</tr>
<tr>
<td>Abdominal distress (e.g., vomiting, diarrhea, pain)</td>
<td></td>
</tr>
<tr>
<td>Skin manifestations (e.g., erythematous macules and papules)</td>
<td></td>
</tr>
<tr>
<td>Arthralgias and arthritis</td>
<td></td>
</tr>
<tr>
<td>Splenomegaly</td>
<td></td>
</tr>
<tr>
<td><strong>Constantly Present</strong></td>
<td></td>
</tr>
<tr>
<td>Elevated IgD (≥100 U/mL) measured on 2 occasions at least 1 mo apart*</td>
<td></td>
</tr>
<tr>
<td>Elevated IgA (≥2.6 g/L)</td>
<td></td>
</tr>
</tbody>
</table>

Diagnosis and Diagnostic Tests

HIDS is diagnosed on the basis of a combination of characteristic clinical findings and continuously elevated IgD concentrations (>100 IU/mL) (Table 97-4). There are numerous caveats concerning IgD serum concentration, however: Values may be normal in young patients (especially patients younger than 3 years old), persistently normal levels have been reported in a few patients with HIDS, and patients with other familial autoinflammatory syndromes may also have elevated IgD concentrations, although these are usually only slightly elevated. More than 80% of HIDS patients combine a high concentration of IgD with high IgA levels. During fever attacks, a brisk acute-phase response is observed including leukocytosis, high levels of SAA and CRP, and activation of the cytokine network. The diagnosis of HIDS can be confirmed by DNA analysis of the mevalonate kinase gene. The best approach is to start with screening for the two most prevalent mutations, V377I and I268T. If this screening is negative, but the clinical suspicion remains high, sequencing of the entire gene can be considered. A good alternative is the measurement of urinary mevalonic acid concentrations during an attack, which are slightly elevated. Gas chromatography–mass spectroscopy is necessary to detect this slight increase, however. The measurement of mevalonate kinase enzyme activity is complicated and time-consuming and should be reserved for research purposes.

Treatment

There is no established treatment regimen for HIDS. Anakinra is effective in reducing disease severity in a number of case reports and is currently the most promising therapy. A double-blind, placebo-controlled, crossover trial of the HMG-CoA reductase inhibitor simvastatin showed a beneficial effect of this drug, with a reduction in number of days of illness in five out of six patients; however, in clinical practice the beneficial effect is not impressive. Favorable preliminary experience with the TNF antagonist etanercept has been reported.

Some individual patients have been reported to have benefited from treatment with corticosteroids, colchicine, intravenous immunoglobulin, or cyclosporine, but these results have not been repeated in most patients. Thalidomide did not have an effect on disease activity in a placebo-controlled trial.

Outcome

The long-term outcome in HIDS is relatively benign in most patients. In some patients, the fever episodes occur less frequently and become less severe later in life, starting from late adolescence. Joint destruction is rare, but abdominal adhesions are seen, resulting from repeated abdominal inflammation or (unnecessary) diagnostic laparotomy because of suspected acute abdomen.

Until more recently, no cases of amyloidosis had been seen in HIDS patients since its first description in 1984. Since 2004, four HIDS patients 19 to 27 years old have been reported who developed renal failure because of AA amyloidosis. Regular screening for proteinuria also may be advisable in HIDS patients, especially patients with frequent and severe fever episodes.

TUMOR NECROSIS FACTOR RECEPTOR–ASSOCIATED PERIODIC SYNDROME

Epidemiology

TRAPS (MIM 142680) has an autosomal dominant inheritance pattern. It was originally described in a large family from Irish and Scottish descent as “familial Hibernian fever.” It is found primarily in patients from northwestern Europe but also has been described in families from Australia, Mexico, Puerto Rico, Portugal, and the Czech Republic. Any ethnic group may be affected. Other abandoned nomenclature for this syndrome includes “autosomal dominant familial periodic fever” and “familial perirecticular amyloidosis.”

Etiology

Mutations are found in the gene for the type I TNF receptor (TNFRSF1A), which is located on the short arm of chromosome 12. These are mainly single-nucleotide missense substitutions, located in exons 2, 3, and 4, which encode for the extracellular domain of TNFRSF1A. Many of these mutations disrupt one of the highly conserved cysteine residues involved in extracellular disulfide bonds of the 55-kD type I TNF receptor protein (Figure 97-9) (for details, see the online mutation database available at http://fmf.igh.cnrs.fr/infevers/).

There are some general genotype-phenotype correlations, especially when mutations are grouped in cysteine and noncysteine mutations. Noncysteine mutations have, overall, a lower penetrance than cysteine mutations, and amyloidosis is seen far more often in association with cysteine mutations. Two missense mutations in TNFRSF1A, P46L and R92Q, have a particularly low penetrance and
are found in approximately 1% to 10% of control chromosomes.78-80 R92Q has been observed in higher prevalence in a group of patients with arthritis. It is thought that the clinical manifestations of patients with an R92Q mutation depend on other so-far-unidentified modifying genes, environmental factors, or both.77,78

Pathogenesis

TNF is a pleiotropic molecule, which induces cytokine secretion, activation of leukocytes, fever, and cachexia. Activation of the TNF receptor by TNF causes cleavage and shedding of its extracellular part into the circulation, where it acts as an inhibitor of TNF. However, TRAPS-associated mutations in the TNF receptor lead not to an increase but rather to a loss of TNF-signaling function including less binding of TNF,81,82 less cell surface expression,82,83 and decreased TNF-induced NfκB-activation.84,86,87 The mutated TNFRSF1A is retained intracellularly, pooled in the endoplasmic reticulum,82,83,87,88 Mutant TNFR1 cannot associate with the wild-type version but can form aggregates by self-interaction.82,83 This cytoplasmic receptor aggregation results in ligand-independent signaling.87,89 Mitochondrial-derived reactive oxygen species appear to mediate this effect.90 This new hypothesis also might offer an explanation for the observation that blocking IL-1β works better in some TRAPS patients than blocking TNF.87,91

An alternative hypothesis, the “shedding hypothesis,” was postulated on the finding of reduced shedding, which leads to prolonged TNF signaling and uncontrolled inflammation. Not all TRAPS mutations cause decreased shedding, however, and although serum concentrations of the shedded soluble TNFRSF1A in TRAPS patients during periods without symptoms are often found to be significantly reduced compared with normal subjects, this is not always the case.4 The hypothesis of reduced shedding, although attractive by its simplicity, is not supported as the sole cause of the fever attacks in TRAPS, and additional mechanisms seem to be at work.

Clinical Features

The clinical features can vary much more between individual TRAPS patients than is generally seen in FMF or HIDS.74,77 The age of onset can vary, even within the same family, with a documented range of 2 weeks to 53 years old.74,92 There is also a large variation in duration and frequency of the fever episodes in TRAPS. On average, attacks last 3 to 4 weeks and recur two to six times each year, but episodes also may be limited to a few days (see Figure 97-1).

Although the index patient, through whom the diagnosis is made, often displays well-defined inflammatory attacks, affected family members may have less typical symptoms such as episodic mild arthritis. During inflammatory attacks, a high, spiking fever can be accompanied by skin lesions, myalgia and arthralgia, abdominal distress, and ocular symptoms. The most common cutaneous manifestation is a centrifugal, migratory, erythematous patch, which may overlie a local area of myalgia (Figure 97-10),93 but urticarial plaques also may be seen.
Myalgia is often located primarily in the muscles of the thighs, but it may migrate during the fever episode, affecting all of the limbs and the torso, face, and neck. Arthralgia primarily affects large joints including hips, knees, and ankles. Frank synovitis is rarer, and when it does occur it is nonerosive, asymmetric, and monoarticular.\textsuperscript{74} Abdominal pain occurs in 92\% of TRAPS patients during inflammatory attacks; other gastrointestinal symptoms often seen include vomiting and constipation. Ocular involvement is characteristic in TRAPS, and it may involve conjunctivitis, periorbital edema, or periorbital pain in one or both eyes. Severe uveitis and iritis have been described, and any TRAPS patient with ocular pain should be examined for these complications.\textsuperscript{74,93} Other, less frequently observed symptoms during fever attacks in TRAPS are chest pain, breathlessness, pericarditis, and testicular and scrotal pain, which may be caused by inflammation of the tunica vaginalis. One case report described a patient who presented with psychosis without fever.\textsuperscript{94} It has been suggested from observation in one of the first families with TRAPS that this disorder is associated with an increased incidence of indirect inguinal hernias,\textsuperscript{95} but this has not been shown in other patients. Lymphadenopathy is rare in TRAPS.

**Diagnosis and Diagnostic Tests**

As in the other familial autoinflammatory syndromes, laboratory investigations during inflammatory attacks show a clear acute-phase response, and even in between fever attacks, such an inflammatory response may be measured. The IgD level may be elevated, but the value is almost always less than 100 IU/mL.\textsuperscript{92,95}

Hull and colleagues\textsuperscript{92} proposed a set of clinical diagnostic criteria for TRAPS (Table 97-5). These criteria are not validated by epidemiologic measures, but they may be used as a first step in evaluation of patients. TRAPS is ultimately a genetic diagnosis, defined by a missense mutation in the gene for TNFRSF1A. Clinical penetrance of TRAPS mutations is not 100\%, however, even for cysteine mutations, and asymptomatic carriers are common. Also, the finding of an R92Q or P46L variant in this gene would pose a difficulty. Because they have many characteristics of a polymorphism rather than a direct disease-causing mutation (see etiology), it is debatable whether such a finding should lead to a diagnosis of TRAPS.

**Table 97-5  Diagnostic Indicators of Tumor Necrosis Factor Receptor–Associated Periodic Syndrome**

<table>
<thead>
<tr>
<th>1. Recurrent episodes of inflammatory symptoms spanning ( \geq 6 ) mo duration (several symptoms generally occur simultaneously)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Fever</td>
</tr>
<tr>
<td>b. Abdominal pain</td>
</tr>
<tr>
<td>c. Myalgia (migratory)</td>
</tr>
<tr>
<td>d. Rash (erythematous macular rash occurs with myalgia)</td>
</tr>
<tr>
<td>e. Conjunctivitis or periorbital edema</td>
</tr>
<tr>
<td>f. Chest pain</td>
</tr>
<tr>
<td>g. Arthralgia or monoarticular synovitis</td>
</tr>
<tr>
<td>2. Episodes last ( \geq 5 ) days on average (although variable)</td>
</tr>
<tr>
<td>3. Responsive to glucocorticosteroids but not colchicine</td>
</tr>
<tr>
<td>4. Affects family members in autosomal dominant pattern (although may not always be present)</td>
</tr>
<tr>
<td>5. Any ethnicity may be affected</td>
</tr>
</tbody>
</table>


**Treatment**

In mild cases of TRAPS, NSAIDs are often sufficient. NSAIDs and glucocorticoids in high doses (\( \geq 20 \) mg/day of oral prednisone) alleviate the symptoms of fever and inflammation in most TRAPS patients, although they do not alter the frequency of attacks. They can be used beneficially at times of attack, and glucocorticoids usually can be tapered in the course of 1 or 2 weeks, as tolerated. There is no response to colchicine or immunosuppressive drugs such as azathioprine, cyclosporine, thalidomide, or cyclophosphamide.\textsuperscript{74}

Anti-TNF and anti-IL-1 biologic therapy is much more effective in reducing symptoms. Use of etanercept, a fusion product of TNFRSF1B (the receptor that is not defective) has been partially successful.\textsuperscript{74,76,86,97} A study with twice-weekly administration of etanercept (25 mg for adults or 0.4 mg/kg for children) in nine TRAPS patients with various mutations revealed an overall 66\% response rate as determined by decreased number of attacks over a 6-month period.\textsuperscript{74} Another study with the same dosage of etanercept for 24 weeks in seven TRAPS patients also showed a clear beneficial effect without serious adverse events.\textsuperscript{96} A similar regimen of etanercept reversed the nephrotic syndrome in a patient with amyloidosis.\textsuperscript{98}

Drewe and colleagues\textsuperscript{99} described one patient whose symptoms were resistant to administration of etanercept, who responded favorably to use of oral sirolimus (4 to 6 mg daily). Infliximab, a monoclonal antibody against TNF, has been shown to be less effective than etanercept in TRAPS and seems to cause increased symptoms.\textsuperscript{96,99,100} Intravenous infusion of a synthetic TNFRSF1A fusion protein was tried in one patient by Drewe and colleagues\textsuperscript{96} but this seemed to provoke a severe attack.

Blocking IL-1\( \beta \) with anakinra is even more effective than blocking TNF in several case reports and case series.\textsuperscript{91,101-103}

**Outcome**

Reactive AA amyloidosis is the principal systemic complication of TRAPS. It occurs in about 15\% to 25\% of patients\textsuperscript{78,104} and generally leads to renal impairment. Amyloidosis in a patient with TRAPS places other affected family members at high risk for this complication. It is principally associated with TNFRSF1A mutations affecting cysteine residues.\textsuperscript{71} Because proteinuria is the initial manifestation of renal amyloidosis, it is advisable to screen urine samples from TRAPS patients regularly by dipstick examination, especially affected family members of a TRAPS patient with amyloidosis.

**CRYOPYRIN-ASSOCIATED PERIODIC SYNDROME**

Cryopyrin-associated periodic syndrome (CAPS) encompasses three clinical syndromes that all have been traced to
mutations in one common gene: Muckle-Wells syndrome (MWS), familial cold autoinflammatory syndrome (FCAS), and chronic infantile neurologic cutaneous and articular syndrome (CINCA), also known as neonatal-onset multisystemic inflammatory disease (NOMID). After the recognition of the genetic defect, it became clear that there is considerable overlap among these three disorders.\textsuperscript{107} FCAS, MWS, and CINCA/NOMID might represent a spectrum of disease, with FCAS the mildest and CINCA the most severe form. Given their common genotype, they are discussed under one heading here, but clinical features and outcomes are dealt with separately.

**Epidemiology**

All three syndromes are rare, autosomal dominantly inherited syndromes. Most articles on FCAS, first described in 1940, describe large families from Europe and North America with extensive pedigrees, but sporadic cases have been described. There seems to be a founder effect in American families of Northern European extraction.\textsuperscript{106} MWS was first described in 1962 and has since been described in large families, although it does occur in isolated cases and small nuclear families. CINCA is rare, and, to date, some 70 cases and only a few families have been described.

**Etiology**

The first indications that MWS and FCAS are allelic stem from early linkage studies showing that FCAS and MWS were linked to the same region on the long arm of chromosome 1 (1q44).\textsuperscript{107,108} In 2001, the gene for FCAS and MWS was identified. In a large-scale, positional cloning effort using three families with FCAS and one family with MWS, missense mutations in a new gene were found.\textsuperscript{109} This gene, CIAS1, encodes for a protein denoted cryopyrin at discovery, but now by consensus known as NLRP3 (other names that have been used are NALP3 and PYPAF). Later studies showed that CIAS1 mutations were also associated with CINCA.\textsuperscript{110,111} Practically all mutations are missense mutations found in exon 3 of the CIAS1 gene, which encodes for the NOD domain of NLRP3.\textsuperscript{112} (Figure 97-11) (for details, see the online mutation database available at http://fmf.igh.cnrs.fr/infevers/).

**Pathogenesis**

NLRP3 or cryopyrin was a previously unknown protein at the time of the discovery of the mutations involved in these syndromes. Since that time, it has become the focus of numerous studies, which have led to a new concept—the inflammasome.\textsuperscript{113} NLRP3 is a member of the nucleotide-binding oligomerization domain–leucine-rich repeat (NOD-LRR) protein family.\textsuperscript{114} It consists of a pyrin domain (PYD), a NOD (also known as NACHT domain), and a LRR domain. NLRP3 is mainly expressed in monocytes and neutrophils but is also found in human chondrocytes.\textsuperscript{109,111,115}

NLRP3 is thought to be an intracellular sensor of pathogens or danger signals, regulating innate immunity. On stimulation with various ligands, which include bacterial RNA, imidazoquinolone compounds, gram-positive bacterial toxins nigericin and maitotoxin, adenosine triphosphate, and uric acid crystals, NLRP3 forms interactions with adapter proteins ASC and cardinal, which results in a multiprotein complex termed the NLRP3 inflammasome. The NLRP3 inflammasome activates caspase-1, which subsequently cleaves pro-IL-1β to the active IL-1β.\textsuperscript{116,117}

There is conflicting evidence for a role of NLRP3 in regulation of transcription factor NFkB.\textsuperscript{1} Possibly, the ultimate effect on NFkB is determined by interaction of multiple proteins. Four more recent publications by independent groups that each developed an NLRP3-deficient mouse showed no effect, however, on NFkB activation in these mice, whereas they did show a clear deficiency in caspase-1–mediated IL-1β activation.\textsuperscript{1}

Monocytes from patients with mutations in the NOD of NLRP3 show increased activation of caspase-1 and subsequently increased release of IL-1β.\textsuperscript{117,118} The key role of IL-1β in CAPS is confirmed by the success of treatment with the IL-1 blocker anakinra in all three clinical syndromes (see treatment section).

The exact effect of the NLRP3 mutations is still unclear. An attractive hypothesis involves a possible autoinhibitory loop of NLRP3.\textsuperscript{118} Mutations in the NOD could interfere

---

**Figure 97-11** Cryopyrin protein, containing an N-terminal pyrin domain, a nucleotide binding site (NACHT), and a leucine-rich repeat (LRR) domain. Indicated are missense mutations identified in patients with familial cold autoinflammatory syndrome (red), Muckle-Wells syndrome (green), or chronic infantile neurologic cutaneous and articular syndrome (blue), and mutations found in common in two or all of these clinical syndromes (black). For complete listing of all currently known mutations, see INFEVERS website: http://fmf.igh.cnrs.fr/infevers/.
seen. Most patients (84%) also report conjunctivitis during a fever episode. Other symptoms include myalgia, profuse sweating, drowsiness, headache, extreme thirst, and nausea.

A typical feature of FCAS is the requirement of cold exposure to trigger the symptoms. The delay between cold and onset of symptoms varies from 10 minutes to 8 hours. When Hoffman and colleagues provoked an inflammatory attack in FCAS patients by generalized cold exposure in a cold room, they saw that patients developed rash, fever, and arthralgia within 1 to 4 hours. The occurrence of these symptoms could be blocked by pretreatment with the IL-1β inhibitor anakinra.

The subsequent fever attack varies in length, depending on the degree of cold exposure; generally it lasts a few hours to a maximum of 3 days. These episodes start at an early age, with 95% of patients having had their first fever episode in the first year of life—60% even within the first days of life. The symptoms tend to become less severe with advancing age.

Type AA amyloidosis complicated by renal insufficiency has been described in at least three FCAS families.

**Muckle-Wells Syndrome**

MWS (MIM 191900) is a rare autosomal dominant inflammatory disorder with incomplete penetrance. Patients have recurrent episodes of fever, abdominal pain, myalgia, urticarial rash (Figures 97-14 and 97-15), and conjunctivitis, frequently accompanied by arthralgia, arthritis with limb pain, or both. Attacks start in adolescence and can be provoked by hunger, fatigue, and sometimes exposure to cold. The inflammatory episodes generally last 24 to 48 hours (see Figure 97-1) and start with ill-defined malaise and transient chills and rigor, followed by aching or lancinating pains in the distal limbs and larger joints. Arthralgia is a common feature of the attacks, but synovitis of the large joints is less common. The rash consists of usually aching and sometimes pruritic erythematous papules 1 to 7 cm in diameter. In a few cases, genital and buccal aphthous ulcers have been seen. Ocular symptoms include uveitis and conjunctivitis. Symptoms typically start in adolescence, although they have been reported at an earlier age. Late-onset development of perceptive deafness is common in MWS. Bone

**Clinical Features and Outcome**

**Familial Cold Autoinflammatory Syndrome**

FCAS (MIM 120100) is characterized by episodes of rash, fever, and arthralgia after generalized exposure to cold (see Figure 97-1). The disease occurs in large families as an autosomal dominant inherited disorder with almost complete penetrance. The rash usually starts on the exposed extremities and, in most episodes, extends to the remainder of the body. It consists of erythematous macules and plaques (Figures 97-12 and 97-13), urticarial lesions, and sometimes petechiae and can cause a burning or itchy sensation. In one case report, FCAS was associated with Raynaud’s disease. In some cases, localized edematous swelling of extremities is reported. Arthralgia, present in 93% of cases, most often affects the hands, knees, and ankles but can also involve feet, wrists, and elbows. Frank arthritis is not

**Figure 97-12** Fine, confluent, erythematous macules on the upper leg of a patient with familial cold autoinflammatory syndrome. (Courtesy Dr. Johnstone, Medical College of Georgia, Augusta, Ga.)

**Figure 97-13** Detail of upper leg with fine, confluent, erythematous macules in familial cold autoinflammatory syndrome. (Courtesy Dr. Johnstone, Medical College of Georgia, Augusta, Ga.)

**Figure 97-14** Urticarial skin rash in a patient with Muckle-Wells syndrome. (Courtesy Dr. D. L. Kastner, National Institutes of Health, Bethesda, Md.)
involvement such as clubbing of nails and pes cavus can be seen as well. Most often, patients have a positive family history for the disease, which is indicative of autosomal dominant inheritance, but isolated cases have been reported. The most feared complication of the inflammatory attacks is type AA amyloidosis, which affects the kidneys first, leading to proteinuria and subsequent rapid progression to renal failure.

**Chronic Infantile Neurologic Cutaneous and Articular Syndrome**

CINCA or NOMID (MIM 607115) is a rare congenital disorder defined by the presence of the triad of (1) neonatal-onset skin lesions, (2) chronic aseptic meningitis, and (3) recurrent fever along with joint symptoms. The key clinical feature of CINCA is a skin rash accompanied by peculiar joint manifestations and central nervous system involvement. The symptoms in CINCA begin right after birth or in the first months of life with a generalized skin rash. The disease follows an unpredictable course with persistent non-pruritic and migratory rash with fever, hepatosplenomegaly, and lymphadenopathy. Central nervous system involvement is not obvious from the outset, although some patients present with seizures, spasticity, or transient episodes of hemiplegia. In most patients, there are signs of chronic persistent aseptic meningitis. Cerebrospinal fluid analysis may show mild pleocytosis, and there may be increased intracranial pressure. Brain imaging shows mild ventricular dilation, prominent sulci, central atrophy, and, in longstanding cases, calcifications of fauces and dura.

In older children, headache is often a prominent feature as a sign of chronic meningitis. Mental retardation is present in almost all cases. Progressive sensorineural impairment leading to high-frequency hearing loss can be seen in a few cases. Ocular manifestations are prominent, with optic disc changes such as optic disc edema, pseudopapilledema, and optic atrophy, as well as anterior segment manifestations such as chronic anterior uveitis. These symptoms may lead to visual impairment. Hoarseness, especially in older children, is typical. Joint and bone symptoms are a prominent feature of CINCA, and these manifest as bone inflammation, which gives rise to major arthropathies secondary to epiphyseal and metaphyseal disorganization. Growth cartilage alterations such as enlarged epiphyses and patellar overgrowth can be an impressive feature of the disease (Figures 97-16 and 97-17). Erosive changes occur, especially in the phalanges of hands and feet. There are typical dysmorphic features such as frontal bossing and a saddle nose. These common physical features are the reason CINCA patients give the impression that totally unrelated patients are siblings. The prognosis of these patients is grave; 20% die in childhood because of infections, vasculitis, and amyloidosis.

**Diagnosis and Diagnostic Tests**

Diagnosis starts with a thorough patient and family history (see Table 97-1). Hoffman and colleagues suggested a set...
Table 97-6  Diagnostic Criteria for Familial Cold Autoinflammatory Syndrome

1. Recurrent intermittent episodes of fever and rash that primarily follow generalized cold exposures
2. Autosomal dominant pattern of disease inheritance
3. Age of onset < 6 mo
4. Duration of most attacks <24 hr
5. Presence of conjunctivitis associated with attacks
6. Absence of deafness, periorbital edema, lymphadenopathy, and serositis


BLAU SYNDROME/EARLY-ONSET SARCOIDOSIS

Epidemiology

Blau syndrome, also known as familial granulomatous arthritis (OMIM 186580), and early-onset sarcoidosis (OMIM 609464) (BS/EOS) are now recognized as the same disorder. “Pediatric granulomatous arthritis” has been suggested as a new name to describe this syndrome, although this might, erroneously, give the impression that the disease occurs only in children. Little is known about its epidemiology, although it is thought to occur worldwide.

Etiology

The inheritance pattern of BS/EOS is autosomal dominant. In many cases, a de novo mutation is found, which explains the relatively high incidence of sporadic cases. These sporadic cases were often classified as early-onset sarcoidosis precisely because of the absence of affected relatives, but Blau syndrome and early-onset sarcoidosis have now been shown to be caused by mutations in the nucleotide-binding oligomerization domain 2/caspase recruitment domain 15 gene (NOD2/CARD15). Mutations are mostly located in exon 4 of NOD2/CARD15. The predominant mutations are two missense mutations at position 334 (R334Q and R334W) (for details, see the online mutation database available at http://fmf.igh.cnrs.fr/infevers/).

Pathogenesis

The NOD2/CARD15 protein is considered to be an intracellular sensor for pathogenic components, analogous to the Toll-like receptors. Activation of NOD2/CARD15 results in a wide array of downstream effects that are still not well understood, including activation of NFκB and mitogen-activated protein kinase pathways, turning on an innate immune response of diverse cytokines (e.g., IL-1β) and defensins.

Seven of the nine different mutations of NOD2/CARD15 linked to BS/EOS are located in the NOD domain of the protein, similar to the mutations in cryopyrin (NLRP3) in the cryopyrin-associated syndromes. The two most common mutations affect a codon at a homologous position in the NOD domain to the location of the cryopyrin R260W mutation. This suggests a similar pathophysiologic effect on the function of the protein.

Polymorphisms in another part of this same NOD2/CARD15 gene on chromosome 16 are associated with increased susceptibility to Crohn’s disease; the risk of developing Crohn’s disease is increased 40-fold in individuals homozygous for these polymorphisms. Whether these polymorphisms result in a gain or loss of function of this protein is debated. There are some shared features between the two diseases: Both are characterized by granulomatous inflammation, and although bowel inflammation is not seen in Blau syndrome, Crohn’s disease can manifest with uveitis, arthritis, and skin rash.

of diagnostic criteria for FCAS after studying six large families with this syndrome (Table 97-6), but these have not been validated in an independent cohort. Laboratory examination during a fever episode in CAPS shows an acute-phase response with polymorphonuclear leukocytosis and increased erythrocyte sedimentation rate, but this does not differentiate among the periodic fever disorders. Symptoms such as an urticarial rash after cold exposure highly favor a diagnosis of FCAS. The ice cube test (i.e., holding an ice cube to a patch of skin to provoke urticaria), which is diagnostic in acquired cold urticaria, is negative in FCAS. Typical facial features such as frontal bossing and a long pediatric history including chronic aseptic meningitis point to CINCA/NOMID.

Genetic testing of the CIAS1 gene can subsequently help to establish the genetic diagnosis. Usually, exon 3 of this gene is screened for mutations. There seems to be genetic heterogeneity in CINCA because not all patients have CIAS1 mutations.

Treatment

Since the advent of IL-1 inhibition, it is the treatment of choice for patients with severe forms of cryopyrin-associated periodic syndrome. Previously, high-dose oral corticosteroids were often used and found to be beneficial in some patients. NSAIDs, disease-modifying antirheumatic drugs, and cytotoxic drugs generally do not help.

IL-1 inhibition by anakinra was shown to be beneficial for patients with any of the cryopyrin-associated periodic syndromes in a number of case reports and case series. The largest study was by Goldbach-Mansky and colleagues, who studied 18 patients with CINCA/NOMID, 12 of whom had mutations in the cryopyrin gene. All of them had a rapid and sustained response to daily subcutaneous injection of anakinra (1 to 2 mg/kg body weight), with decrease of symptoms, acute-phase response, and leptomeningeal lesions as seen on MRI. Mirault and colleagues described improvement of sensorineural deafness in a patient with MWS on treatment with anakinra.

Lachmann and colleagues published the results of a trial of the new IL-1β antibody canakinumab in patients with CAPS, demonstrating the efficacy of an injection of 150 mg canakinumab once every 2 months. The first reports of the new IL-1 inhibitor IL-1-trap are also promising.
Clinical Features and Outcome

The clinical phenotype of BS/EOS consists of recurrent granulomatous inflammations. The three typical sites affected are joints, eyes, and skin. The granulomatous arthritis is most often polyarticular, with a synovitis or tenosynovitis.\(^\text{134}\) The uveitis associated with this disorder tends to follow a chronic, persistent course. It can be an acute anterior uveitis, but it often extends to a panuveitis.\(^\text{136}\) Cataracts, secondary glaucoma, and significant visual impairment can result. Involvement of the skin results in a papular, erythematous skin rash with associated dermal granulomas, usually generalized and intermittent, on trunk and extremities.\(^\text{142}\) Other symptoms include campylodactyly (contracture of multiple interphalangeal joints), cranial neuropathies, fever, and arteritis.\(^\text{137}\) In some severely affected patients, granulomatous inflammation can disseminate at an advanced stage into a systemic disease, with granulomas in liver, lung, and kidney.\(^\text{136}\) Onset is generally before age 5. In familial cases, genetic anticipation is often observed (i.e., the course of disease tends to be more severe in later generations). The major long-term complications are joint deformity and visual impairment.\(^\text{134}\)

Diagnosis

The most important aspect of diagnosis is the histologic evidence of granulomas at the site of inflammation. This evidence can be obtained by biopsy of any involved site, of which skin is least invasive. One study showed that skin biopsy was diagnostic in all cases with the typical skin rash, whereas synovial biopsy was not positive in all patients, perhaps owing to sampling error.\(^\text{136}\) Genetic testing is available for NOD2/CARD15 mutations, but in some series, not all patients with a typical clinical phenotype carried a mutation in this gene.\(^\text{137}\)

Treatment

There are no controlled studies of management of BS/EOS patients. There tends to be a poor response to NSAIDs. A good response to the TNF inhibitor infliximab was reported in case reports,\(^\text{143,144}\) one of which also noted that etanercept did not have the same effect. IL-1 blockade by anakinra was also reported as beneficial in a case description,\(^\text{145}\) although in two other cases anakinra was not effective.\(^\text{146}\) Thalidomide was effective in two children in Japan.\(^\text{147}\) The panuveitis is usually managed by topical, subconjunctival, or systemic corticosteroids.\(^\text{134}\)

Etiology/Pathogenesis

Wise and colleagues\(^\text{149}\) identified mutations in the CD2-binding protein 1 (CD2BP1) gene as the cause of PAPA syndrome. CD2BP1, also known as proline-serine-threonine phosphatase interacting protein 1 (PSTPIP1), is highly expressed in neutrophils\(^\text{112}\) (see the online mutation database available at http://fmf.igh.cnrs.fr/infevers/).

PSTPIP1 can form interactions with pyrin, the protein mutated in FMF.\(^\text{150}\) The mutations in PAPA syndrome result in hyperphosphorylation of PSTPIP1,\(^\text{150-152}\) which increases the strength of the interaction between PSTPIP1 and pyrin. This increased interaction of PSTPIP1 and pyrin leads to increased IL-1β production. This activity correlated with a higher IL-1β production in response to lipopolysaccharide stimulation of peripheral blood leukocytes from a PAPA patient ex vivo compared with a healthy control.\(^\text{150}\) It places PAPA syndrome in the same pathogenic pathway as FMF. Other studies report an increased production of TNF.\(^\text{153,154}\) Dysregulated apoptosis may also be involved.\(^\text{155}\)

Clinical Features and Outcome

The episodic inflammation in this syndrome includes, as the name aptly indicates, symptoms of pyogenic sterile arthritis, pyoderma gangrenosum, and severe cystic acne. Lesions generally occur at the site of mild physical trauma, but sometimes no obvious trigger can be discerned.\(^\text{148}\) The inflammation can be severe and eventually may lead to destruction of joints, muscle, and skin. Fever is not prominent in this syndrome. Onset is usually from age 1 to 16 years.\(^\text{148,156}\) The acne generally starts early in puberty and persists in adulthood.

Diagnosis

No specific diagnostic test exists. Diagnosis is based on a finding of the typical constellation of symptoms and a positive family history. A specialized DNA diagnostics department would be able to perform the genetic test for PAPA syndrome. At this time, it is unknown whether this genetic test would detect all patients or whether other genes could be involved.

Treatment

Diverse anecdotal evidence is only available on treatment options in PAPA syndrome. High-dose steroids generally have a positive effect on the pyoderma gangrenosum but may be associated with increased acne.\(^\text{148}\) Pyogenic arthritis is often responsive to glucocorticoids intra-articularly and orally.\(^\text{156}\) Varying results have been reported with anticytokine treatment. The TNF inhibitor etanercept was reported to be beneficial,\(^\text{153,157}\) as was the IL-1 inhibitor anakinra.\(^\text{150,158,159}\) Stichweh and colleagues\(^\text{160}\) reported on a patient with severe pyoderma gangrenosum, however, who did not respond to either etanercept or anakinra, but in whom the other TNF inhibitor infliximab did prove successful.

PYOGENIC STERILE ARTHRITIS, PYODERMA GANGRENOsum, AND ACNE SYNDROME

Epidemiology

Pyogenic sterile arthritis, pyoderma gangrenosum, and acne (PAPA) syndrome (MIM 604416) is an autosomal dominant disorder first described by Lindor and colleagues.\(^\text{148}\) Fewer than 20 families have been reported.
DEFICIENCY OF THE IL-1 RECEPTOR ANTAGONIST

Deficiency of the IL-1 receptor antagonist (DIRA) is the latest discovery of a monogenetic disorder in the group of autoinflammatory diseases. It was found in nine patients from six families, one from Newfoundland, Canada, three from the Netherlands, one from Puerto Rico, and one consanguineous family from Lebanon.

All patients were homozygous for mutations affecting the gene encoding the IL-1 receptor antagonist (IL-1ra), designated IL1RN. These mutations resulted in an absence of secretion of IL-1ra. IL-1ra is an endogenous inhibitor of the inflammatory cytokines IL-1α and IL-1β. The patients showed signs of unopposed IL-1 signaling leading to overproduction of other inflammatory cytokines and chemokines.

The clinical phenotype is present at birth or in the first month of life. Features include primarily cutaneous pustulosis and sterile pustular osteomyelitis. Patients do not present with fever. Two children died of multiorgan failure secondary to severe inflammatory response syndrome (SIRS) at the ages of 2 months and 21 months, and a third child died at 9.5 years of complications of pulmonary hemosiderosis with progressive intestinal fibrosis.

The diagnosis is made by a combination of the clinical features and an impressive response to anakinra. Anakinra is a synthetic version of IL-1ra and thus supplements the missing protein in DIRA. Treatment with anakinra suppresses all disease symptoms, although in one patient inflammatory markers remained elevated.

CONCLUSION

The familial autoinflammatory syndromes are characterized by recurrent episodes of fever and inflammation. This group of disorders should be considered in a patient with a history of years of such inflammatory attacks with symptom-free intervals in between (except for CINCA/NOMID, in which some symptoms and morphologic features persist). Dysregulation of the IL-1β pathway is central to many familial autoinflammatory syndromes, especially CAPS, FMF, and DIRA. Increasing availability of IL-1 inhibitors has revolutionized treatment options in many of these diseases.

The discovery of the causative genes has had an enormous impact in the field of periodic fevers. This discovery has been made possible because of the accurate phenotypic characterization of patients with periodic fever. Careful analysis and proper clustering of these patients is indispensable to allow the elucidation of the genetic background and the evaluation of possible treatment options (Table 97-7). Central periodic fever registries have afforded the opportunity to appreciate previously unrecognized symptoms, to give insight into the long-term prognosis, and to allow better evaluation of drug regimens. Despite these efforts at classification, however, many patients with periodic fever do not fall in one of the previously mentioned disease categories. It is to be expected that in the future other periodic fever syndromes and corresponding genes will be discovered.

Table 97-7 Summary of Treatment Options

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Treatment Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMF</td>
<td>Colchicine</td>
</tr>
<tr>
<td></td>
<td>In refractory cases or intolerance of colchicine: intravenous colchicine; IL-1 inhibition (anakinra)</td>
</tr>
<tr>
<td>HIDS</td>
<td>Simvastatin; etanercept; IL-1 inhibition (anakinra)</td>
</tr>
<tr>
<td>TRAPS</td>
<td>NSAIDs; etanercept; IL-1 inhibition (anakinra)</td>
</tr>
<tr>
<td>CAPS</td>
<td>IL-1 inhibition (anakinra, canakinumab, IL-1 trap)</td>
</tr>
<tr>
<td>BS/EOS</td>
<td>Corticosteroids; infliximab</td>
</tr>
<tr>
<td>PAPA</td>
<td>High-dose steroids; IL-1 inhibition (anakinra); etanercept; infliximab</td>
</tr>
</tbody>
</table>

BS/EOS: Blau syndrome/early-onset sarcoidosis; CAPS, cryopyrin-associated periodic syndrome; FMF, familial Mediterranean fever; HIDS, hyper-IgD syndrome; IL-1, interleukin-1; NSAIDs, nonsteroidal anti-inflammatory drugs; PAPA, pyogenic sterile arthritis, pyoderma gangrenosum, and acne; TRAPS, tumor necrosis factor receptor–associated periodic syndrome.

Selected References


References


Osteoarthritis (OA) is a degenerative joint disease, occurring primarily in older persons, characterized by erosion of the articular cartilage, hypertrophy of bone at the margins (i.e., osteophytes), subchondral sclerosis, and a range of biochemical and morphologic alterations of the synovial membrane and joint capsule. Pathologic changes in the late stages of OA include softening, ulceration, and focal disintegration of the articular cartilage; synovial inflammation may also occur. Typical clinical symptoms are pain and stiffness, particularly after prolonged activity.

In industrialized societies OA is the leading cause of physical disability, increases in health care usage, and impaired quality of life. The impact of arthritic conditions is expected to grow as the population both increases and ages in the coming decades. Despite its prevalence, the precise etiology, pathogenesis, and progression of OA remain beyond our understanding, primarily due to multiple confounding factors. Without a clear-cut picture of how OA arises at the cellular or molecular level, many still consider it a result of "wear and tear," an inevitable consequence of aging. In fact, the accompanying biochemical, structural, and metabolic changes in joint cartilage have been well documented. It is now known, for example, that cytokines, mechanical trauma, and altered genetics are involved in its pathogenesis and that these factors can initiate a degenerative cascade that results in many of the characteristic alterations of articular cartilage in OA. More recently it has become apparent that OA is a disease process that affects the entire joint structure including cartilage, synovial membrane, subchondral bone, ligaments, and periarticular muscles. OA is thus better considered a group of overlapping disorders of various etiologies arising from a combination of systemic factors (e.g., genetics) and local factors (e.g., biomechanically or biochemically mediated events) that gradually converge to produce a condition with definable morphologic and clinical outcomes.

OA may be classified as primary or secondary according to its cause or major predisposing factor; what they all have in common is altered cartilage physiology. Primary OA is the most common type and has no identifiable etiology or predisposing cause. Secondary OA, although it has an identifiable underlying cause, is pathologically indistinguishable...
from primary OA. The most common causes of secondary OA are metabolic conditions (e.g., calcium crystal deposition, hemochromatosis, acromegaly); anatomic factors (e.g., leg length inequality, congenital hip dislocation); traumatic events (e.g., major joint trauma, chronic joint injury, joint surgery); or the sequelae of inflammatory disorders (e.g., ankylosing spondylitis, septic arthritis). In the case of secondary OA arising from inflammatory joint disease, cartilage degeneration most likely results initially from degradative enzymes released from the synovium or leukocytes within the joint space and later from the mechanical attrition of a biomechanically altered extracellular matrix. Distinguishing between primary and secondary OA may be difficult because the clinical presentation and symptoms are often so similar.

**ETIOLOGIC FACTORS IN OSTEOARTHRITIS**

Major factors affecting degree of risk for developing OA include age, joint location, obesity, genetic predisposition, joint malalignment, trauma, and gender.

**Age**

Age is the risk factor most strongly correlated with OA. OA is the most common chronic disease in later life; more than 80% of persons older than age 75 years are affected, and OA increases progressively with age at all joint sites. Radiologic changes in OA increase as individuals age, although these changes do not always correlate with clinical symptoms or disability. Although clearly an age-related disease, OA is not, however, an inevitable consequence of aging. Age-related morphologic and structural changes in articular cartilage include fraying, softening, and thinning of the articular surface; decreased size and aggregation of matrix proteoglycans; and loss of matrix tensile strength and stiffness. These age-related tissue changes are most likely due to a decrease in the chondrocytes’ ability to maintain and repair the tissue, as chondrocytes themselves undergo age-related decreases in mitotic and synthetic activity, exhibit decreased responsiveness to anabolic growth factors, and synthesize smaller and less uniform large aggregating proteoglycans and fewer functional link proteins.

There also appears to be a direct correlation between chondrocyte apoptosis and cartilage degradation leading to OA. Age does appear to be an independent factor that predisposes articular chondrocytes to apoptosis because the expression levels of specific proapoptotic genes (Fas, FasL, caspase-8, and p53) is higher in aged cartilage.

**Obesity**

Obesity is another important risk factor for OA. Greater body mass index in both women and men has been shown to be associated with an increased risk of knee (but not hip) OA. An increase in mechanical forces across weight-bearing joints is probably the primary factor leading to joint degeneration. Obesity not only increases the forces at weight-bearing joints but may also change posture, gait, and physical activity level, any or all of which may further contribute to altered joint biomechanics. The majority of obese patients exhibit varus knee deformities, which results in increased joint reactive forces in the medial compartment of the knee, thereby accelerating the degenerative process.

Particularly in elderly obese individuals, heavy physical activity is an additional risk factor for the development of knee OA, whereas light to moderate activity does not appear to increase risk for knee OA and may in fact alleviate symptomatic knee OA by reducing body mass index. Similarly, weight loss can reduce both radiographic knee OA progression and clinical symptoms. Recent evidence indicates that in obese patients with OA, significant weight loss dramatically improves functional status, with short-term results equivalent to those of patients who have undergone joint replacement.

Yet recent work suggests that the association between obesity and osteoarthritis lies beyond the mechanical loading from a higher body mass index. Adipose tissue is now recognized as a metabolically active contributor to inflammatory cascades. Activated adipose tissue increases the synthesis of proinflammatory cytokines including leptin, adiponectin, resistin, interleukin-1 (IL-1), IL-6, and tumor necrosis factor (TNF), while some regulatory cytokine levels such as IL-10 are decreased. Specifically, the discovery of the obesity gene and its product leptin may have important implications for the onset and progression of OA and increase our understanding of the link between obesity and OA. The fact that women have a greater proportion of total body fat and relatively higher levels of adipose-derived systemic leptin concentrations than men may partially account for the gender disparity in OA patients. Leptin, however, is produced not only by adipose cells but also by osteoblasts and chondrocytes, suggesting that local leptin production may play a role in OA. Significant levels of leptin were observed in the cartilage and osteophytes of subjects with OA, yet few chondrocytes produced leptin in the cartilage of healthy subjects. Leptin has also been demonstrated to induce anabolic activity in the chondrocytes of rats and thus may ultimately confer structural joint changes.

**Joint Location**

Although osteoarthritis occurs most commonly in weight-bearing joints, age affects joints differentially. A study comparing tensile fracture stress of cartilage in the femoral head and in the talus showed that it decreased progressively with age in the former but not in the latter. Joint-specific age-related viability in articular cartilage may explain why OA is more common in hip and knee joints with increasing age but occurs rarely in the ankle.

**Genetic Predisposition**

Because of the prevalence of OA in the general population and its extensive clinical heterogeneity, the genetic contribution to its pathogenesis has been difficult to analyze. Two major cohorts (the Framingham Study and the Baltimore Longitudinal Study on Aging) support a genetic contribution to OA, with evidence for a major recessive gene and a multifactorial component, representing either polygenic or environmental factors. Twin pair and family risk
studies have indicated that the heritable component of OA may be on the order of 50% to 65%. Moreover, family, twin, and population studies have indicated differences among genetic influences that determine the site of OA (hip, spinal, knee, hand). Further evidence supporting a genetic predisposition to OA is the demonstration of a significantly higher concordance for OA between monozygotic twins than between dizygotic twins. Genetic studies have identified multiple gene variations associated with an increased risk of OA.

Inherited forms of OA may be caused by mutations in several genes that are expressed in cartilage including those encoding types II, IV, V, and VI collagens, as well as cartilage oligomeric matrix protein (COMP). Recently it was reported that mice deficient in the type IX collagen gene and the matrilin 3 gene (human equivalent condition not yet reported) developed age-dependent OA-like changes in the knee and temporal mandibular joints. Moreover, candidate genes in OA have been identified that are not structural proteins. The haplotype of a vitamin D receptor (VDR) that plays a vital role in controlling bone mineral density appears to be associated with a twofold risk of knee OA—although the VDR locus is close to the COL2A1 locus on chromosome 1q, so the association may be due to linkage disequilibrium with the latter. In addition, the locus of the insulin-like growth factor 1 (IGF-1) gene has been associated with radiographic OA, as has an aggrecan polymorphic allele with hand OA.

In population studies, genome-wide linkage scans have highlighted up to seven chromosomal regions that may harbor OA susceptibility genes. Chromosome 2q was positive in several scans, suggesting that this chromosome is likely to harbor one or more susceptibility genes. In a study of affected sibling pairs, a region of linkage stretching from 2q12 to 2q21 was reported for OA of the distal interphalangeal joint and a previous study of affected sibling pairs in the United Kingdom demonstrated a broader region of linkage, stretching from 2q12 to 2q31. Two IL-1 genes (IL1A and IL1B) and the gene encoding IL-1Ra (IL1RN) are located on chromosome 2q13 within a 430-kb genomic fragment. Given the importance of IL-1 in the perpetuation of cartilage damage in OA, it is possible that a proportion of the genetic susceptibility to OA may be encoded for by variation in the activity of interleukins, and that for chromosome 2q this susceptibility could reside within the IL-1 gene clusters. Loughlin and colleagues however, have provided evidence that the IL-1 gene cluster harbors susceptibility for knee OA, but not for hip OA. These and other epidemiologic studies have highlighted potential differences in the degree of OA heritability among different joint groups and between the two sexes.

Genomic and postgenomic technology, in addition to defining susceptibility genotypes, will also lead to the discovery of genes and gene products that are overexpressed in OA tissues and that contribute to disease pathogenesis and progression. Studies of differential gene expression in diseased tissue, in addition to elucidating pathogenic processes that lead to novel therapies, could also have two additional benefits: (1) identification of unique biomarkers that can be used for OA diagnosis or management and (2) identification of candidate susceptibility genotypes such as polymorphic variations of cytokines or growth factors that may predispose to disease progression.

**Joint Malalignment and Trauma**

Joint malalignment or trauma may lead to rapid development of OA, or it may initiate a slow process that results in symptomatic OA years later. Probably as a result of progressive reduction in periarticular blood flow and the resultant decrease in rate of remodeling at the osteochondral junction, joints become increasingly congruent with age. Altered joint geometry may interfere with nutrition of the cartilage, or it may alter load distribution, either or which may result in altered biochemical composition of the cartilage, irrespective of age. Local factors such as stresses related to joint use and joint deformity also influence the development of OA.

Joint incongruence (e.g., malreduced intra-articular fractures, developmental dysplasia of the hip, recurrent dislocation of the patella) can lead to early-onset OA. Repetitive, high-impact sports are strongly associated with joint injury and increase the risk for lower limb OA. Repetitive trauma at a subfracture level has been shown to accelerate remodeling in the zone of calcified cartilage, with reduplication of the tidemark and thinning of the noncalcified zone, resulting in stiffening of the subchondral bone, increased wear of the overlying cartilage, and ultimately development of OA. Regular exercise is important in maintaining articular cartilage structure and metabolic function. Recreational running and low-impact activities have not been shown to increase the risk of OA.

Articular cartilage is remarkably resistant to damage by shear forces; it is, however, highly vulnerable to repetitive impact loading. When joints are subjected to in vitro cyclic loads that are easily borne by subchondral bone, cartilage degeneration still results. This vulnerability accounts for the high frequency of OA in shoulders and elbows of pneumatic drill operators and baseball pitchers, ankles of ballet dancers, metacarpophalangeal joints of boxers, and knees of basketball players. The risk for knee osteoarthritis among participants in sports, however, may be more closely related to previous knee injury than to participation in sport alone.

The major forces on articular cartilage, in addition to weight bearing, are due to the contraction of the muscles that stabilize or move the joint. For example, in normal walking, 4 to 5 times the body weight may be transmitted through the knee, and in squatting it is up to 10 times the body weight. Articular cartilage is believed to be too thin to be an effective shock absorber under these high loads. What protects the joint under physiologic conditions of impact loading is joint motion, with the associated lengthening of muscles under tension and deformation of the subchondral bone. Cancellous subchondral bone functions as a major shock absorber due to its material properties. Two-thirds of subchondral bone stiffness derives from bony trabeculae, about one-third from intraosseous fluid. In the normal unloaded joint, the opposing surfaces are not congruous; under loading, both the cartilage and the bone deform so that a larger proportion of the opposing surfaces comes into contact, increasing joint congruity and resulting in a force distribution over the largest possible area.
Excessive loads may cause microfractures of subchondral trabeculae that heal via callus formation and remodeling, resulting in stiffer-than-normal bone that is less effective as a shock absorber and predisposing articular cartilage to degeneration.

Whether subchondral sclerosis precedes the onset of OA or is a change secondary to cartilage degeneration is not known. Indirect evidence supports the theory that biomechanical changes in subchondral bone may be important in OA.63,65,66 Foss, for example, reported cases of femoral osteoporosis (which is associated with softening and greater compliance of the subchondral bone) that may have protected the hip from OA.67 Conversely, in vitro studies have shown that stiffening of the cancellous bone with methacrylate, reducing its deformability, leads to cartilage degeneration with repetitive impact loading.68

Gender

Women are about twice as likely as men to develop OA. Although women have a lower prevalence of OA than men before age 50 years, there is a marked increase in prevalence among women after 50, particularly in the knee.69 Radiographic and interview data from the National Health and Nutrition Examination Survey (NHANES III), a representative cross-sectional health examination survey of the U.S. population, reported that the lifetime prevalence of radiographic knee osteoarthritis was 37.4% and that the prevalence of symptomatic knee osteoarthritis was 12.1% in adults age 60 years and older; prevalence was greater among women than men (42.1% vs. 31.2%), and women had significantly more Kellgren-Lawrence grade 3 and 4 changes (12.9% vs. 6.5% in men).70 Women have a greater number of joints involved and are more likely to exhibit clinical symptoms of morning stiffness, joint swelling, and nocturnal pain. The gender differences in OA incidence after age 50 may be the result of postmenopausal estrogen deficiency. Articular chondrocytes possess functional estrogen receptors, suggesting that these cells can be regulated by estrogen. Nuclear estrogen receptors (ERs) have been detected in articular chondrocytes of humans,71,72 rats,73 monkeys,74 pigs,75 and human growth plate chondrocytes.76

Recent epidemiologic studies have linked estrogen replacement therapy (ERT) with a lower-than-expected risk of knee and hip OA in postmenopausal women. Clinical investigations of the association between OA and hormonal level in women have involved measurement of circulating estrogen levels in postmenopausal women, general radiographic evaluation of postmenopausal women, and examination of the effect of ERT on such variables as knee OA and cartilage volume.77-80 In a study of more than 4000 women age 65 years or older that assessed pelvis radiographs for hip OA, Nevitt and colleagues77 showed that women using oral estrogen faced a significantly reduced risk of hip OA. Estrogen users for 10 years or longer had a greater reduction in risk of developing hip OA than users for less than 10 years. Zhang and colleagues,79 using weight-bearing radiographs in female participants in the Framingham Osteoarthritis Study (n = 831; mean age, 73 years; age range, 63 to 93) to study the rate of knee OA, reported a modest but nonsignificant greater protective effect for radiographically detected OA in women who were on ERT. A prospective cohort study using anteroposterior weight-bearing knee radiographs of women (mean age, 71 years; age range, 63 to 91) from the Framingham Study categorized patients into three groups according to estrogen use at biennial examination: never users (n = 349), past users (n = 162), and current users (n = 40). When both incident and progressive radiographic knee OA cases were combined, current ERT users were found to have 60% less risk for knee OA than never users. Wluka and colleagues80 reported on the longer-term use of ERT and its association with knee cartilage volume (measured by magnetic resonance imaging [MRI]) in postmenopausal women; results showed that after adjusting for confounders, women using long-term ERT had more knee cartilage than controls. Beneficial effects of ERT on the severity of knee OA in ovariectomized monkeys have also been demonstrated.81

Changes in Osteoarthritis

Morphologic Changes

In early OA, the articular cartilage surface becomes roughened and irregular, and superficial clefts within the tissue become apparent. Histologically, the surface is fibrillated and small cracks are apparent but limited to the upper layers of the surface zone. These changes show evidence of mechanical wear and are often accompanied by matrix swelling and chondrocyte proliferation or, to a limited extent, apoptosis near the articular surface. Proteoglycan distribution is altered, as revealed by histochemical staining. As OA progresses, the articular cartilage surface becomes more irregular and the superficial clefts within the tissue become enlarged and extend beyond the superficial zone and into the middle zone of cartilage. The previously isolated and focal portions of damaged cartilage become increasingly contiguous. Portions of the superficial zone are completely missing in focal lesions. As the condition worsens, the clefts deepen, surface irregularities increase, and the articular cartilage eventually ulcerates, exposing the underlying bone. As the disease continues to progress, the joint articulates on exposed bone, causing eburnation and thickening of the bone. The eburnated bone becomes denser and more metabolically active in response.

Early Reparative, Proliferative, and Hypertrophic Changes

Attempts at local self-repair can be seen as an initial increase in the number of chondrocytes in the form of clusters or clones, with as many as 50 or more cells in a cluster89 (Figure 98-1). In healthy cartilage, chondrocytes are quiescent and neither commit to proliferation nor to further hypertrophic differentiation. In contrast, in the early stages of OA, proliferating chondrocytes in clusters express higher levels of matrix proteins including aggrecan and type II collagen,83 as well as stem cell markers90 and markers of hypertrophic differentiation.85 The chondrocyte clusters with proliferative and hypertrophic markers are a hallmark of OA and are also prominent in animal models of OA and three-dimensional cell cultures.85 In addition to being a repair response, chondrocyte clusters are thought to contribute to
OA pathogenesis and progression through the release of matrix-degrading enzymes, growth factors, and inflammatory cytokines that affect surrounding chondrocytes and joint tissues.85

Osteophyte Formation

Osteophytes consist of newly formed fibrocartilage and bone and are most often formed at the peripheral margins of joints at the interface between cartilage and the periosteum. Osteophytes are thought to arise through chondrogenic differentiation of progenitor cells, most commonly from within the periosteum.86 As such, osteophytes may be a cellular repair response to the altered growth factor environment after joint injury, and in certain cases osteophytes can contribute to the stability of the joints.87,88 The connection between osteophyte formation and the repair response is further strengthened by animal models of OA in which the early stages of osteophyte formation can be histologically apparent as early as 3 days after joint injury. Many growth factors that enhance in vitro chondrogenic differentiation of stem cells are also present in osteophytes, including members of the transforming growth factor (TGF)-β superfamily,89 bone morphogenetic proteins (BMPs),90 IGF, and fibroblast growth factor,86 and TGF-β can induce osteophyte formation in animal models.91 As OA progresses, the osteophytes can limit movement and become painful.

Hypocellularity

A reduction in cell number is observed in aging cartilage, and the reduced synthetic capability of hypocellular cartilage is a contributing factor to the initiation and progression of OA. In healthy adult femoral cartilage, the cell density varies between 24,000/mm³ in the surface zone and 8000/mm³ in the deep zone, with an average of 1.65% of the cartilage volume consisting of cells.92 This number is greatly reduced in OA, through cell death via necrosis or apoptosis.93 Necrosis can result from direct mechanical damage to cells and is not generally an active process, whereas apoptosis is an active energy-consuming process. Apoptotic chondrocyte death can be initiated by many factors that are involved in the initiation and progression of arthritis including mechanical damage or injury, changes in cell-matrix interactions, oxidative stress resulting from nitric oxide or other reactive oxygen species, impaired mitochondrial function, and signal transduction pathways such as CD95/CD95 ligand. These pathways ultimately converge, and the execution of apoptotic cell death is mediated by the activation of caspases. Inhibition of apoptosis by interfering with caspase activation following injury is being explored as a chondroprotective intervention to prevent secondary OA following injury. Compromised regulation of cellular senescence and autophagy can also cause cell death. The role of autophagy in inflammation and pathogenesis has been gaining the attention of researchers.94 In some circumstances cell death can be mediated by the autophagic degradation of the damaged organelles. In cartilage, autophagy can be protective, and its reduction in OA corresponds with increased apoptotic markers.95

ALTERATIONS IN CARTILAGE MATRIX METABOLISM

Biochemical Changes

These morphologic changes are accompanied by changes in the biochemical composition of the cartilage, which vary from early to later stages in the disease process. In early OA, the water content of the articular cartilage significantly increases, causing the tissue to swell and altering its biomechanical properties. This phenomenon suggests that there has been weakening of the collagen network; the type II collagen fibers have a smaller diameter than normal cartilage, and the normally tight weave in the midzone is slackened and distorted.96,97

In later stages of OA, type I collagen concentration within the extracellular matrix increases and the proteoglycan concentration falls to 50% or less than normal, with less aggregation and shorter glycosaminoglycan side chains.98,102 Keratan sulfate concentration decreases, and the ratio of chondroitin-4-sulfate to chondroitin-6-sulfate increases, reflecting synthesis by chondrocytes of a proteoglycan profile more typical of immature cartilage.103 Proteoglycan concentration in the cartilage diminishes progressively

Figure 98-1  Histologic sections of normal (left) and osteoarthritic (right) articular cartilage obtained from the femoral head. The osteoarthritic cartilage has surface irregularities, with clefts to the radial zone and cloning of chondrocytes.
that is important in the pathogenesis of OA.

Metabolic Changes

Calcium crystals (e.g., calcium pyrophosphate dihydrate, basic calcium phosphate crystals) are commonly found in the cartilage of the elderly, and often crystal arthropathy coexists with OA.127 That calcium crystals play a role in causing or worsening OA is supported by clinical and laboratory studies, but the relationship is complex and it is unclear whether these crystals are directly involved in OA pathogenesis.128-130 Pyrophosphate (PP) is produced from adenosine triphosphate (ATP) by the exoenzyme nucleoside pyrophosphohydrolase.131 High levels of PP in synovial fluid from OA patients correlate directly with the severity of joint damage.132,133 Young or proliferating chondrocytes are a major source of PP, whereas resting chondrocytes from normal adult cartilage secrete little.134 Thus the increased PP secretion in OA cartilage might indicate increased chondrocyte metabolic activity toward matrix repair.135 The presence of CPPD may alter the biomechanical properties of the cartilage extracellular matrix and lead to cartilage breakdown. Hemochromatosis ( hemosiderin), Wilson’s disease (copper), ochronotic arthropathy (homogentisic acid polymers), gouty arthritis (crystals of monosodium urate), and calcium pyrophosphate dihydrate (CPPD) crystal deposition disease are further examples of conditions in which the abnormal entity may alter the cartilage extracellular matrix, leading to either direct or indirect chondrocyte injury by increasing the stiffness of the tissue and thereby precipitating the development of OA.

Anabolic Factors (Transforming Growth Factor-β, Bone Morphogenetic Proteins) and Cartilage Repair

TGF-β is essential for the formation and maintenance of cartilage (reviewed in Reference 120). Interference with TGF-β function in cartilage leads to OA in multiple animal models121 and in human genetic susceptibility to OA.122 TGF-β affects cartilage homeostasis at multiple levels: It enhances stem cell chondrogenesis to increase the pool of cells available for cartilage synthesis, and it increases matrix production in existing chondrocytes. TGF-β also increases synthesis of anticytotoxic compounds such as TIMPs and PAI-1, proteins that inhibit the activation of latent proteinases in cartilage. TGF-β attenuates the cellular response to inflammatory cytokines including IL-1β and TNF-α and TGF-β signaling through the Smad2/3 pathway can inhibit terminal differentiation and hypertrophy in chondrocytes. However, TGF-β activity can also contribute to the pathogenesis and progression of OA. For example, TGF-β signaling induces osteophyte formation in animal models. In aging cells and cells in which the ratio of the TGF-β receptors ALK1 and ALK5 is altered, TGF-β can also have opposing effects such as activating MMP-13 and inducing terminal hypertrophic differentiation in chondrocytes.123

BMPs are structurally related to TGF-β but generally activate a different set of receptors and intracellular signaling molecules. BMPs influence all stages of embryonic chondrogenesis and can promote chondrogenic differentiation of adult MSCs.124 Recent genetic evidence suggests that impaired BMP signaling affects OA susceptibility, with the most progress reported for BMP-1 (GDF-5).125,126 There is good evidence that a reduction of BMP-7 (OP-1) is evident in osteoarthritic cartilage,127 with possible regulation occurring through both inhibitory microRNA128 and promotor methylation.129 Supplementation of BMP-7 in joints reduces experimental arthritis in animals and was proven safe in a phase I clinical trial, with no findings of ectopic bone formation in the enrolled patients during the study.130 However, as described for TGF-β earlier, BMP signaling through certain receptors can also enhance terminal differentiation and hypertrophy in chondrocytes, processes that are hallmarks of OA progression.

Anabolic Factors (Transforming Growth Factor-β, Bone Morphogenetic Proteins) and Cartilage Repair

TGF-β is essential for the formation and maintenance of cartilage (reviewed in Reference 120). Interference with TGF-β function in cartilage leads to OA in multiple animal models121 and in human genetic susceptibility to OA.122 TGF-β affects cartilage homeostasis at multiple levels: It enhances stem cell chondrogenesis to increase the pool of cells available for cartilage synthesis, and it increases matrix production in existing chondrocytes. TGF-β also increases synthesis of anticytotoxic factors such as TIMPs and PAI-1, proteins that inhibit the activation of latent proteinases in cartilage. TGF-β attenuates the cellular response to inflammatory cytokines including IL-1β and TNF-α and TGF-β signaling through the Smad2/3 pathway can inhibit terminal differentiation and hypertrophy in chondrocytes. However, TGF-β activity can also contribute to the pathogenesis and progression of OA. For example, TGF-β signaling induces osteophyte formation in animal models. In aging cells and cells in which the ratio of the TGF-β receptors ALK1 and ALK5 is altered, TGF-β can also have opposing effects such as activating MMP-13 and inducing terminal hypertrophic differentiation in chondrocytes.123

BMPs are structurally related to TGF-β but generally activate a different set of receptors and intracellular signaling molecules. BMPs influence all stages of embryonic chondrogenesis and can promote chondrogenic differentiation of adult MSCs.124 Recent genetic evidence suggests that impaired BMP signaling affects OA susceptibility, with the most progress reported for BMP-1 (GDF-5).125,126 There is good evidence that a reduction of BMP-7 (OP-1) is evident in osteoarthritic cartilage,127 with possible regulation occurring through both inhibitory microRNA128 and promotor methylation.129 Supplementation of BMP-7 in joints reduces experimental arthritis in animals and was proven safe in a phase I clinical trial, with no findings of ectopic bone formation in the enrolled patients during the study.130 However, as described for TGF-β earlier, BMP signaling through certain receptors can also enhance terminal differentiation and hypertrophy in chondrocytes, processes that are hallmarks of OA progression.

Metabolic Changes

Early OA is characterized by increased synthesis of proteoglycans, collagen, noncollagenous proteins, hyaluronate, and cell replication.134,135 This “activation” of chondrocytes is thought to be an attempt to repair the cartilage matrix, although it is not always effective and yields a matrix of inferior quality that is more susceptible to degradation.136 However, both anabolic and catabolic processes increase as cells attempt to repair or maintain tissue integrity138 and it is the imbalance between synthesis and degradation that is important in the pathogenesis of OA.117 In later stages of OA, there is a decrease in the synthesis of matrix per cell and a decrease in cell number. Furthermore, the quality of the synthesized matrix is reduced with respect to the composition and distribution of their glycosaminoglycans, the size of the proteoglycan subunit, and their ability to aggregate with hyaluronic acid.139,140,141,142,143 In addition to the reduced matrix production and hypocellularity, there is increased synthesis and activation of matrix degrading enzymes, as well as an overall decrease in the concentrations of enzyme inhibitors such as tissue inhibitors of metalloproteinases (TIMPs). Eventually OA progresses when the chondrocytic anabolic repair processes cannot keep pace with catabolic processes.98,104 The complex interaction between matrix synthesis and degradation explains why OA typically is slowly progressive and at times remains static by morphologic criteria, but ultimately it results in the overall degradation of cartilage matrix. The following sections summarize the major anabolic and catabolic factors relevant to cartilage degradation and the pathogenesis of OA.

**Classes of Proteinases (Metalloproteinases, Aggrecanases, Serine and Cysteine Proteases)**

In OA, the synthesis and secretion of matrix-degrading enzymes by chondrocytes markedly increases. There are four classes of proteases, which are grouped by the catalytic mechanisms of peptide bond cleavage: metalloproteinases, cysteine proteinases, serine proteinases, and aspartyl proteinases. Of these, the first three have clearly defined roles in the degradation of cartilage during the progression of OA.

**Metalloproteinases**

Metalloproteinases have an enzymatic site that requires a metal ion (often zinc) for activity. Cartilage contains two families of metalloproteinases, namely ADAMTSs and MMPs. In OA, the cytokines IL-1 and TNF stimulate the synthesis and secretion of many proteases and MMPs (Figure 98-2). IL-1 is synthesized by mononuclear cells (including synovial lining cells) in the inflamed joint and by chondrocytes as an autocrine activity. The enzymes stimulated by IL-1 and TNF include latent collagenase, latent stromelysin, latent gelatinase, aggrecanase, and tissue plasminogen activator (TPA). Plasminogen is either synthesized by the chondrocyte or enters the matrix by diffusion from the synovial fluid. TPA converts plasminogen to plasmin, a serine protease that can activate latent cartilage-degrading enzymes. A downstream mediator of IL-1- and TNF-induced cartilage degradation that is gaining increasing attention is hypoxia-inducible factor 2α (Hif2α). Hif2α is a transcription factor strongly upregulated in OA cartilage and mouse models of OA. It directly induces the expression of many cartilage-degrading enzymes including MMPs 1, 3, 9, and 12, as well as a disintegrin and metalloproteinase with thrombospondin motif (ADAMTS-4) and (indirectly) ADAMTS-5. In addition, a high level of Hif2α decreases the protective role of autophagy, increases the extent of cell death, and activates the RUNX2 and IHH pathways that further contribute to cartilage matrix degradation. The different classes of proteinases that are activated by these cytokines in OA are discussed in more detail later.

**Catabolic Factors and Cartilage Degradation**

In addition to new matrix synthesis, cartilage remodeling also involves a degree of proteolysis. This occurs via the induction of an array of proteases, particularly matrix metalloproteinases (MMPs). In OA, the cytokines IL-1 and TNF stimulate the synthesis and secretion of many proteases and MMPs (Figure 98-2). IL-1 is synthesized by mononuclear cells (including synovial lining cells) in the inflamed joint and by chondrocytes as an autocrine activity. The enzymes stimulated by IL-1 and TNF include latent collagenase, latent stromelysin, latent gelatinase, aggrecanase, and tissue plasminogen activator (TPA). Plasminogen is either synthesized by the chondrocyte or enters the matrix by diffusion from the synovial fluid. TPA converts plasminogen to plasmin, a serine protease that can activate latent cartilage-degrading enzymes. A downstream mediator of IL-1- and TNF-induced cartilage degradation that is gaining increasing attention is hypoxia-inducible factor 2α (Hif2α). Hif2α is a transcription factor strongly upregulated in OA cartilage and mouse models of OA. It directly induces the expression of many cartilage-degrading enzymes including MMPs 1, 3, 9, and 12, as well as a disintegrin and metalloproteinase with thrombospondin motif (ADAMTS-4) and (indirectly) ADAMTS-5. In addition, a high level of Hif2α decreases the protective role of autophagy, increases the extent of cell death, and activates the RUNX2 and IHH pathways that further contribute to cartilage matrix degradation. The different classes of proteinases that are activated by these cytokines in OA are discussed in more detail later.
factor signaling. Once transcribed, the transcript stability and translation of several metalloproteinases are regulated by microRNA (miR). For example, microRNA-27b (miR-27b) regulates MMP-13 expression. Also, miR-140 levels are decreased in OA, which reduces its repression of ADAMTS-5. Several of these miRNAs are controlled by the same cytokines and growth factors that maintain cartilage homeostasis including IL-1 and TGF-β. Although the miR field is still in its infancy, it is already evident that several miRs are involved in the pathogenesis of OA by affecting transcript stability and protein translation. Once translated into proteins, almost all MMPs are expressed as inactive proenzymes (zymogens) that require further processing for full proteolytic activity. Most MMPs contain an N-terminal prodomain that blocks or otherwise inhibits the catalytic site. The primary MMP activators are serine- and cysteine-dependent proteases (e.g., the plasminogen/plasmin cascade or furin-like proprotein convertases and cathepsin B, respectively), as well as membrane-type MMPs. Activated MMPs can then be inactivated nonspecifically by α2 macroglobulin and with more specificity by the tissue inhibitor of metalloproteinase (TIMP) family of proteins.

MMPs have historically been further divided into three groups on the basis of their substrate specificities; collagenases cleave across all three chains of the native triple-helical collagens, whereas gelatinases preferably cleave denatured collagen and stromelysins have much broader substrate specificities. However, there is considerable overlap in substrates between these classifications; for example, MMP-1 (interstitial collagenase), MMP-3 (stromelysin 1), and MMP-13 (collagenase-3) are all capable of cleaving aggrecan core protein. MMPs can degrade other cartilage extracellular matrix molecules in addition to collagen. If combined with plasmin, which has the capability of activating many MMPs, MMPs can rapidly destroy cartilage altogether.

Collagenases

Collagenases typically make the first cleavage in triple-helical collagen, allowing its further degradation by other proteases. Degradation of collagens is thought to be the first irreversible step in the pathogenesis of OA because it significantly reduces the mechanical properties of cartilage. The best-studied MMPs capable of cleaving native collagens are MMP-13 and MMP-1, although MMP-8 and MMP-28 are also involved in breaking down type II collagen. Of the MMPs that degrade native collagen, MMP-13 may be the most important in OA because it preferentially degrades type II collagen. Mice deficient in MMP-13 resist developing OA and have reduced cartilage erosion even after proteoglycan loss. Expression of MMP-13 greatly increases in OA; overall collagenase activity also markedly increases in human osteoarthritic cartilage cultures, suggesting that it is a major factor in OA progression and cartilage matrix degradation. The resultant collagen fragments may then be susceptible to further cleavage by such other enzymes as MMP-2 (gelatinase A), MMP-9 (gelatinase B), MMP-3, and cathepsin B (a cysteine proteinase).

Aggrecanases

The aggrecanases are metalloproteases that belong to a family of extracellular proteases known as ADAMTS. Two aggrecanases, ADAMTS-4 and ADAMTS-5, appear to be major enzymes in cartilage degredation in arthritis. The roles of these enzymes appear reversed in human and mouse OA—whereas ADAMTS-4 is predominantly associated with aggrecan degradation in human OA, in mouse models of OA ADAMTS-5 is more important. Reombinant ADAMTS-4 and ADAMTS-5 cleave aggrecan at five distinct sites along the core protein, and all resultant fragments have been identified in cartilage explants undergoing matrix degradation. The G1 region of aggrecan is highly resistant to proteases; however, a glutamate-alanine bond within the extended region between G1 and G2 is remarkably susceptible to proteolytic degradation. Aggrecanase proteolytic activity can be modulated by altered expression, by activation by proteolytic cleavage at a furin-sensitive site, by binding to the aggrecan substrate through the C-terminal thrombospondin motif, by activation through post-translational processing of a portion of the C-terminus, and by inhibition of activity by the endogenous inhibitor, tissue inhibitor of metalloproteinase 3 (TIMP-3). ADAMTS-4 and ADAMTS-5 activity is also detected in joint capsule and synovium and may be upregulated in arthritic synovium at either the message level or through posttranslational processing. In addition to ADAMTS-4 and ADAMTS-5, several MMPs are also capable of cleaving aggrecan in vitro (MMP-1, MMP-2, MMP-3, MMP-7, MMP-8, MMP-9, MMP-13, and MMP-28), as are ADAMTS-1, ADAMTS-8, ADAMTS-9, and ADAMTS-15. It has been shown that ADAMTS-7 and ADAMTS-12 both bind to and degrade COMP (a prominent noncollagenous protein in cartilage) and that the latter is highly upregulated in osteoarthritic cartilage. ADAMTS-4, ADAMTS-19, and ADAMTS-20 also have been shown to degrade COMP in vitro; however, their in vivo activity in osteoarthritis has yet to be determined.

A specific hyaluronidase has not been found in articular cartilage matrix, but one or several lysosomal enzymes that can cleave both hyaluronic acid and chondroitin-6-sulfate have been implicated. There are six or seven potential hyaluronidases in the human genome, and of these, hyaluronidase-1, hyaluronidase-2, hyaluronidase-3, and PH-20 are likely to be active in cartilage. However, experimental evidence to date suggests that these enzymes are either exclusively lysosomal or can exist both in the lysosome and at the plasma membrane. Thus although there is convincing evidence of extracellular degradation of hyaluronan, the precise role of the hyaluronidases in this process has not yet been clearly identified. The decrease in chondroitin sulfate chain length in osteoarthritic cartilage may be due to digestion by synovial fluid hyaluronidase, which may diffuse into the matrix as its permeability increases. Evidence to support this theory is the finding that the hyaluronic acid concentration in OA cartilage is low, even though its rate of synthesis is considerably greater than normal. These degradative enzymes serve to disrupt the proteoglycan aggregate. The early result of the MMP-induced tissue degradation is thinning of the collagen fibers,
Loosening of the tight collagen network, and the consequent cartilage matrix swelling seen in OA.

**Enzyme Inhibitors (Tissue Inhibitor of Metalloproteinases, Plasminogen Activator Inhibitor-1)**

The balance of active and latent enzymes is controlled to some extent by at least two enzyme inhibitors: tissue inhibitor of metalloproteinases (TIMP) and plasminogen activator inhibitor-1 (PAI-1). TIMP and PAI-1 are synthesized in increased amounts under the regulation of transforming growth factor-β. If insufficient concentrations of or degraded TIMP or PAI-1 are present in the matrix along with active enzymes, then increased matrix degradation occurs. Expression profiling of all known members of the MMP, ADAMTS, and tissue inhibitor of metalloproteinases (TIMP) gene families in normal cartilage and cartilage from patients with OA has revealed that several members are regulated in OA. Genes that showed increased expression in OA were MMP-2, MMP-9, MMP-12, MMP-13, MMP-16, MMP-28, ADAMTS-2, ADAMTS-12, ADAMTS-14, ADAMTS-16, and TIMP-3 (all at P < .05). Genes with decreased expression in OA were MMP-1, MMP-3, MMP-10, ADAMTS-1, ADAMTS-5, ADAMTS-9, and ADAMTS-15; and TIMP-1 and TIMP-4. These results illustrate the complexity of the events that occur within the extracellular matrix regarding regulation of tissue-degrading enzymes.

**Alterations in Matrix Synthesis**

Much of what is known about changes in the extracellular matrix in early OA comes from animal models (e.g., the rabbit partial meniscectomy model, the canine anterior cruciate ligament–deficient model). These animal models represent secondary OA, produced by internal derangement of the joint, and therefore may not precisely simulate the state of affairs in primary OA. In a recent review, models for spontaneous osteoarthritis were reviewed and their preclinical utility was compared. Of particular interest is an iodoacetate model, validated as the first primary model of osteoarthritis, in which intra-articular injection of iodoacetate in rats has been shown to lead to cartilage degeneration associated with pain and manifesting as time- and concentration-dependent alterations in hindlimb weight bearing. That not all animal models of osteoarthritis are equivalent becomes clear when one considers the differences in therapeutic response between young and old animals and between spontaneous and surgical models.

In the dog model, the first alteration seen within days after joint destabilization is an increase in cartilage water content. Initially, water content increases locally, in the tibial plateau and femoral condyle cartilage, but it soon spreads to the entire joint cartilage. Proteoglycans are more readily extractable from the matrix of experimental animals than from that of controls, and this is reflected in the serum biomarker profile after acute injury in patients. These matrix changes are also seen in spontaneously occurring dog and steer OA and in experimentally induced rabbit OA. The increase in water content in osteoarthritic cartilage is due to loss of the collagen network’s elastic restraint, enabling the hydrophilic proteoglycans to swell more than normally. In early-stage OA, proteoglycan concentration may increase and the cartilage may consequently become thicker than normal and exhibit increased staining for proteoglycans. Shortly after the increase in cartilage water, newly synthesized proteoglycans are characterized by a higher proportion of chondroitin sulfate and a lower proportion of keratan sulfate and proteoglycan aggregation is impaired. These abnormal changes in extracellular matrix occur before fibrillation or any other gross morphologic changes are evident and result in a generalized decrease in stiffness that occurs in grossly normal cartilage adjacent to fibrillated areas. As OA progresses, focal cartilage ulcerations develop. Proteoglycan loss is accompanied by a decrease in its ability to aggregate, persistence in abnormal glycosaminoglycan composition, and a decrease in chondroitin sulfate chain length. Once proteoglycan loss reaches a critical threshold, water content, which initially increased, falls below normal.

**Chondrocyte Senescence**

The term senescence was originally coined to describe the replicative limit of primary cells cultured in vitro, and the concept has evolved to include premature senescence or senescence-like changes in differentiated postmitotic cells. Senescent cells have a unique phenotype that affects both the proteins expressed by senescent cells and their responses to extracellular stimuli. Furthermore, the impact of senescence is not limited to the cells undergoing senescence; rather, the senescent cells have an impact on surrounding cells through the paracrine effects of the altered protein secretions.

Premature chondrocyte senescence in OA is thought to be enhanced by oxidative injury. The inflammatory cytokines IL-1β and TNF that are active in OA induce oxidative stress in chondrocytes, thereby contributing to the potential for oxidative injury that leads to senescence. Oxidative injury is also induced with excessive mechanical shear and injurious loading of explants, and this may contribute to the accumulation of senescent chondrocytes in aged cartilage. TGF-β can influence both growth arrest and senescence in cultured cells, and senescent cells can overexpress TGF-β, but this effect has mostly been demonstrated in endothelial cells and not cells of mesenchymal origin. In summary, although it is clearly evident that many of the cellular mechanisms controlling senescence are also active in disease progression including OA, a causal link between senescence and OA has yet to be identified.

**BIOMECHANICS AND DISEASE MECHANISMS OF OSTEOARTHRITIS**

**Biomechanical Changes**

Two long-standing biomechanical theories of the pathogenesis of OA hold that mechanical stresses injure chondrocytes, causing them to release degradative enzymes, and that mechanical stresses initially damage the collagen network (as opposed to the cells per se). In either case, the result is that matrix breaks down. Extracellular matrix breakdown in osteoarthritic cartilage leads to (1) loss of
compressive stiffness and elasticity, resulting in greater mechanical stress on chondrocytes; and (2) an increase in hydraulic permeability, resulting in loss of interstitial fluid during compression and increased diffusion of solutes through the matrix (including the movement of degradative enzymes and their inhibitors). One important consequence is disruption of normal fluid film joint lubrication and loading dynamics due to alterations in inflammatory synovial fluid.\textsuperscript{195-197} Joint friction, wear, lubrication, and contact mechanics are further negatively affected by the loss of cartilage proteoglycans and superficial zone protein (also called lubricin).\textsuperscript{196-202}

**Response of Cartilage to Mechanical Injury**

The response of normal articular cartilage to injury typically results in suboptimal repair; these injuries can often result in secondary OA.\textsuperscript{203,204} Articular cartilage produces a repair tissue with neither the original structure nor properties of normal cartilage.\textsuperscript{205-208} Chondrocytes in areas surrounding an injured zone are unable to migrate, proliferate, repopulate, or regenerate repair tissue with similar structure, function, and biomechanical properties of normal hyaline cartilage.

That articular cartilage lacks regenerative power has a long history of documentation.\textsuperscript{210} In 1851 Redfern reported that articular cartilage wounds healed with fibrous tissue, which he believed arose from chondrocyte intercellular substance.\textsuperscript{211} Fisher and Ito in the 1920s proposed that cartilage repair is affected by fibrous tissue resulting from proliferation of cells from bone marrow, synovial membrane, and occasionally surrounding articular cartilage.\textsuperscript{212,213} It was later observed that the fibrous tissue regenerate subsequently transforms into fibrocartilage, with occasional foci of imperfect hyaline cartilage.\textsuperscript{214-217} The common findings of these investigators were that articular cartilage lacks regenerative potential and that the regenerative fibrous tissue and fibrocartilage tissue must have originated from undifferentiated mesenchymal tissue arising from bone marrow, synovium, or the superficial layer of articular cartilage.\textsuperscript{210}

One reason the reparative process of cartilage significantly differs from those of other tissues is that it is avascular. The healing response in vascularized tissues consists of three main phases: necrosis, inflammation, and repair.\textsuperscript{203,208} Cartilage undergoes the initial phase of necrosis in response to injury, although typically less cell death occurs than in vascularized tissues because of chondrocytes’ relative insensitivity to hypoxia.\textsuperscript{203,208} The inflammatory phase, primarily mediated (in other tissues) by the vascular system, is largely absent in partial-thickness injuries (i.e., lesions that do not cross the tidemark), and the repair phase is severely limited given the lack of vascularity and a preceding inflammatory response. Thus no local hyperemia results, no fibrin network is given the lack of vascularity and a preceding inflammatory response, and no inflammatory cells, which have mitotic activity and replication, adult chondrocytes have little potential for replication and intrinsic repair.\textsuperscript{216,217} Articular cartilage lesions that cross the tidemark may undergo extrinsic repair via differentiation and proliferation of mesenchymal stem cells from para-articular connective tissues; typically, however, a fibrocartilaginous regenerate results.\textsuperscript{218}

There are three categories of articular cartilage injury: (1) microdamage or repetitive trauma to the matrix and cells; (2) partial-thickness or superficial injuries or chondral fractures, articular surface injuries that do not penetrate the subchondral plate; and (3) osteochondral (full-thickness or deep penetrating) injuries, which extend through the tidemark and into the underlying subchondral bone.\textsuperscript{203,205,208} The host response to each type of injury differs in both timing and quality of repair.

Microdamage to the chondrocytes and/or extracellular matrix without gross disruption of the articular surface can be caused by a single severe impact or repetitive blunt trauma.\textsuperscript{203,205,208} Repetitive loading of rabbit cartilage produces a surface loss of proteoglycans and an increase in chondrocyte metabolic activity.\textsuperscript{219} Proteoglycans become more easily extractable from the articular cartilage, with a greater percentage of nonaggregated forms.\textsuperscript{189} The cellular, metabolic, and biochemical changes after repetitive blunt trauma resemble those in the early stages of OA: increased hydration; cellular degeneration and/or death; disruption of the collagen ultrastructure resulting in marked variation in the size and arrangements of fibers, fissuring and ulceration of the articular surface, thickening of the subchondral bone, and softening of the cartilage with loss of its compressive and tensile stiffness.\textsuperscript{189,205,210-213} Trauma induces the release of degradative enzymes and proinflammatory factors (e.g., nitric oxide, TNF, IL-1) that frequently cause degradation of the surrounding matrix.\textsuperscript{219,214,215} Eventually the material properties of the cartilage are altered—cartilage matrix thins and subchondral bone stiffens—which in turn often accelerate the degenerative process.\textsuperscript{189} The point at which accumulated microdamage becomes irreversible is unknown, although it has been demonstrated that lost proteoglycans and matrix components may be restored if damage to chondrocytes and the collagen network is limited and the repetitive trauma halted.\textsuperscript{205}

Necrosis of neighboring chondrocytes follows chondral fractures and superficial lacerations, injuries that do not cross the tidemark.\textsuperscript{203,208,226} Within 48 to 72 hours, surviving chondrocytes bordering the defect exhibit increased synthesis rates of extracellular matrix molecules and type II collagen, sometimes accompanied by cell proliferation and formation of clusters or clones in the periphery of the injured zone.\textsuperscript{189,208,210,217} The increased metabolism and mitotic activity is transient, however, and is followed by a fall in metabolic rate back to normal levels, typically resulting in a suboptimal repair.\textsuperscript{189,208} Chondrocytes proliferating on the border of the injured zone do not migrate into the defect, which remains unfilled by the newly synthesized matrix.\textsuperscript{189,205,229} In some cases, superficial lacerations in otherwise normal joints may not progress to full-thickness loss of cartilage or OA.\textsuperscript{189}

Lesions that cross the articular cartilage tidemark and disrupt the underlying subchondral plate elicit the three-phase repair response normally encountered in vascularized tissues. A hematoma forms in the defect that becomes organized into a fibrin clot, activating an inflammatory response.
Mechanotransduction and Gene Expression

Chondrocytes can sense and respond to mechanical stimuli via several regulatory pathways (e.g., upstream signaling, transcription, translation, post-translational modification, vesicular transport). Chondrocytes can remodel extracellular matrix in response to alterations in functional demand, as physical forces influence the synthesis, assembly, and degradation of the extracellular cartilage matrix. High magnitude or duration loads can also cause chondrocyte death and collagen damage, and chondrocytes in the superficial zone appear to be more vulnerable to load-induced injury than those in the middle and deep zones. Normal stimuli help chondrocytes maintain the extracellular matrix; abnormal stimuli can disrupt this balance.

Mechanotransduction influences the cell-mediated feedback between physical stimuli, the molecular structure of newly synthesized matrix molecules, and the resulting biomechanical tissue properties. Cell-matrix interactions via integrins are believed to be one of the important mediators in mechanotransduction in chondrocytes. In a study on the expression of COMP to long-term cyclic compression, it was found that uniaxial unconfined dynamic compression significantly upregulated COMP expression, which could be blocked by incubation with anti-α5β1 integrin antibodies. Studies of bovine articular cartilage explants showed that cyclic loading increased protein synthesis by as much as 50% above free-swelling controls and had an inhibitory influence on proteoglycan synthesis, whereas static compression reduced biosynthetic activity. Fibronectin and COMP were the most affected noncollagenous extracellular proteins; static compression caused a significant increase in fibronectin synthesis versus free-swelling control levels, and cyclic compression caused a significant increase in synthesis of COMP and fibronectin.

Human articular chondrocytes use the α5β1 integrin as a mechanoreceptor. Mechanical stimulation initiates a signal cascade involving stretch-activated ion channels, the actin cytoskeleton, and focal adhesion complex molecules. The result is an anabolic response, manifested by increased aggrecan and decreased MMP-3 expression. Mechanical stimulation also activates Rho and Rho kinase pathways, which are linked to changes in the actin cytoskeleton. Stimulation of the Rho/ROCK pathway in this context is an anabolic stimulus that leads to nuclear translocation and activation of Sox9, which is a "master regulator" of cartilage gene expression. Indian hedgehog (Ihh) protein is a key signaling molecule that controls chondrocyte proliferation and differentiation and may also be an essential mediator of mechanotransduction in cartilage. Cyclic mechanical stress was shown to induce Ihh expression by chondrocytes.

Dysregulation of these anabolic pathways is thought to contribute to the progression of OA. For example, although the α5β1 integrin is also mechanosensitive in OA chondrocytes, the downstream signaling pathways differ, leading instead to increased cartilage breakdown. Integrins and integrin-associated signaling pathways are at least partly regulated by mechanical stimulation by activation of plasma membrane aminoplasmin-sensitive Ca2+-activated K+ channels; the result is membrane hyperpolarization after cyclic mechanical stimulation. Chondrocytes from normal articular cartilage exhibit membrane hyperpolarization response to cyclic pressure-induced strain, whereas chondrocytes from osteoarthritis cartilage respond by membrane depolarization and exhibit no changes in aggrecan or MMP-3 messenger RNA following mechanical stimulation. The different signaling pathways responding to mechanical stimulation in healthy versus OA chondrocytes may affect the disease outcome.

In addition to cell and matrix deformation, fluid flow is also sensed by chondrocytes. A study using a tissue shear-loading model to uncouple fluid flow from cell and matrix deformation demonstrated that deformation of cells and pericellular matrix alone stimulated protein and proteoglycan synthesis.

ABNORMALITIES OF BONE

Osteophyte Formation

Osteophytes—bony proliferations at the joint margins and in the floor of cartilage lesions—are in part responsible for the pain and restriction of joint movement in OA. Human osteoarthritic joint osteophytes synthesize cartilage with significant amounts of type I collagen and nonaggregating proteoglycans. In experimentally induced OA, osteophytes may develop even though the articular cartilage appears grossly normal.
from penetration of blood vessels into the basal layers of degenerating cartilage or as a result of abnormal healing of stress fractures in the subchondral trabeculae near the joint margins.\textsuperscript{252,253} In the OA dog model, periarticular osteophyte formation begins in the marginal zone, where synovium merges with peristeum and articular cartilage, as early as 3 days after induction of knee instability.\textsuperscript{254} Bony proliferation may result from venous congestion; in human hip OA, phlebography has demonstrated the formation of medullary varices, presumably due to changes in the medullary sinuses, which may be compressed by subchondral cysts and thickened subchondral trabeculae.\textsuperscript{255,256} Subchondral cysts in OA may be created by entry of synovial fluid under pressure through defects in the cartilage or may develop in necrotic areas of subchondral bone.\textsuperscript{257} The increased venous pressure caused by the cysts and remodeled trabeculae may account for some of the pain in OA. Immobilization and glucocorticoids (but not bisphosphonates) have been shown to decrease the size and prevalence of osteophytes in experimental models of OA.\textsuperscript{258-261}

Subchondral Bone Sclerosis

Increased remodeling and hardening of subchondral bone is evident early in osteoarthritis and can sometimes be detected before loss of cartilage thickness is evident radiologically.\textsuperscript{262} An advancing tidemark of calcified cartilage is observed in OA, which changes the mechanical interface between subchondral bone and cartilage and is associated with increased subchondral vascularity. The increased calcification results in a thinning cartilage layer, which increases the mechanical stress in the adjacent cartilage. In combination with subchondral bone sclerosis, the altered mechanical environment and rapid bone remodeling may be contributing factors in cartilage degradation and OA pathogenesis.

Bone Marrow Lesions

Bone marrow lesions (detectable by MRI) are associated with OA, and studies suggest that they contribute to the pain felt by OA patients.\textsuperscript{263} The presence of bone marrow lesions is predictive of OA progression, the development of cartilage defects and degradation, and the need for joint replacement.\textsuperscript{264} The relationship between bone marrow lesions and cartilage degeneration in OA is still under investigation. For example, a longitudinal study also found that the presence of cartilage defects predicts progression of bone marrow lesions. The study authors suggest that it “remains unclear whether bone marrow lesions precede, accompany, or follow cartilage damage and volume loss in OA.”\textsuperscript{264}

Fibrosis and necrosis of the underlying bone marrow are common histologic features of the lesions visible by MRI, as are high bone turnover, altered trabecular structure, and edema. Unlike cartilage loss, bone marrow lesions are not a permanent structural change in OA. Several reports demonstrate that these lesions can resolve or at least regress, although it is more commonly observed that bone marrow lesion scores increase over time.\textsuperscript{265} It has been postulated that edema-like lesions are less severe and reversible, whereas more advanced fibrotic and necrotic lesions are not.\textsuperscript{266} There is a possibility that synovial fluid entering the subchondral bone marrow through intra-articular defects leads to an altered growth factor and cytokine environment that affects bone turnover. These findings suggest that there may be an altered biomechanical property of the subchondral tissues, which in turn would affect the biomechanical stresses experienced by the adjacent cartilage.

ROLE OF INFLAMMATORY MEDIATORS IN DISEASE PROGRESSION

Although levels of inflammatory cytokines such as IL-1, IL-6, and TNF are elevated in the serum of patients with knee osteoarthritis,\textsuperscript{267} these and other classic inflammatory mediators are activated within the joint tissues themselves during the course of OA. These cytokines autacatlytically stimulate their own production and induce chondrocytes to produce proteases, chemokines, nitric oxide, and eicosanoids such as prostaglandins and leukotrienes. The action of these inflammatory mediators within cartilage is predominantly to drive catabolic pathways, inhibit matrix synthesis, and promote cellular apoptosis. Thus although osteoarthritis is not conventionally considered an inflammatory disease, “inflammatory” mediators from the affected tissues perpetuate disease progression and therefore represent potential targets for disease modification.

Inflammatory Molecules Produced by Articular Cartilage

Cytokines and Chemokines

A characteristic feature of established osteoarthritis is the increased production of proinflammatory cytokines such as IL-1β and TNF by articular chondrocytes. Both IL-1β and TNF exert comparable catabolic effects on chondrocyte metabolism, decreasing proteoglycan collagen synthesis and increasing aggregan release via the induction of degradative proteases.\textsuperscript{142,268-273} IL-1β and TNF also induce chondrocytes and synovial cells to produce other inflammatory mediators such as IL-8, IL-6, nitric oxide, and prostaglandin E\textsubscript{2}. The actions of both cytokines are in part mediated by the activation of the transcription factor nuclear factor κB (NFκB), which further increases their own expression and that of other catabolic proteins such as inducible nitric oxide synthase and COX-2, thus creating an autacatlytic cascade that promotes self-destruction of articular cartilage\textsuperscript{274-276} (Figure 98-3). Mounting evidence suggests that NFκB mediates chondrocytes’ proinflammatory stress responses in OA and also controls these cells’ differentiation, making NFκB activating factors attractive OA therapeutic targets for many reasons.\textsuperscript{277}

IL-1β and TNF are both synthesized intracellularly as precursors, converted through proteolytic cleavage to their mature forms by caspasess—membrane-bound IL-1β—converting enzyme (ICE) and TNF-converting enzyme (TACE)—and released extracellularly in their active forms.\textsuperscript{278} The expression of both ICE and TACE has been shown to be upregulated in OA cartilage.\textsuperscript{251-253} Inhibitors of both ICE and TACE are of interest as future therapeutic small-molecule antagonists of downstream IL-1β and TNF expression, respectively.
The actions of IL-1 are dependent on the engagement of two specific cell surface receptors (IL-1Rs), designated type I and type II. The type I receptor, which spans the plasma membrane, is responsible for signaling transduction, whereas the type II receptor is a “decoy” receptor, expressed at the cell membrane but unable to signal. A relative deficit of the ratio of IL-1Ra (the competitive inhibitor to the IL-1/IL-1R complex) to IL-1 has been described in OA synovial tissue, which may permit increased IL-1 activity. Addition of IL-1Ra, or soluble type I and II IL-1 receptors, to OA explant cultures blocks prostaglandin E2 (PGE2) synthesis, collagenase production, and nitric oxide (NO) production; addition of these antagonists in culture also results in an increase in aggrecan content, likely by inhibiting degradation of newly synthesized molecules. One recent study showed that chondrocytes with abnormal morphology, from macroscopically intact cartilage representative of early OA, demonstrated increased levels of cell-associated IL-1 and loss of type VI collagen.

Encouraging results with IL-1Ra have also been reported in vivo, where gene therapy or the intra-articular administration of IL-1Ra has been shown to retard the progression of OA in experimental animal models. Other lines of evidence also point to IL-1β as an essential link in the pathogenesis of cartilage damage including proteoglycan loss by intra-articular injection of IL-1. Clinical trials using IL-1β antagonists in OA are few and have been inconclusive. One multicenter trial with double-blinded doses of intra-articular IL-1Ra to 14 OA patients resulted in decreased pain without significant adverse events or acute injection reactions. No long-term studies of structure-modifying effects of such therapies have been reported, however.

Osteoarthritic cartilage is also the site of increased production of both CXC and CC chemokines. These include IL-8, monocyte chemoattractant protein-1 (MCP-1), and RANTES (regulated on activation, normal T-cell expressed and secreted; aka CCL5) chemokine, as well as the receptors CCR-2 and CCR-5. The expression of chemokines is low or undetectable in normal chondrocytes unless stimulated with cytokines such as IL-1 or IL-17. Chemokines are detected by immunohistochomistry in the superficial and midzones of the tissue, as has been demonstrated for such other inflammatory mediators as inducible nitric oxide synthase (iNOS), IL-1β, and TNF. RANTES induces expression of its own receptor, CCR-5, suggesting an autocrine/paracrine pathway of the chemokine within the cartilage. MCP-1 and RANTES promote chondrocyte catabolic activities including induction of nitric oxide synthase, increased MMP-3 expression, inhibition of proteoglycan synthesis, and enhancement of proteoglycan release. Consistent with these effects, treatment of normal articular cartilage with RANTES increases the release of glycosaminoglycans and profoundly reduces the intensity of safranin O staining.

### Proteinases

A presumed key action of cytokines and chemokines produced in osteoarthritis is to promote cartilage proteolysis via the induction of a wide array of proteases, in particular MMPs. The two main families of MMP enzymes are (1) the collagensases that break down type II collagen (especially MMP-1, MMP-8, MMP-13, and MMP-28) and proteoglycans (MMP-3, which also cleaves pro-MMPs into their active forms) and (2) the aggrecanase (ADAMTS) family, which mediate aggrecan degradation in cartilage. Both families of MMPs are expressed in OA cartilage at lesion sites, where it is presumed that they play a major role in degradation of the extracellular matrix. A recent comprehensive analysis of specific metalloproteinases that are overexpressed by OA cartilage and synovium has revealed several MMPs and ADAMTSs that may be candidates as targets for disease modification. Among the most interesting is MMP-13, which is overexpressed in murine and human OA cartilage and is the most efficient protease capable of cleaving type II collagen. Similarly, the aggrecanase ADAMTS-5 has surfaced as the aggrecanase required for aggrecan loss in experimental OA and inflammatory joint disease. It has been demonstrated that mice lacking ADAMTS-5 are protected from developing OA.

The expression and degradation of noncollagenous proteins and nonaggregating proteoglycans are also altered in OA cartilage and may have a direct or indirect effect on modulating the catabolic state of the chondrocyte.

---

**Figure 98-3** Immunostaining of osteoarthritic cartilage specimen for inducible nitric oxide synthase (A) and interleukin-1β (B). Note intense staining of chondrocytes in the superficial zones for both inflammatory proteins. (From Melchiorri C, Meliconi R, Frizziero L, et al: Enhanced and coordinated in vivo expression of inflammatory cytokines and nitric oxide synthase by chondrocytes from patients with osteoarthritis, Arthritis Rheum 41:2165–2174, 1998.)
groups of molecules are likely to have important structural and/or biologic functions.\textsuperscript{294,295} From their interactions with other extracellular matrix constituents, they can influence the supramolecular assembly of the cartilage matrix and as a result affect the physical properties of the tissue; by interacting directly with chondrocytes and/or neighboring cells, they can provide biologic signals on matrix properties and thereby influence cellular function.\textsuperscript{296}

**Nitric Oxide**

Nitric oxide, produced by the inducible isoform of nitric oxide synthase (iNOS), is a major catabolic factor produced by chondrocytes in response to proinflammatory cytokines such as IL-1β and TNF.\textsuperscript{287} Considerable evidence indicates that the overproduction of nitric oxide by chondrocytes plays a role in the perpetuation of cartilage destruction in OA.\textsuperscript{297,299} (Figure 98-4). Increased concentrations of nitrites have been demonstrated in the synovial fluid of patients with OA, and iNOS has been demonstrated in OA synoviocytes and chondrocytes by in situ hybridization and immunohistochemistry.\textsuperscript{300,301} Although normal cartilage does not express iNOS or produce nitric oxide without stimulation by cytokines such as IL-1, OA cartilage explants spontaneously produce large amounts of nitric oxide;\textsuperscript{307} iNOS is also upregulated from chondrocytes by cartilage compression.\textsuperscript{302,303}

Nitric oxide exerts multiple effects on chondrocytes that promote articular cartilage degradation.\textsuperscript{304} These include (1) inhibition of collagen and proteoglycan synthesis;\textsuperscript{304} (2) activation of metalloproteinases;\textsuperscript{305} (3) increased susceptibility to injury by other oxidants (e.g., H₂O₂);\textsuperscript{306} and (4) apoptosis.\textsuperscript{308} Several studies have implicated nitric oxide as an important mediator in chondrocyte apoptosis, a feature common in progressive OA.\textsuperscript{306,307} Immunohistochemistry of joint tissue obtained from patients with OA reveals colocalization of iNOS protein and apoptosis in articular cartilage cells.\textsuperscript{308} There is evidence that apoptosis results from the formation of peroxynitrite, a toxic free radical produced by the reaction of nitric oxide and superoxide anion. Peroxynitrite reacts with tyrosine residues on proteins, which can be detected by antibodies to nitrotyrosine. Immunostaining of OA cartilage reveals that chondrocytes that are highly positive for IL-1β also stain for nitrotyrosine, consistent with overproduction of peroxynitrite and oxidative damage.\textsuperscript{309} The importance of nitric oxide has been corroborated in animal models of OA. In the Pond-Nuki canine model, the inhibition of nitric oxide reduced the progression of cartilage lesions.\textsuperscript{309}

**Transforming Growth Factor-β**

In most respects, TGF-α acts as a counterregulatory molecule that opposes the effects of inflammatory mediators in cartilage. TGF-β\textsubscript{1} has been shown to downregulate proteolytic MMP-1 and MMP-13, as well as IL-1 and TNF receptors on OA chondrocytes.\textsuperscript{310} TGF-β\textsubscript{3} selectively suppresses the cleavage of type II collagen by collagenases in OA cartilage in culture and limits MMP and proinflammatory cytokine expression.\textsuperscript{258} Recent studies using the murine knee osteoarthritis model indicate that TGF-β\textsubscript{1} protects articular cartilage, as histologic staining for this molecule revealed a lack of TGF-β\textsubscript{1} in damaged cartilage in comparison with normals. Although premature osteophyte chondrocyte clusters express high levels of TGF-β\textsubscript{1}, other data suggest that bone morphogenetic protein 2 (BMP-2) is more responsible for the development of late osteophyte development.\textsuperscript{311} Of note, TGF-β\textsubscript{3} may also exert selected catabolic effects via the stimulation of ADAMTS-4 expression.\textsuperscript{310}

**Hyaluronic Acid**

Hyaluronic acid (HA) has been investigated as a marker of cartilage degradation that can be detected in synovial fluid and serum (see Biomarkers of Osteoarthritis later)\textsuperscript{312} but it also appears to play a role in limiting the progression of arthritis. A recent study by Karna and colleagues found that HA in vitro counteracts the ability of IL-1β to inhibit collagen biosynthesis. At the transcriptional and posttranscriptional levels, chondrocyte cultures with IL-1β upregulated collagen synthesis markers, whereas HA negated this effect.\textsuperscript{313} The same group found that HA similarly protects against IL-1-induced inhibition of collagen synthesis in human skin fibroblasts at the level of the insulin-like growth factor I receptor.\textsuperscript{314}

**Prostaglandins**

The expression of inducible cyclooxygenase-2 (COX-2) is increased in OA chondrocytes, which spontaneously produce PGE\textsubscript{2} ex vivo.\textsuperscript{299} The effects of prostaglandins on chondrocyte metabolism are complex, and include enhanced type II collagen synthesis, activation of metalloproteinases, and promotion of apoptosis.\textsuperscript{315} In cartilage explants, IL-1β induces COX-2 expression and PGE\textsubscript{2} production correlates with proteoglycan degradation. COX-2 inhibition prevents IL-1β–induced proteoglycan degradation, which can be reversed by the addition of PGE\textsubscript{2} to cultures.\textsuperscript{16} In

![Figure 98-4](image-url) Arthroscopic view of osteoarthritic lesion of the femoral condyle (designated chondropathy). Note that proliferative synovitis is localized to the area of the osteoarthritic lesion. (Courtesy Maxime Dougados.)
contrast, in vitro evidence has accumulated that selected nonsteroidal anti-inflammatory drugs (NSAIDs) may interfere with proteoglycan synthesis. Of note, another study concluded that up to 30% of PGE\(_2\) expression in OA synovial tissue stems from the COX-1 pathway. Whether any differences exist between the effects of COX-1- and COX-2–derived prostaglandins on cartilage metabolism is unknown.

**F-spondin**

A novel mediator in OA cartilage, F-spondin is a neuronal extracellular matrix glycoprotein that appears to regulate cartilage degradation through the TGF-β and PGE\(_2\) pathways. One recent article reported a sevenfold increase in expression in human OA cartilage and a significant upregulation in knee cartilage surgical specimens from rats with OA. Addition of F-spondin in vitro to OA cartilage tissue led to increased levels of activated TGF-β and production of PGE\(_2\), as well as accelerated collagen degradation and reduced proteoglycan synthesis, which are both dependent on these two molecules.

**Alterations in Bone**

The inflammatory mediators produced by bone in OA are less well understood than those produced by cartilage and synovium. Biomechanical and biochemical factors seem to influence the remodeling, but the underlying pathogenesis has yet to be identified. Nitric oxide plays a role in bone cell function, which could have implications for OA insofar as it contributes to alterations in subchondral bone. The endothelial isoform, endothelial cell nitric oxide synthase (ecNOS), is constitutively expressed in bone, where it seems to play a key role in regulating osteoblast activity and bone formation. ecNOS also mediates the effects of mechanical loading on the skeleton, where it acts along with prostaglandins to promote bone formation and suppress bone resorption.

In contrast, such proinflammatory cytokines as IL-1 and TNF induce iNOS in bone cells, and NO derived from this pathway potentiates bone loss. Osteophyte formation and subchondral bone remodeling likely result from local production of anabolic growth factors such as insulin-like growth factor 1 (IGF-1) and mostly TGF-β, which is highly expressed in osteocytes of the femoral head in OA patients. Areas of increased radionuclide uptake (“hot spots”) on bone scintigraphy have also been reported to identify OA joints more likely to progress by radiographic criteria—and/or to require surgical intervention—over a 5-year period.

**Alterations in Synovial Tissue**

Synovial lining inflammation and synovial effusions have emerged as another key feature of OA pathophysiology. OA has traditionally been classified as a noninflammatory arthritis, largely because the synovial fluid leukocyte count in OA is typically fewer than 2000 cells/mm\(^3\). Such parameters can be misleading because low-grade inflammatory processes nevertheless occur in osteoarthritic synovial tissues and contribute to disease pathogenesis, and some degree of synovitis has been observed even in early OA. This localized synovitis may be subclinical because arthroscopic studies suggest that localized proliferative and inflammatory changes of the synovium occur in up to 50% of OA patients, and the activated synovium may produce proteases and cytokines that accelerate damage to the adjacent cartilage.

More recently, musculoskeletal ultrasound has provided a reliable noninvasive method to detect both synovial hypertrophy and even small effusions in OA patients, using both grayscale and power Doppler methods.

Many of the clinical symptoms and signs seen in OA joints (e.g., joint swelling and effusion, stiffness, occasionally redness) reflect synovial inflammation. Synovial histologic changes include synovial hypertrophy and hyperplasia with an increased number of lining cells, often accompanied by infiltration of the sublining tissue with scattered foci of lymphocytes. In contrast to rheumatoid arthritis (RA), synovial inflammation in OA is mostly confined to areas adjacent to pathologically damaged cartilage and bone. This activated synovium can release proteases and cytokines that may accelerate destruction of nearby cartilage.

Although synovial macrophages and macrophage-produced mediators result in key inflammatory cascades and cartilage destruction in both OA and RA, it appears there may be differences between the diseases in levels of the cytokines and how they mediate destruction of the cartilage.

As described earlier, the metalloproteinases that degrade cartilage are produced not only by the cartilage itself but also by the synovium. Although cartilage destruction might be directed by the chondrocytes, some degree of synovitis exists in patients even with mild OA. A comprehensive analysis by Davidson and colleagues reported several proinflammatory genes significantly elevated in the synovium that had not previously been reported. Cartilage breakdown products, derived from the articular surface as a result of mechanical or enzymatic destruction of the cartilage, can provoke the release of collagenase and other hydrolytic enzymes from synovial cells and macrophages (Figure 98-5).

Cartilage breakdown products are also believed to result in mononuclear cell infiltration and vascular hyperplasia in the synovial membrane in OA.

A consequence of the low-grade inflammatory processes is the induction of synovial IL-1β and TNF, which are likely contributors to the degradative cascade. There are also reports of increased numbers of immune cells in synovial tissue including activated B cells and T lymphocytes, as well as evidence that OA patients express cellular immunity to the cartilage proteoglycan link protein and C1 domain.

Histologic changes in OA synovium usually demonstrate mild or moderate synovitis characterized by an increase in the number of inflammatory mononuclear cells in the sublining tissue including activated B cells and T lymphocytes.

A recent study of 10 patients with early OA (arthroscopic specimens) and 15 patients undergoing total knee arthroplasty revealed that synovial tissues from early OA had higher levels of IL-1β and TNF and increased mononuclear cell infiltration compared with late OA.

Along the lines of acute episodes of OA, the inflammation may result from crystal-induced synovitis (either calcium apatite or calcium pyrophosphate dihydrate). “Milwaukee shoulder syndrome” represents a rapidly destructive form of OA with evidence of inflammation in the synovial
membrane but minimal synovial fluid leukocytosis. It is typically associated with rotator cuff degeneration, severe shoulder OA, and hydroxyapatite crystal deposition in the synovial membrane. The synovial fluid typically contains few cells and high levels of active collagenase. It is theorized that crystals released from the degenerating tendons trigger the release of collagenase from synovial mononuclear cells that leads to cartilage breakdown; cartilage breakdown products then further activate release of enzymes from the synovium. This inflammation is typically associated with increases in synovial IL-1 and TNF that further proteinate the degradative cascade.

**Biomarkers of Osteoarthritis**

The aims of biomarker research in OA are early detection before irreversible damage occurs, predicting OA progression, and monitoring response to therapeutic intervention. The OA Biomarkers Network, funded by the National Institutes of Health, has established the “BIPED” biomarker classification with five separate categories of surrogate markers: burden of disease, investigative, prognostic, efficacy of intervention, and diagnostic. Examples of burden of disease and prognosis for hip and knee OA include serum COMP, serum hyaluronic acid, and urinary CTXII. A recent survey revealed that this classification scheme was not adopted as often as anticipated, and therefore the authors undertook a systematic review of the primary OA biomarker literature to standardize the findings according to the BIPED classification.

The major structural components in cartilage are type II collagen and aggrecan, which are relatively unique to cartilage. Additional constituents of articular cartilage include the proteins COMP, cartilage intermediate layer protein (CILP), cartilage link protein, matrilin, minor collagens (types I, V, VI, IX, and XI), and hyaluronic acid. In healthy cartilage these molecules have a relatively slow turnover rate, whereas OA is characterized by enhanced synthesis and enzymatic degradation of most of these molecules. Therefore much focus in OA biomarkers has been on detecting markers of cartilage matrix synthesis and degradation. The biomarkers have been made more specific through the detection of “neo-epitopes,” or degraded cartilage matrix fragments generated by specific proteases enhanced in OA. The turnover of subchondral bone is also enhanced in OA, although biomarkers of bone turnover have proven less specific in identifying OA, perhaps because of the continuous bone remodeling in the skeleton as a whole. In addition to markers of matrix degradation and turnover, recent advances in the proteomics and microRNA fields have enabled detection of new OA biomarkers using unbiased screening of serum constituents in arthritic patients. The slow progression of primary OA is causing researchers to refocus the search for OA biomarkers on more rapidly progressing secondary OA after acute injury.

Although OA is not considered an inflammatory disease in the traditional sense, it definitely involves inflammatory processes, and these hold promise as biomarkers for OA. For example, elevated levels of the inflammatory marker C-reactive protein (CRP) appear to be predictive of radiographic progression of long-term knee OA. In a study of 1025 women, higher levels of serum CRP were associated with a statistically significant increase in both prevalent and incident knee OA and greater knee OA severity; women with unilateral knee OA had higher CRP levels than those with unilateral knee OA. Compared with women who did not develop knee OA, those who did had a higher baseline CRP than at their index test 2.5 years earlier. In another study of women (age 44 to 67 years), CRP levels were statistically greater in 105 women with knee OA than in 740 women without OA. These results must be interpreted with caution because a recent study concluded that after adjustment of CRP levels to the patient’s body mass index, serum CRP levels were no longer correlated to OA. In summary, these and other studies report that inflammation markers such as CRP are modestly but significantly increased in early knee OA and can be predictive of OA that will progress over time. Several studies have identified COMP levels as helpful assessors of the potential for, presence of, and progression of OA. This noncollagenous extracellular matrix protein, synthesized by both cartilage and synovium with TGF-β stimulation, is abundant in articular cartilage. Both the degradation and tissue distribution of COMP exhibit marked differences in normal and OA human knee articular cartilage. Synovial fluid COMP levels are higher in individuals with knee pain or injury, anterior cruciate ligament or meniscal injury, and OA than in...
demographically matched healthy individuals. Similarly, serum levels of COMP are often higher in patients with more rapidly progressive joint damage, 225,289,291 with some studies showing this specifically in hip OA 283,359 and knee OA. 290 Although COMP is one of the most useful serum markers for OA, the lack of specificity to OA and relatively high natural variations in serum COMP levels currently necessitate use of additional markers. The development of specific reagents to detect degradative products of COMP may increase its utility as an OA biomarker. 290 As with other biomarkers based on anabolic and catabolic processes in cartilage, serum COMP levels indicate a specific stage of OA.

Hyaluronic acid (HA) is a marker of cartilage degradation that can be detected in synovial fluid and serum, 312 even though the majority of circulating HA originates from extracartilaginous sources. HA levels reflect synovial activity, whereas proteoglycan levels reflect cartilage turnover. 312,360 In addition, higher serum HA levels have been correlated with the number of joints involved and degree of clinical disability. These findings support the theory that serum HA levels reflect synovial hyperactivity. An animal model in which the anterior cruciate ligament was transected to induce OA also demonstrated a rise in serum HA level within 7 days after joint injury, a rise sustained at 13 weeks; the rise in synovial fluid HA levels correlated with serum levels. 361 Serum HA levels, which may also serve as a predictor of OA disease progression, 293,362 correlated with radiographic evidence of disease progression over a 5-year time interval because patients whose disease progressed had higher levels than at the outset. 43

**SUMMARY**

Osteoarthritis, although classically conceived of as a degenerative consequence of aging, is a disease with an increasingly well-characterized molecular pathophysiology. Biomechanical factors, particularly in the context of genetic predisposition, obesity, and malalignment, result in chemical alterations within the joint that promote cartilage degradation. Early, anabolic changes, characterized by proliferation of chondrocytes and increased matrix production, are followed by a predominantly catabolic state, characterized by decreased matrix synthesis, increased proteolytic degradation of matrix, and chondrocyte apoptosis. Many of the features of the chondrocyte in the catabolic state are related to the production of inflammatory mediators by synovium and chondrocytes that act locally to perpetuate cartilage degradation. Although current treatments improve the signs and symptoms of disease, further characterization of the altered metabolism in synovium, cartilage, and bone that promote disease progression should lead to future treatments that prevent structural damage in osteoarthritis.

**Selected References**


Osteoarthritis (OA), the most common form of arthritis, is found worldwide and is strongly associated with aging. It affected 27 million adults in the United States in 2005.1 OA of the knee and hip has significant functional impact due to effects on ambulation and mobility and is associated with considerable medical costs, accounting for 97% of the 455,000 total knee replacements and 83% of the 233,000 total hip replacements for arthritis in 2004.2 Given the aging of our society and the obesity epidemic, the burden of OA can only be expected to increase over the next 20 years.3

OA affects all of the structures in and around a joint and should be considered a failure of the total joint. Historically, the emphasis in OA research has been on cartilage degeneration, but more recent work has expanded this view to an improved understanding of the role of subchondral bone, synovium, ligaments/fpondens, meniscus, muscle, and nerve tissues in the disease process.4,5 The late- or end-stage joint, clinically recognizable as OA, likely represents a final common pathway of many different factors including genetics, environment, and biomechanical contributors.

**Epidemiology of Osteoarthritis**

OA can be defined pathologically, radiographically, or clinically. The American College of Rheumatology (ACR) criteria for the classification of OA of the hand, hip, and knee are shown in Table 99-1.6,8 For the hand, only clinical criteria are used, with a sensitivity of 92% and specificity of 98%. For the hip, the sensitivity and specificity of the ACR criteria are estimated to be 91% and 89%, whereas for the knee they are 91% and 86%, respectively. Due to their high specificity, these criteria are most useful for differentiating OA from inflammatory arthropathy, but less so for differentiating early OA from healthy controls. Compared with radiographic definitions, the ACR criteria tend to underestimate prevalence of OA.9,10

The presence of radiographic OA usually requires the presence of a definite osteophyte or joint space narrowing on plain radiographs, although magnetic resonance imaging (MRI)-based definitions are under development.11 Clinical OA is usually defined by abnormalities on physical examination consistent with OA such as nodal changes in the hands, limited and painful internal rotation of the hip, or crepitus with knee movement. Symptomatic OA is usually defined as pain, aching, or stiffness in a joint with radiographic OA. Definitions can vary according to joint site, frequency or intensity of symptoms, and time span over which symptoms are assessed.

**Prevalence of Radiographic Osteoarthritis**

Because of the multiple different definitions of OA, prevalence estimates vary. The first National Health and Nutrition Examination Survey (NHANES I) found that 12% of the U.S. population had clinically defined OA in at least one joint. A few population-based studies have estimated the prevalence of radiographic OA at the knee, recently reviewed by Lawrence and colleagues.1 Despite variations based on radiographic technique and participant age, these studies estimate that radiographic knee OA is present in 14% to 37% of U.S. adults and is more frequent in women.12,13 For radiographic hip OA, available estimates vary more widely, from less than 1% to 27%, depending on the population being studied.13 Radiographic hand OA is common, especially in older individuals, but it may not be symptomatic or functionally limiting. One large study found radiographic OA of at least one joint in 67% of women and 55% of men aged 55 years and older, with 28% having radiographic OA in at least two of three joint sites (distal or proximal interphalangeal or carpometacarpal).14 A recent review of radiographic findings in foot OA reported a prevalence of 12% to 35% for first metatarsophalangeal OA.15

**Prevalence of Symptomatic Osteoarthritis**

A summary of prevalence estimates for symptomatic OA (symptoms in the presence of radiographic OA) at the hands, knees, and hips is presented in Table 99-2.1 On the basis of these findings, Lawrence and colleagues estimated that more than 9 million U.S. adults are affected by symptomatic knee OA and more than 13 million have symptomatic hand OA.
Table 99-1 American College of Rheumatology Radiologic and Clinical Criteria for Osteoarthritis

<table>
<thead>
<tr>
<th>Site</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand</td>
<td>1. Hand pain, aching, or stiffness on most days of prior mo 2. Hard tissue enlargement of ≥2 of 10 selected joints* 3. Fewer than 3 swollen MCP joints 4. Hard tissue enlargement of ≥2 DIP joints 5. Deformity of ≥2 of 10 selected joints*</td>
</tr>
<tr>
<td>Knee</td>
<td>1. Knee pain for most days of prior mo 2. Crepitis with active joint motion 3. Morning stiffness lasting ≤30 min 4. Bony enlargement of the knee on examination 5. Age ≥38 yr</td>
</tr>
</tbody>
</table>

Table 99-2 Prevalence of Symptomatic Osteoarthritis (OA)

<table>
<thead>
<tr>
<th>Site (Age in Yrs)</th>
<th>Source</th>
<th>% with Symptomatic OA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>Hands (≥26)</td>
<td>Framingham</td>
<td>3.8</td>
</tr>
<tr>
<td>≥26</td>
<td>Framingham</td>
<td>4.6</td>
</tr>
<tr>
<td>≥45</td>
<td>Framingham</td>
<td>5.9</td>
</tr>
<tr>
<td>≥60</td>
<td>Johnston County</td>
<td>13.5</td>
</tr>
<tr>
<td>≥25</td>
<td>NHANES III</td>
<td>10.0</td>
</tr>
<tr>
<td>Hip (≥25)</td>
<td>Johnston County</td>
<td>8.7</td>
</tr>
</tbody>
</table>

Table 99-3 Etiologies of Secondary Osteoarthritis

<table>
<thead>
<tr>
<th>Etiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolic</td>
</tr>
<tr>
<td>Crystal-associated arthritis</td>
</tr>
<tr>
<td>Calcium pyrophosphate or apatite deposition</td>
</tr>
<tr>
<td>Acromegaly</td>
</tr>
<tr>
<td>Ochronosis</td>
</tr>
<tr>
<td>Hemochromatosis</td>
</tr>
<tr>
<td>Wilson’s disease</td>
</tr>
<tr>
<td>Hyperparathyroidism</td>
</tr>
<tr>
<td>Ehlers-Danlos</td>
</tr>
<tr>
<td>Gaucher’s disease</td>
</tr>
<tr>
<td>Diabetes</td>
</tr>
<tr>
<td>Mechanical/Local Factors</td>
</tr>
<tr>
<td>Joint trauma (e.g., ACL tear)</td>
</tr>
<tr>
<td>Fracture through joint</td>
</tr>
<tr>
<td>Prior joint surgery (i.e., meniscectomy, ACL)</td>
</tr>
<tr>
<td>Charcot joint (neuropathic arthropathy)</td>
</tr>
<tr>
<td>Inflammatory</td>
</tr>
<tr>
<td>Rheumatoid arthritis or other inflammatory arthropathies</td>
</tr>
<tr>
<td>Crystalline arthropathy (gout)</td>
</tr>
<tr>
<td>History of septic arthritis</td>
</tr>
</tbody>
</table>

Primary and Secondary Osteoarthritis

Histologically, osteoarthritis was considered to be “primary” in the absence of an injury history or other joint disease and “secondary” if a predisposing disorder was present (Table 99-3). However, as more and more local risk factors for OA have been identified (such as femoroacetabular impingement at the hip and malalignment at the knee) and a broader range of associated factors have been discovered (genetic, biomechanical, and environmental factors), the division between primary and secondary is less clear. Many individuals who develop secondary OA are likely predisposed to the condition with or without the identified inciting event; other individuals who have a disorder that is linked to secondary OA may not develop clinical OA. It may be most useful to think of OA as a common pathway through which an individual’s genetics, history of injury or other joint damage, mechanical factors, and psychosocial milieu act on the joint, in some cases leading to an “end-stage” or “failed” joint.

**CLINICAL FEATURES**

General Symptoms and Signs

OA most commonly affects the knees, hands, feet, hips, and spine. These joints may be symptomatic or may be affected only on radiographs. Individuals with OA generally describe pain in the joint(s) that is worse with activity, with limited morning stiffness (<30 minutes), and pain and stiffness with rest. This stiffness after inactivity, or “gelling” phenomenon, is often a main complaint, although morning stiffness is generally less severe and of shorter duration than that seen...
in systemic inflammatory arthropathies. Affected joints in OA often demonstrate bony enlargement and crepitus on examination, with concomitant reductions in range of motion. There may be soft tissue swelling or effusion, although these tend to be much less dramatic than in inflammatory arthritis. Pain complaints may be more or less than expected on the basis of structural damage. 18,19 The effects of depression, disturbed sleep, and other psychosocial factors on the pain experience in OA are being increasingly recognized. 20-23

Joint-Specific Symptoms and Signs: Knee

Knee OA is characterized by the insidious onset of pain with gelling and limited range of motion. Individuals with knee OA often describe pain and limitation with walking, transferring (as from seated to standing), and especially stair climbing. These complaints are often associated with a sensation of instability or “giving out” at the knee. A “locking” sensation at the knee can be a consequence of stiffness, loose bodies in the joint space, or meniscal lesions. Knees with OA often have demonstrable crepitus and bony enlargement. Pain may be elicited by palpation of the medial or lateral joint line, or both. Effusions, when present, are often cool and generally without redness. They can be associated with popliteal bursa enlargement (Baker’s cyst) when large. Associated pain over the anserine bursa, or even the greater trochanter, is often seen in knee OA and may be related to altered biomechanics. 24 Appreciation of such soft tissue symptoms is important because these may be amenable to corticosteroid injections with subsequent pain relief.

Malalignment, most often varus, is often seen in severe disease but can be present even in fairly mild/early disease. Clinically evident varus thrust may be a risk factor for progression. 25 Other signs of severe disease include flexion deformities or joint instability. Quadriceps weakness represents an early modifiable risk factor for knee OA progression, particularly in women. 26,27 and in late stages of disease may be apparent as muscle atrophy. 28,29 Alterations in proprioception and vibratory sense have been demonstrated in association with knee OA, although the relation of these factors to progression and pain is still unclear. 30-32

Patellofemoral OA can strongly contribute to pain and disability at the knee but is often overlooked. 33 OA of the patellofemoral joint is characterized by pain with ascending or descending stairs and is often located anteriorly. It can be seen in isolation or in association with tibiofemoral OA. The relation between patellofemoral OA and commonly seen patellofemoral pain disorders in younger individuals remains to be elucidated. 34

Joint-Specific Symptoms and Signs: Hip

OA of the hip can present with groin pain, which is fairly specific, 35 but may be described more vaguely as pain in the thigh, buttck, low back, or even in the ipsilateral knee. Therefore it is important to assess for other causes of pain in the “hip” including spinal pain (lumbar disk degeneration, spinal stenosis, facet joint OA, sacroiliac pain); trochanteric bursitis; altered gait due to knee pathology; meralgia paresthetica (lateral femoral cutaneous nerve entrapment); thigh claudication from vascular causes; or even intrapelvic causes. It is also important to consider other causes of hip and groin pain such as occult femoral neck fracture or avascular necrosis. Persons affected by hip OA have limitations in walking, bending, and transferring, as well as stair climbing. Internal rotation of the affected hip is often limited and can be quite painful, even in early disease, often evident to the patient as difficulty putting on socks, tying shoes, or trimming toenails. Visible deformity, hip flexion contracture, or severe limitations of range of motion are indicators of more advanced disease, which may also be associated with shortening of the affected limb due to superior migration of the femoral head. In young individuals presenting with groin pain that is worsened by sitting, with pain and limitation when internally rotating and adducting the hip in the flexed position, femoracetabular impingement is a consideration. 36

Joint-Specific Symptoms and Signs: Hand

The hands often provide the first clue to a diagnosis of OA. Evident bony enlargement of the distal interphalangeal (DIP) joints, called Heberden’s nodes, and proximal interphalangeal (PIP) joints, called Bouchard’s nodes, can be appreciated (Figure 99-1). These nodes may be acutely inflamed with warmth and tenderness or may be bland, hard enlargements and are often more marked in the dominant hand. Some patients, most commonly elderly women, have erosive osteoarthritis characterized by episodic inflammation, pain, and swelling. It remains a matter of debate whether this type of OA is part of the continuum or is itself a separate disease entity. 37,38 OA involvement of the first carpometacarpal (CMC) joint is particularly problematic and can lead to significant pain, limitations in functionality of the hand, and reduced grip strength. 39,40 CMC squaring, representing deformity of the joint due to osteophyte formation and joint space narrowing, can be seen on examination. Bilateral involvement of multiple joints, both within (multiple PIPs) and across (both DIPs and PIPs) joint groups, is frequent. Metacarpophalangeal joints are affected more commonly than previously recognized, 41-43 although prominent involvement of these joints should prompt consideration of inflammatory arthropathies or secondary causes of OA such as hemochromatosis. 44 Again, soft tissue findings such as deQuervain’s tenosynovitis should be considered because they are associated with hand OA, can mimic or aggravate symptoms, and are amenable to conservative management.

Joint-Specific Symptoms and Signs: Spine

Facet joint arthritis is a common finding, is associated with back pain, and has many features in common with OA of other diarthrodial joints, 45 although few studies of this finding in association with OA at other sites have been reported. 46 Osteophytosis of the spine is nearly ubiquitous in older individuals, although it is often asymptomatic. 47 Lumbar disk degeneration (LDD), characterized by disk space narrowing, end-plate sclerosis, and herniation, is often seen in association with radiographic osteoarthritis, 48 although the relationship between LDD and osteoarthritis of other joint sites remains controversial. 49 Clinical
susternoclavicular joints than the glenohumeral joint itself. Other maladies common in older adults such as subacromial bursitis, rotator cuff pathology, and adhesive capsulitis should be assessed. Rotator cuff damage, in particular, can predispose to glenohumeral OA. Pathology at the cervical spine can present as pain in the shoulder region, so the shoulder evaluation should include an assessment of the cervical spine. Milwaukee shoulder syndrome is a destructive arthropathy of the glenohumeral joint associated with large effusions, which on aspiration often reveal high red blood cell content and basic calcium crystals.

Joint-Specific Symptoms and Signs: Other Joints

Osteoarthritis of the foot has been largely neglected in the literature to date. OA of the first metatarsophalangeal (MTP) joint is characterized by pain and hallux valgus (bunion) deformity and has radiographic features (joint space narrowing, osteophytosis) similar to OA at other joints. Loss of function at the first MTP joint due to ankylosis (hallux rigidus) can lead to altered gait. Other joints that can be affected by OA but are often not part of epidemiologic studies include the ankles (talonavicular and subtalar) and temporomandibular joint (TMJ). Elbow OA is rare and may be the result of trauma, vibration damage, or other entities such as pseudogout.

Polyarticular Osteoarthritis

It has been recognized for more than 100 years that OA can co-occur in multiple joint sites. Although the term generalized OA (GOA) has been used to describe this disease entity, there is no universally understood or accepted definition of what constitutes GOA (Figure 99-2). Kellgren and Moore in 1952 provided the first clinical description of GOA, involving primarily Heberden’s nodes and the CMC joints, with the spine, knees, hips, and feet involved in descending frequency. Later studies have defined GOA as more than three or more than five joint sites affected, affected joint counts, multiple hand joint involvement, nodal hand OA with other joint involvement, or summed scores of OA across multiple joints. It remains unclear which joints should be included in such definitions, particularly whether lumbar disk degeneration and hip OA are part of the same disease entity as other joints. Clinically, it is important to recognize and consider that a patient with OA in one joint is likely to have involvement at other sites and to assess the impact of that individual’s overall OA burden on their functional status, rather than focusing on a single joint.

DIAGNOSTIC TESTING

Diagnostic Approach

The diagnosis of OA is a clinical one, and laboratory testing is rarely required. Similarly, if the diagnosis is clear on the basis of history and clinical findings, radiographs are often not required. The purpose of additional diagnostic testing in OA is primarily to exclude potentially treatable underlying conditions such as metabolic or inflammatory arthropathies.
Laboratory Testing

There is rarely an indication for testing rheumatoid factor, antinuclear antibodies, or other serologic studies in the setting of clinically suspected OA, and such tests should be reserved for a patient in whom there are findings suggestive of an inflammatory arthropathy. Complete blood count, chemistry panel including glucose and creatinine, and liver function tests should be obtained before initiation of pharmacologic therapy for OA, especially in older individuals with comorbid medical conditions, due to an increased risk of adverse events in this population. In cases where there is prompt involvement of the metacarpophalangeal (MCP) joints, evaluation for hypothyroidism and hemochromatosis may be warranted.

Synovial Fluid

The synovial fluid in OA is typically normal or mildly inflammatory, appearing clear and colorless to slightly yellowish. The leukocyte cell count is typically less than or equal to 2000 cells/mm³ (<2 cells seen across 10 high-power fields). Fluid is often obtained in the course of a symptomatic joint injection, although diagnostic aspiration may be done in the setting of an effusion. Concomitant calcium pyrophosphate crystal disease can be identified, although other calcium crystals, or hydroxyapatite, are not seen on routine preparations.

Molecular Biomarkers

Quantitative determination of the products of cartilage and bone metabolism has provided several putative biomarkers of OA pathophysiology. As an example, one such biomarker, urinary C-telopeptide fragments of type II collagen (u-CTXII), is associated with the occurrence and progression of radiographic OA, independent of other risk factors. Such biomarkers, obtained from serum, urine, or synovial fluid, may eventually provide a method for early diagnosis and monitoring of treatment effect, although they are currently used primarily in research. Interested readers are referred to a comprehensive review of biomarkers by van Spil and colleagues.

Imaging: Conventional Radiography, General Considerations

Conventional radiography is widely available and relatively inexpensive, making this modality useful to confirm the diagnosis and exclude others, primarily when there is clinical uncertainty regarding the diagnosis, when the joint involved is atypical, or when evidence exists to suggest other diagnoses such as inflammatory arthritis, fractures, Paget’s disease, osteonecrosis, infection, or malignancy. Radiographs of joints affected by osteoarthritis typically show osteophytes, joint space narrowing, sclerosis, and cysts of subchondral bone. The Kellgren-Lawrence (KL) grading system remains the most commonly used for research purposes. KL grades range from 0 (no osteophytes or joint space narrowing) to 4 (severe joint space narrowing with subchondral sclerosis); a grade of 2 is generally considered diagnostic of osteoarthritis. The KL system relies on osteophytes and may underestimate OA in individuals with a more atrophic subtype (where joint space narrowing is more severe than osteophytosis, particularly at the hip). Other grading systems, such as the Osteoarthritis Research Society International (OARSI) grading system, view osteophytes and joint space narrowing separately and assign separate scores (Figure 99-3). The choice of grading system depends on the goal of the study and the joint site of interest.

Imaging: Conventional Radiography, Specific Joint Issues

Imaging of the knees should be bilateral and in weight bearing, and for clinical purposes it should generally be anteroposterior (AP). Lateral, posteroanterior, or sunrise views may be necessary in some circumstances. Alignment, although best determined from full-limb views, can be estimated from routine knee films. An AP pelvis view should be obtained to assess hip OA and can be supine or weight bearing. Additional views such as frog-leg or lateral images may be indicated for specific hip pathology or to assess for femoroacetabular impingement or congenital abnormalities. An atrophic subtype of hip OA, characterized by joint space narrowing with either mild or absent osteophytosis, has been described, and alternative radiographic definitions for hip OA (other than KL grading) have been proposed and used in research to address this issue. For hand OA, a posteroanterior view of both hands including the
a response to findings on MRI, are overused and generally ineffective; and the cost associated with these studies, they are generally not indicated in routine clinical practice at present. Ultrasound is gaining in popularity and may have a role at the bedside in detecting small effusions, identifying early cartilage changes, differentiating inflammatory from noninflammatory arthropathies, and as a therapeutic adjunct to allow more accurate aspirations and placement of intra-articular injections.

OUTCOME

Performance Measures and Functional Assessment

Although not routinely used in clinical practice, there are several validated tools for assessing functional status in OA that are widely used in research settings. Results on standardized performance-based tests such as determination of walking speed and time to stand from a chair five times unassisted have shown to correlate with the severity and number of affected joints in OA. However, due to a lack of effective treatment options in OA; growing evidence that surgical interventions such as knee arthroscopy, which are often

Figure 99-3 Examples of semiquantitative radiographic assessment with use of the Kellgren-Lawrence and Osteoarthritis Research Society International (OARSI) grading schemes. A, Kellgren-Lawrence grade 3. No lateral femoral and tibial osteophytes are seen (OARSI grade 0). A medial femoral osteophyte: OARSI grade 1 (white arrow); a medial tibial osteophyte: OARSI grade 2 (white arrowhead); lateral tibiofemoral joint-space width: OARSI grade 0; and medial tibiofemoral joint-space narrowing: OARSI grade 2 (black arrows) are depicted. B, Kellgren-Lawrence grade 2. A lateral femoral osteophyte: OARSI grade 2 (white arrow); a lateral tibial osteophyte: OARSI grade 2 (white arrowheads); a medial femoral osteophyte: OARSI grade 3 (short black arrows); a medial tibial osteophyte: OARSI grade 2 (black arrowhead); a normal lateral tibiofemoral joint-space width: OARSI grade 0; and medial tibiofemoral joint-space width: OARSI grade 1 (long black arrows) are shown. (From Guermazi A, Hunter DJ, Roemer FW: Plain radiography and magnetic resonance imaging diagnostics in osteoarthritis: validated staging and scoring, J Bone Joint Surg Am 91(Suppl 1):54–62, 2009.)
and Australian/Canadian Hand Osteoarthritis Index (AUSCAN, for hand-specific function), are often used to assess the functional impact of OA on affected individuals for research purposes.

**Time to Total Joint Replacement**

Total joint replacement is a “hard” endpoint in OA, but its use as an outcome measure is hampered by a variety of factors affecting an individual patient’s decision to undergo this procedure including insurance coverage, procedure availability, comorbid conditions, ability to rehabilitate the joint, and patient preference. Therefore efforts are ongoing to identify a composite index that would define the “need for total joint replacement,” which could be used as an outcome measure.

**Mortality in Osteoarthritis**

Moderate evidence indicates increased mortality in individuals with OA compared with the general population. The apparent modestly increased mortality risk, determined through a review of nine studies, appears to be primarily due to cardiovascular and gastrointestinal causes. One study showed an increased mortality risk with increasing numbers of joint groups affected by OA, as well as reduced survival in individuals with hand, bilateral knee, or cervical spine involvement, but not in those with OA of the hip, foot, or lumbar spine. Contributors to increased mortality risk include a combination of reduced physical activity among individuals with OA, comorbid conditions, and/or adverse medication effects such as are seen with acetaminophen and nonsteroidal anti-inflammatory drugs.

---

**Table 99-4**

**Western Ontario and McMaster Universities Arthritis Index (WOMAC)**

<table>
<thead>
<tr>
<th>Pain Subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much pain do you have …</td>
</tr>
<tr>
<td>Walking on a flat surface?</td>
</tr>
<tr>
<td>Going up or down stairs?</td>
</tr>
<tr>
<td>At night while in bed?</td>
</tr>
<tr>
<td>Sitting or lying?</td>
</tr>
<tr>
<td>Standing upright?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stiffness Subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td>How severe is your stiffness …</td>
</tr>
<tr>
<td>After first waking in the morning?</td>
</tr>
<tr>
<td>After sitting, lying down, or resting later in the day</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function Subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td>What degree of difficulty do you have with …</td>
</tr>
<tr>
<td>Descending stairs?</td>
</tr>
<tr>
<td>Ascending stairs?</td>
</tr>
<tr>
<td>Rising from sitting?</td>
</tr>
<tr>
<td>Standing?</td>
</tr>
<tr>
<td>Bending to floor?</td>
</tr>
<tr>
<td>Walking on a flat surface?</td>
</tr>
<tr>
<td>Getting in or out of a car?</td>
</tr>
<tr>
<td>Going shopping?</td>
</tr>
<tr>
<td>Putting on socks or stockings?</td>
</tr>
<tr>
<td>Rising from bed?</td>
</tr>
<tr>
<td>Taking off socks or stockings?</td>
</tr>
<tr>
<td>Lying in bed?</td>
</tr>
<tr>
<td>Getting in and out of the bath?</td>
</tr>
<tr>
<td>Sitting?</td>
</tr>
<tr>
<td>Getting on or off the toilet?</td>
</tr>
<tr>
<td>Heavy domestic duties?</td>
</tr>
<tr>
<td>Light domestic duties?</td>
</tr>
</tbody>
</table>

SUMMARY

OA is a highly prevalent and debilitating condition, with increasing prevalence along with the aging of our society and the ongoing obesity epidemic. OA is readily recognized on clinical grounds, although further diagnostic testing and imaging are indicated in some cases to assess for potentially treatable underlying conditions. Pain is the most common reason for a patient with OA to seek medical advice, and examination often reveals loss of function. Assessment of the patient’s pain symptoms and functional limitations, in the framework of that individual’s goals, is necessary to formulate an appropriate plan of care in the absence of disease-modifying medications. Current research in OA, in addition to ongoing efforts in both pharmacologic and non-pharmacologic therapeutic development, is focused on optimizing incident case and progressor definitions and understanding the whole-body burden of OA.

References


63. Adams R: A treatise on rheumatic gout, or chronic rheumatic arthritis of all the joints, ed 2, London, 1873, John Churchill.


The references for this chapter can also be found on www.expertconsult.com.
Treatment of Osteoarthritis

CARLOS J. LOZADA

Osteoarthritis (OA) is the most common form of arthritis. It is often referred to by other names such as arthrosis, osteoarthrosis, or simply arthritis. Because its incidence increases with age, OA is becoming an increasingly important health issue with the “graying” of the world’s population.

OA can be defined radiographically or clinically. The most useful definition, however, includes symptoms and radiographic changes. If a purely radiographic definition is used, it can be demonstrated that almost all individuals older than 75 years have OA. Although the epidemiology of OA is well covered in Chapter 99, it has been estimated that between 10% and 30% of those affected with OA are significantly disabled, making OA the leading cause of chronic disability in the United States. This leads to substantial direct and indirect costs.

Traditional treatment paradigms for OA have conceded the inexorable progression of the disease and concentrated on pain management. A simplistic but potentially useful algorithm is provided in Figure 100-1. As the population ages, there will be increasing societal pressure on physicians, particularly rheumatologists, to improve the available treatments for OA. Researchers have turned to the investigation of agents that might delay the progression of OA. Particular investigational agents have included collagenase inhibitors, nutritional supplements, and polysaccharides, although many novel molecular entities are now under exploration.

PATIENT ASSESSMENT

Appropriate management of OA begins with an accurate diagnosis. As with most rheumatic illnesses, obtaining a good history is of paramount importance. Symptoms should be carefully described, particularly pain. Duration, location, and any alleviating or exacerbating factors should be ascertained. Distinct features such as stiffness or gelling and the description of events such as “locking” or “giving way” of a joint can help direct the physical examination.

The physical examination seeks to confirm the diagnostic suspicion and establish the causes of symptoms. Laboratory evaluations are not helpful in establishing the diagnosis of OA but can help in excluding alternative diagnoses. They are also useful in determining which therapeutic approaches are appropriate for a particular patient because conditions such as renal insufficiency or anemia can be identified.

Radiographs are not necessary for diagnosis in the majority of patients but can identify coexistent conditions such as chondrocalcinosis that may require further workup or modification of the therapeutic plan.

SOURCE OF PAIN

The main symptom in OA is pain. It has many potential sources in and around the joint. These include focal synovitis, synovial effusions, subchondral bone pain receptors, and periarticular tendons and bursae. Factors complicating the determination of the source of pain may include varus or valgus deformity, weight issues, and the emotional impact of chronic pain. Once the source or sources of pain are accurately identified, a treatment plan can be formulated.

MANAGEMENT

The management of OA can be divided into nonpharmacologic interventions (Table 100-1), pharmacologic interventions, and surgical options. Pharmacologic interventions can be further subdivided into symptomatic therapy and potential structure- or disease-modifying therapy.

Nonpharmacologic Interventions

Psychosocial Interventions

As in other types of arthritis, patient education is an important first step in OA therapy. The patient should be an integral part of the decision-making team. To do this effectively, the patient should understand the nature of OA including its natural history and treatment options. It is often reassuring for the patient to realize that OA is a common, slowly progressive ailment and is not typically as disabling or deforming as the inflammatory arthritides. A significant number of patients have already tried non-prescription medications or nutraceutical remedies before...
seeing a physician and will want to discuss these options. Physicians should emphasize that treatment includes nonpharmacologic and pharmacologic interventions. Organizations such as the Arthritis Foundation can be valuable sources of information geared toward patients and can provide helpful reading materials.

Some patients may develop significant emotional disturbances related to the pain and changes in normal daily activities that can stem from OA. These may include mood disorders such as depression or sleep disorders. Worsened measures of mental health have been associated with increased OA pain and risk of flares. Suspicion of either condition should lead to an evaluation by a psychiatrist or a primary physician who regularly manages these types of disorders.

### Weight Loss

Obesity is an important risk factor in the development of OA of the knee. Further, higher body mass index (BMI) has been associated with an increased risk of progression of OA of the knee. This can be compounded by malalignment—namely, varus and valgus deformities that modulate the effect of weight on knee OA. In one study, BMI was associated with OA severity in those with varus deformity but not in those with valgus.

Regimens of weight loss and exercise have been associated with improvement in pain and disability in OA of the knee. Weight loss alone has been associated with a decrease in the odds of developing symptomatic knee OA. One study suggested that a reduction in the percentage of body fat, rather than weight, may be significant in reducing pain from OA of the knee. The symptom-relieving effects of weight loss have been shown to last as long as 1 year.

### Table 100-1  Nonpharmacologic Management of Osteoarthritis

<table>
<thead>
<tr>
<th>Conventional Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient education</td>
</tr>
<tr>
<td>Arthritis self-help courses</td>
</tr>
<tr>
<td>Weight loss</td>
</tr>
<tr>
<td>Temperature modalities</td>
</tr>
<tr>
<td>Exercise</td>
</tr>
<tr>
<td>Orthotics</td>
</tr>
<tr>
<td>Modified activities of daily living</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unconventional Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcutaneous electrical nerve stimulation</td>
</tr>
<tr>
<td>Pulsed electromagnetic fields</td>
</tr>
<tr>
<td>Static magnets</td>
</tr>
<tr>
<td>Acupuncture</td>
</tr>
<tr>
<td>Spa therapy</td>
</tr>
<tr>
<td>Yoga</td>
</tr>
</tbody>
</table>
combination of weight loss and exercise can be superior to either intervention alone.\textsuperscript{13}

**Temperature Modalities**

Topical applications of heat or cold can be a helpful adjunct to the therapeutic plan. These are more effectively used in superficial joints such as the knees than in deep ones such as the hip. An acute injury such as a sprained ankle calls for cold applications for the first 2 to 3 days.\textsuperscript{18} In a setting of chronic pain, most patients prefer warm applications, although if superior pain relief is obtained from cold applications, these can be continued.

Warm applications can be in the form of warm soaks or heating pads. Individual sessions should not exceed a temperature of 45° C or last more than approximately 30 minutes.\textsuperscript{19} The application of warmth should be avoided over certain areas such as close to the testicles and in patients with poor vascular supply, neuropathy, or cancer. Benefits of warm applications include decreased pain and stiffness, along with relief of muscle spasm and prevention of contractures.

**Exercise**

Periarticular structures, particularly muscles, influence the expression of OA. This is likely due to their role in providing stability to the joints and in dampening some of the forces acting across joints. Quadriceps muscle weakness has been postulated as a risk factor for OA of the knee.\textsuperscript{20} Quadriceps strengthening exercises have been advanced as fundamental to the management of conditions such as chondromalacia patellae.\textsuperscript{21}

Both the dynamic and isometric exercise arms of a 16-week study of patients with knee OA showed equivalent improvement in symptoms and physical functioning.\textsuperscript{22} Walking can be beneficial, and supervised fitness-walking regimens can improve function in those with OA of the knee.\textsuperscript{23} Home-based exercise interventions also significantly improve symptoms in those with knee OA.\textsuperscript{23,24} Finally, community-based aquatic exercise programs such as aquatic aerobics have merit.\textsuperscript{25}

**Orthotics and Bracing**

Orthotics—ranging from insoles to braces—can be effective in providing symptomatic relief and are probably underused by most physicians. Studies have demonstrated that lateral wedged insoles provide substantial relief to those with medial compartment knee OA, particularly those with varus deformity.\textsuperscript{26} In some studies, those with milder symptoms obtained greater benefit.\textsuperscript{27} Knee braces have been evaluated as well. Valgus bracing of patients with medial compartment OA can reduce pain and increase levels of activity.\textsuperscript{28} In one study, medial taping of the patella reduced the pain of those with patellofemoral compartment OA by 25%.\textsuperscript{29}

Heel lifts have been tried in those with hip OA. In an uncontrolled study, most patients reported diminished symptoms. Time to improvement lengthened with the radiographic stage of OA.\textsuperscript{30} For those with calcaneal spurs or foot joint OA in general, appropriate athletic-type footwear is recommended. A good athletic shoe should provide medial arch support and calcaneal cushioning, as well as good mediolateral stability.

Those with carpometacarpal joint arthritis should initially be offered conservative management including the use of splints. In one trial, 70% of patients treated with a 7-month intervention that included the use of splints were able to improve their symptoms considerably and avoid surgical intervention.\textsuperscript{31}

**Cane/Walking Aid**

The appropriate use of a cane (walking stick) can be an important adjunct, particularly in OA of the hip. It has been estimated that a cane can provide up to a 40% reduction in hip contact forces during ambulation.\textsuperscript{32} The cane should be used in the hand contralateral to the affected hip or knee\textsuperscript{13} and should be advanced with the affected limb while walking. The appropriate cane size is that which results in a 20-degree flexion of the elbow during use.\textsuperscript{34} A useful approximation is a cane that is equal to the distance from the floor to the patient’s greater trochanter.

**Modification in Activities of Daily Living**

Physician advice and occupational therapy can provide useful insights into modifications of daily activities to reduce OA symptoms. These interventions can range from using an elevated toilet seat or shower bench in someone with lower extremity OA to using appliances designed to open jars in patients with hand OA. Assistance from occupational therapists can be valuable.

**Other Interventions**

Other modalities have been tried in OA. These are unconventional and include magnetic field application, acupuncture, and yoga-based regimens. These are not accepted as standard therapy for OA, but some deserve further study. A significant number of these interventions are being used by patients on their own and should be formally studied not only for evidence of any benefit but also to ensure that there are no harmful effects.

Studies of transcutaneous electrical nerve stimulation (TENS) have generally been small. A review of TENS studies in OA of the knee concluded that a trend toward symptom improvement existed, warranting larger, well-controlled studies.\textsuperscript{35} In one randomized, controlled study, patients had initial symptom reduction, but at 1-year follow-up, only two patients continued to use the device.\textsuperscript{36} TENS use for 3 weeks was compared with three weekly hyaluronic acid injections in 60 patients with OA of the knee. Pain relief was observed in both groups through the 6 months of follow-up. There was superior improvement in the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) physical function subscale score for the hyaluronic acid group.\textsuperscript{37}

Pulsed electromagnetic fields have been tested in double-blind, placebo-controlled trials. These fields are applied through the daily use of a brace-type device. In one study, a primary endpoint of pain reduction was not achieved.\textsuperscript{38} Another study did not meet its primary endpoint but
reported an improvement in knee stiffness in subjects younger than 65 years, without an accompanying reduction in pain.39

The use of static magnets in chronic knee pain has become popular with some patients. In one double-blind, randomized, placebo-controlled trial of 43 patients, the WOMAC pain and physical function subscales, along with a 50-foot walk, demonstrated a statistically significant benefit of static magnets at 2 weeks.40 Another 29-patient double-blind, placebo-controlled trial in knee OA reported a benefit over placebo after 4 hours of use, but there were no significant differences between groups at 6 weeks of continued treatment.41 The potential mechanism for any effect remains unclear, and larger, longer-term studies are necessary before any clinical benefit can be postulated.

Acupuncture is being formally tested in a National Institutes of Health (NIH)–sponsored multicenter clinical trial. It has been difficult to develop appropriate controls to test acupuncture’s clinical efficacy. Most recent studies have tried to employ “sham” methods in the control arm such as the use of blunted, telescopic needles.42 Early clinical trials43,44 and one literature review45 concluded that acupuncture shows promise in the treatment of knee pain from OA. A double-blind, randomized, placebo-controlled trial of acupuncture as adjunctive therapy in OA of the knee enrolled 570 patients in two outpatient clinics. Reduction in knee pain in the true acupuncture group was superior to that in the sham acupuncture group at 26 weeks by WOMAC function score, WOMAC pain score, and patient global assessment. Twenty-five percent of the patients in each of the acupuncture groups were unavailable for analysis at 26 weeks, however.46

The most recent and largest randomized, double-blind, placebo-controlled trial of acupuncture in knee OA showed a benefit of both sham and “traditional” methods of acupuncture over physiotherapy and as-needed nonsteroidal anti-inflammatory drugs (NSAIDs); however, there were no significant differences between the sham and “traditional” arms of the studies in terms of OA symptom relief. The beneficial pain-relieving effect seen in slightly more than half the patients in each of these arms appeared to be secondary to the use of the needles themselves rather than the specific locations where they were placed.

Spa therapy also has advocates. It has been touted for low back pain and for lower extremity OA.47 However, randomized, controlled studies are lacking.48 Yoga has also shown some symptomatic benefit in OA of the hands on the basis of limited testing.49

Pharmacologic Interventions

Topical Agents

Topical agents for the management of OA are available without a prescription in the United States (Table 100-2). The two most widely used types are preparations containing capsaicin and those containing topical NSAIDs.

Capsaicin is a pungent ingredient found in red peppers (such as hot chili peppers). The mechanism of action is thought to be through selective stimulation of unmyelinated type C afferent neurons, causing the release of substance P. This release reversibly depletes the stores of substance P, a neurotransmitter of peripheral pain sensations.50 Capsaicin preparations are available in concentrations of 0.025% or 0.075% in either ointment or, more recently, “roll-on” form, and they can be applied up to four times daily. They have been tested in controlled, double-blind studies in OA of the hands and knees.51,52 Patient response is quite variable, with some obtaining significant pain relief and others not being able to tolerate the burning or stinging sensation produced by its application. Usually, the counterirritant sensation decreases gradually with repeated use, but pain relief remains. Although safe overall, capsaicin products can be irritating if they come in contact with mucosal surfaces, particularly the eyes. Patients should wear disposable gloves, if possible, when applying the agents. There may be some reddening of the skin where the compound is applied.

Topical NSAID preparations are popular worldwide for the treatments of OA.53,54 Safety concerns about traditional oral NSAIDs were the driving force in the use of these topical agents,55 although questions remain as to their absorption and the degree of relief obtained. Results of placebo-controlled trials in OA of the knee have been conflicting. Some demonstrated symptomatic relief with topical application of gels containing NSAIDs such as diclofenac,56,57 whereas others showed only trends favoring the NSAID or no difference at all. In one trial, diclofenac gel was compared with placebo in 238 patients with OA of the knee over 3 weeks. The primary outcome was average pain with movement on days 1 to 14. The group on diclofenac gel had statistically superior improvement in this variable compared with those on placebo. WOMAC scores for function, pain, and disability were also significantly superior to placebo at weeks 2 and 3.58 Transdermal diclofenac patches are also available and can be applied twice daily to painful articular locations. Patients were also randomized to receive eltenac or placebo gel over 4 weeks. Eltenac is a noneselective NSAID that is structurally similar to diclofenac. The primary endpoint was global pain on a visual analogue scale (VAS). At 4 weeks, there was a trend, but no statistical difference, favoring the eltenac gel. Two patients in the active treatment group and two in the placebo group had local itching, reddening, or both in the application area. There are also menthol- and salicylate-based over-the-counter topical preparations, but there are

<table>
<thead>
<tr>
<th>Table 100-2</th>
<th>Symptom-Relieving Pharmacologic Therapies for Osteoarthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topical</strong></td>
<td>Capsaicin</td>
</tr>
<tr>
<td><strong>Systemic</strong></td>
<td>Acetaminophen</td>
</tr>
<tr>
<td><strong>Intra-articular</strong></td>
<td>Corticosteroids</td>
</tr>
</tbody>
</table>

NSAID, nonsteroidal anti-inflammatory drug.
no published trials supporting their use in OA. Finally, lidocaine has gained popularity as a topical agent for musculoskeletal pain. Transdermal lidocaine 5% patches are available for management of pain and can be applied to up to three articular and periarticular locations at a time as an analgesic agent for 12-hour periods.

**Systemic Agents**

**Non-narcotic Analgesics.** Acetaminophen (paracetamol) has often been touted as the initial systemic intervention for the management of OA. This is mainly due to its favorable side effect profile but also to a perception of its equivalent efficacy to NSAIDs. This perception derives from studies of OA in which patients were not stratified in terms of degree of symptoms. In one study, acetaminophen 4 g/day was equivalent to ibuprofen 1200 or 2400 mg/day, with the notable exception of pain at rest. A meta-analysis of 10 randomized, controlled trials concluded that acetaminophen is effective in the relief of pain associated with OA. However, the effect was small, and there was no improvement in overall WOMAC score. This suggests that acetaminophen may be effective for the relief of pain and should not be expected to have a strong effect on stiffness or function. More recently, it has been noted that NSAIDs may have superior efficacy in patients with more symptomatic or inflammatory presentations because acetaminophen has no anti-inflammatory effects at approved doses. A recent database review concluded that the available evidence suggests that NSAIDs have superior efficacy in symptomatic relief in those with hip or knee OA and also in those with moderate to severe levels of pain from OA. Particular concerns in patients taking acetaminophen include the concomitant use of alcohol or over-the-counter products containing acetaminophen. Either of these situations can lead to the possibility of hepatic toxicity through toxic metabolites.

**Nonsteroidal Anti-inflammatory Drugs.** NSAIDs are the most commonly prescribed medications for the treatment of OA. Nonselective NSAIDs work through nonspecific inhibition of cyclooxygenase isoforms 1 and 2 (COX-1 and COX-2). COX-1 is constitutively expressed in renal and gastrointestinal (GI) tissues, among others. COX-2 is inducible in inflammatory responses. The major side effects of NSAIDs are GI toxicities (gastritis, peptic ulcer disease) and renal toxicities (interstitial nephritis, prostaglandin inhibition–related renal insufficiency). Because GI tissues have a higher expression of COX-1, a selective COX-2 inhibitor might spare patients the GI side effects. Unfortunately, COX-2 is expressed in renal tissue, and COX-2–specific drugs such as traditional NSAIDs have potential adverse renal effects. This is especially true in those with baseline renal insufficiency. Concerns about cardiovascular risks led to the voluntary withdrawal of rofecoxib from the market in the United States. There have also been concerns about celecoxib at a dose of 200 mg twice daily, owing to an increased relative risk for myocardial infarction in an adenomatous polyp trial; this, however, has not been confirmed in six observational studies. All NSAIDs and COX-2–specific agents have received “black box” warnings in their package inserts addressing cardiovascular risk. Alternative mechanisms of action of NSAIDs such as interference with receptors in the cell membrane phospholipid bilayers have been proposed.

Nonselective NSAIDs are widely used for the management of OA. They include ibuprofen, naproxen, and diclofenac. NSAIDs are usually analgesic at lower doses but have both analgesic and anti-inflammatory effects at their higher recommended doses. They are prescribed either in fixed doses or “as needed” and are quite effective as symptom modifiers; however, they have no structure- or disease-modifying effects. NSAIDs should be used in the smallest dose that provides satisfactory symptom relief because GI toxicity has been linked to dosage. Adverse GI events have also been linked to patient age, previous history of peptic ulcers or bleeding, and the presence of comorbid conditions such as heart disease.

To reduce the potential for adverse GI events, misoprostol can be added to the therapeutic regimen. It is a prostaglandin E analogue that has been shown to reduce the GI side effects of NSAIDs when used at 200 µg three times a day. Diarrhea is a potential side effect. The use of a concomitant proton pump inhibitor may reduce upper GI endoscopic ulceration rates from NSAIDs, although no study has attempted to show a decrease in events such as symptomatic ulcers or bleeds. Over-the-counter doses of H2 blockers and antacids have not been shown to reduce either endoscopic or serious clinical GI events. COX-2–specific inhibitors are the latest drugs used in an attempt to reduce the GI adverse event profile of OA therapy.

COX-2–specific inhibitors are highly selective for COX-2 in vitro. Currently, only one such agent is available in the United States, celecoxib. Others such as etoricoxib are available elsewhere. These agents have been shown to reduce the rate of endoscopic ulceration by more than 50% when compared with nonselective NSAIDs. Celecoxib significantly reduces the rates of symptomatic ulcers, bleeds, perforations, and obstructions in patients not concurrently on aspirin. It remains unclear how substantial the gastrointestinal benefits of these compounds are to patients taking aspirin. Because COX-2–specific agents can inhibit endothelial prostacyclin but do not affect platelet thromboxane, cardiovascular safety remains an area of investigation.

Combination COX-lipoxygenase inhibitors are still investigational. It remains to be seen how these will compare with traditional NSAIDs and with COX-2 inhibitors in terms of both safety and efficacy. Animal studies have hinted at the possibility of a structure- and disease-modifying effect.

**Narcotic Analgesics.** Although several options exist for the management of pain in OA, some patients obtain suboptimal pain relief. If a patient has failed to respond to other nonpharmacologic and pharmacologic modalities and has no additional identifiable causes of pain (such as fibromyalgia), a narcotic analgesic should be considered.

The pain of OA is generally responsive to narcotic analgesics. Because of concerns about potential addiction, appropriate patient selection is important. Narcotic analgesics such as codeine and propoxyphene have been used effectively in patients with OA, especially in combination with non-narcotic analgesics (e.g., acetaminophen). Potential side effects include nausea, constipation, and somnolence.
Tramadol is an oral medication with mild suppressive effects on the mu opioid receptor. It also inhibits the uptake of norepinephrine and serotonin and is not thought to have significant addictive tendencies. It is available alone or in combination with acetaminophen and is not a controlled-schedule medication in the United States. Tramadol has been used for the symptomatic relief of OA. Seizures and allergic reactions are potential side effects. A warning of increased risk of suicide in certain patients, similar to that on antidepressants, has been added to the label. The incidence of nausea can be reduced by slowly escalating the dose until the desired pain relief is achieved.

One study compared tramadol and acetaminophen with the combination of codeine and acetaminophen. Patients with OA or chronic low back pain were randomized to receive tramadol and acetaminophen (37.5 mg and 325 mg, respectively) or codeine and acetaminophen (30 mg and 300 mg, respectively) for 4 weeks. Pain relief and changes in pain intensity were equivalent in both groups. Those on codeine and acetaminophen had a significantly higher incidence of somnolence (24% vs. 17%) and constipation (21% vs. 11%). The tramadol-acetaminophen combination also provides symptomatic relief as add-on therapy in OA patients receiving NSAIDs or COX-2 agents as baseline therapy.

Extended-release narcotic analgesics have been tested in clinical trials in OA. This approach is intended to achieve a lower level of peak-to-trough variability in the plasma concentration of the narcotic. An extended-release, once-a-day preparation of tramadol relieves pain in OA of the knee and hip. Extended-release oxymorphone dosed twice a day also provides relief in those with moderate to severe pain from OA of the hip or knee, as demonstrated by a VAS and the WOMAC composite index, as well as the subscales for pain, stiffness, and physical function.

Transdermal fentanyl, a narcotic analgesic, has been used in the treatment of moderately to severely symptomatic knee and hip OA. It relieved pain and improved function in clinical trials as judged by a VAS and the WOMAC physical function subscale.

**Intra-articular Agents**

**Corticosteroids.** Although there is no role for systemic corticosteroids in OA, local intra-articular corticoid preparations have a long history in the management of OA. Corticosteroids have been shown to downregulate the expression of adhesion molecules. This, in turn, can reduce cellular infiltration into the joint and subsequent inflammation. Corticosteroid injections slow macrophage-like cell infiltration of the synovium in OA. The dose of steroid injected is determined by the volume of the joint being injected, with larger joints such as the knee receiving higher doses. The risk of joint infection is low if proper technique is employed. Postinjection flares due to corticosteroid crystal synovitis can occur.

There is a relative dearth of information from clinical trials of intra-articular corticosteroid injections. However, in one study, symptomatic benefit from corticosteroid injection for OA of the knee was demonstrated in a double-blind trial at 1 and 4 weeks postinjection. Another trial attempted to assess the possible disease-modifying effects of corticosteroids by randomizing 68 patients to corticosteroid or saline injections of the knee every 3 months for 2 years. At the study’s end, there was no significant difference in rate of joint space narrowing; thus no case could be made for a disease-modifying effect of corticosteroid injections. There was a trend favoring pain relief in the corticosteroid group as measured by the pain subscale of the WOMAC. A review of published studies of intra-articular corticosteroid injections in OA concluded that the short-term symptomatic benefits have been well established, with few adverse events, but long-term benefits have not been confirmed.

The specific corticosteroid compound used, the frequency of injections, and other factors related to the use of corticosteroid injections in OA vary widely and are heavily influenced by the training program the rheumatologist attended and where he or she practices. In general, corticosteroid injections are believed to be most effective in patients with evidence of inflammation, effusions, or both. Because of concerns over possible deleterious effects, usually no more than four corticosteroid injections per year are given in a particular joint. Further discussion of arthrocentesis can be found in Chapter 54.

**Hyaluronic Acid Derivatives.** Synthetic and naturally occurring hyaluronic acid derivatives are administered intra-articularly. Although often mentioned as potential structure-modifying agents, these products are presently considered symptom-modifying drugs. Their molecular weights vary (from <100,000 to >1 million Svedberg units), depending on the preparation. They reportedly reduce pain for prolonged periods and may improve mobility. Improvement in overall physical functioning has also been reported. The mechanisms of action are not known. However, there is evidence of an anti-inflammatory effect (particularly at high molecular weight), a short-term lubricant effect, an analgesic effect by direct buffering of synovial nerve endings, and a stimulating effect on synovial lining cells, leading to the production of normal hyaluronic acid.

Several preparations have been approved in the United States for OA of the knee. The injection regimens vary from five weekly injections to one injection depending on the product selected by the clinician for use. Pain relief has been the primary outcome in these studies with relief at 26 weeks postinjection for some. In one study, three weekly hyaluronic acid intra-articular injections provided comparable pain relief to a single corticosteroid intra-articular injection at 1-week follow-up; at 45 days’ follow-up, hyaluronic acid was superior to the corticosteroid. In a Canadian study, 102 patients with OA of the knee were randomized to three weekly intra-articular injections of hylan G-F (Synvisc), hylan G-F plus an NSAID, or NSAID alone. At 26 weeks, both groups receiving hylan G-F were significantly better than the group receiving NSAIDs alone.

Substantial clinical responses to the saline injections used as placebo in hyaluronic acid trials have sometimes made data interpretation challenging. In a double-blind, placebo-controlled trial, 495 patients with knee OA were randomized to receive five intra-articular injections of hyaluronic acid (Hyalgan) given 1 week apart, placebo, or naproxen (500 mg orally twice a day) and followed for 26
NUTRACEUTICALS

Two nutritional supplements—glucosamine and chondroitin sulfate—have received significant attention (Table 100-3). Health food stores and the lay press rather dubiously proposed them as “cures for arthritis.” The mechanism of action of glucosamine sulfate is uncertain. Some in vitro experiments have shown stimulation of the synthesis of cartilage glycosaminoglycans and proteoglycans. Others have shown that glucosamine and N-acetylglucosamine inhibit interleukin (IL)-1β— and tumor necrosis factor (TNF)—induced nitric oxide production in normal human articular chondrocytes. N-acetylglucosamine also suppresses the production of IL-1β and stimulates IL-6 and COX-2.

GLUCOSAMINE

Urinary excretion of glucosamine (and other glycosaminoglycans) has been investigated and found to be elevated in both OA and rheumatoid arthritis. Supplementation with glucosamine sulfate, an intermediate in mucopolysaccharide synthesis, has been tried both orally and intramuscularly as therapy for OA. Glucosamine sulfate (400 mg injected intramuscularly twice weekly for 6 weeks) reduced the severity of disease as judged by the Lequesne index when compared with placebo. A randomized, double-blind, parallel-group study in knee OA compared 500 mg oral glucosamine sulfate three times a day with 400 mg ibuprofen three times a day for 4 weeks. The response to ibuprofen was more rapid, but at 4 weeks, there was no statistically significant difference in the response rate (reduction of at least 2 points in the Lequesne index). No group in the study received higher, anti-inflammatory doses of ibuprofen. An NIH-sponsored trial is currently under way in the United States to more thoroughly study the symptom-relieving and possible structure-modifying properties of glucosamine. Meanwhile, some already advocate the use of glucosamine as part of the first line of therapy for symptomatic OA.

Glucosamine has also been compared with acetylamino-phen in an industry-sponsored trial. In the GUIDE trial, 318 patients with knee OA were randomized to glucosamine sulfate soluble powder 1500 mg once a day, acetaminophen 1000 mg three times a day, or placebo for 6 months. The main efficacy parameter was the 6-month change in the Lequesne index. At 6 months, the glucosamine group achieved significantly greater efficacy versus placebo. Those on acetylamino-phen failed to achieve a statistically significant benefit versus placebo by either the Lequesne index or WOMAC. There was no statistically significant difference between those on glucosamine and those on placebo on the basis of WOMAC outcomes.

Another clinical trial randomized 80 patients with knee OA to either glucosamine sulfate 1500 mg/day or placebo for 6 months. There was no difference between glucosamine and placebo in the primary variable of patients’ global assessment of pain in the affected knee. Another trial used a unique Internet-based recruiting system and followed 205 patients with knee OA randomized to glucosamine sulfate 1500 mg/day or placebo for 12 weeks. The primary endpoint was the pain subscale of the WOMAC. At study conclusion, there was no difference in the groups with regard to pain, physical function, or overall WOMAC scores. Stratification by severity of OA, glucosamine product used, or use of NSAIDs did not alter the results. The Cochrane review of glucosamine therapy in OA analyzed a pool of 20 studies and 2570 patients. Pain and function improved by 28% and 21%, respectively, by the Lequesne index, compared with placebo. There was no improvement in the overall WOMAC pain and function scales. There has been speculation that these inconsistencies in study results may be due to a lack of standardization in glucosamine preparations.

A recent discontinuation trial has added to the uncertainty about glucosamine’s efficacy. It found that 137 patients who had been clinically classified as moderate responders to glucosamine sulfate were equally likely to experience an OA flare whether they continued or discontinued the glucosamine. No statistically significant differences between the groups were noted in pain and WOMAC function scores after 6 months.

Table 100-3 NUTRACEUTICALS FOR OSTEOARTHRITIS

<table>
<thead>
<tr>
<th>Nutraceuticals for Osteoarthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucosamine</td>
</tr>
<tr>
<td>Chondroitin sulfate</td>
</tr>
<tr>
<td>Ginger extracts</td>
</tr>
<tr>
<td>Avocado and soy unsaponifiables</td>
</tr>
<tr>
<td>Cat’s claw</td>
</tr>
<tr>
<td>Shark cartilage</td>
</tr>
<tr>
<td>S-adenosyl methionine</td>
</tr>
</tbody>
</table>
Combination products containing both glucosamine and chondroitin have become popular in the United States, despite a dearth of clinical trial data. One small, placebo-controlled trial randomized patients with knee OA to receive a regimen of glucosamine hydrochloride (1000 mg), chondroitin sulfate (800 mg), and manganese ascorbate (152 mg) twice a day or placebo. Patients were evaluated at baseline and then every 2 months for 6 months using the Lequesne index of OA severity. At 4 and 6 months, those with mild to moderate radiographic OA of the knee showed significant improvement by the Lequesne index compared with those on placebo. In those with severe radiographic OA of the knee, no significant symptomatic benefit could be demonstrated. The study did not evaluate patients for structure or disease modification.

In the Glucosamine/Chondroitin Arthritis Intervention Trial (GAIT), 1583 patients with OA of the knee were randomized to placebo, glucosamine hydrochloride 1500 mg/day, chondroitin sulfate 1200 mg/day, celecoxib 200 mg/day, or glucosamine hydrochloride and chondroitin sulfate. The primary endpoint was the percentage of patients achieving at least 20% improvement on the WOMAC pain subscale at 6 months. The only statistically significant response was seen in those on celecoxib versus placebo (70.1% vs. 60.1%; P = .008). Patients were then stratified for baseline severity by WOMAC pain scores, most of them falling into the mild OA pain category. In a subgroup analysis, in those with moderate to severe OA pain (WOMAC pain, 301 to 400 mm), the combination of glucosamine hydrochloride and chondroitin sulfate was more efficacious than placebo as measured by a dichotomous response rate (positive = 50% improvement in pain): 79.2% versus 54.3% (P = .002). Analysis of the radiographic data from this trial failed to support the notion of a disease-modifying/slowing role for glucosamine, chondroitin sulfate, or the combination of these compounds.

From these results, it appears that patient selection may be important in maximizing any potential benefit from glucosamine or chondroitin therapy. The GAIT study also had a particularly high placebo response rate, which may reflect the enrollment of patients with less symptomatic OA and may have affected the results. It also used a glucosamine hydrochloride preparation instead of the glucosamine sulfate used in most other studies, particularly those that have demonstrated efficacy. This raises the question of whether the choice of glucosamine hydrochloride negatively affected efficacy in the trial. However, one small (142 patients) Chinese trial randomized patients with OA of the knee to glucosamine sulfate 1500 mg/day or glucosamine hydrochloride 1440 mg/day for 1 month. No efficacy differences were noted, with a clear majority of patients in each treatment arm achieving symptomatic improvement by Lequesne scores. The study had no placebo arm. Safety assessments continued for 2 additional weeks, with no significant adverse events reported. At present, it is still unclear whether glucosamine hydrochloride preparations have the same potential clinical benefits as glucosamine sulfate preparations. Additional investigations are necessary.

Two European trials tried to address the subject of disease modification with glucosamine. In one study, 212 patients with OA of the knee were randomized to receive placebo or glucosamine sulfate (1500 mg/day) and were followed prospectively for 3 years. Fluoroscopically positioned, standing anteroposterior radiographs of the knees were taken at enrollment, 1 year, and 3 years. At 3 years, the treatment group had a joint space reduction of 0.06 mm, whereas the placebo group had a reduction of 0.31 mm. Whether this is a clinically meaningful difference in joint space is unclear. Those taking glucosamine also showed symptomatic benefit on the order of 20% to 25%, whereas those taking placebo had a slight worsening of symptoms, as judged by the WOMAC. There were no significant adverse events attributed to the use of the glucosamine sulfate. A second group of researchers randomized 202 patients to receive placebo or glucosamine sulfate (1500 mg/day) for 3 years. The width of the narrowest medial joint space of the tibiofemoral joint was measured serially, using visual assessments with a 0.1-mm graduated magnifying glass on standardized full-extension, weight-bearing anteroposterior radiographs of each knee. At 3 years, there was a significant difference in joint space width, with a decrease of 0.19 mm in the placebo group and an increase of 0.04 mm in the glucosamine sulfate group. Also, significantly greater improvements in the WOMAC score and the Lequesne index were seen in the glucosamine group. The favorable results of these studies have been questioned because of the radiographic technique used to assess joint space. At issue is whether the joint space seen on standing films of the knee might be significantly affected by the symptoms of OA (i.e., pain) and whether a semiflexed film would be preferable. In one study, investigators obtained baseline radiographs (after analgesic or NSAID washout) using both standing-extended and semiflexed, fluoroscopically positioned techniques in 19 patients with knee OA. Radiographs were then repeated 2 to 8 weeks later after reinstitution of analgesic or NSAID therapy. Joint space width increased with effective pain relief in highly symptomatic patients if measured by standing-extended radiographs. Using the semiflexed technique, there were no significant changes in joint space width related to severity of pain or responsiveness to pain therapy. This suggests that data obtained using the standing-extended radiographic technique may need to be revisited because the results may represent a therapeutic intervention’s effect on symptoms (pain) rather than a disease-modifying effect. More recent, ongoing trials have changed to the semiflexed, fluoroscopically positioned knee radiograph to assess potential disease modification.

### Chondroitin Sulfate

Oral chondroitin sulfate, a glycosaminoglycan composed of units of glucosamine with attached sugar molecules (molecular mass of around 14,000), has also been used as therapy for hip and knee OA. Its mechanism of action is unknown. A double-blind, placebo-controlled study included a 3-month treatment phase followed by a 2-month treatment-free phase. The major outcome parameter was NSAID consumption. Those receiving chondroitin sulfate used fewer NSAIDs than the controls both at the completion of treatment and in the treatment-free phase. Another study compared chondroitin sulfate to diclofenac sodium. One group received chondroitin sulfate (400 mg three times a day) and the other diclofenac sodium (50 mg three times a day).
One study evaluated chondroitin sulfate as a disease-modifying intervention. Three hundred patients were enrolled and randomized to chondroitin sulfate 800 mg daily or placebo for 2 years. Joint space width was assessed using anteroposterior semiflexed radiographs. Pain and function were assessed as secondary endpoints. In the placebo group, the mean change in joint space width was 0.07 mm/year, while in the treatment group, the mean change was 0.00. A similar difference was noted when the minimum joint space width was evaluated. The differences were statistically significant (mean joint space width, \( P = .04 \); minimum joint space width, \( P = .05 \)), but the clinical relevance remains unclear. The changes in radiographic progression were not matched by similar differences when pain and function were analyzed. The treatment group achieved improvement in all WOMAC subscales of pain, function, and stiffness, but a statistically significant difference could not be shown. It has been suggested that the overall low baseline WOMAC scores created difficulties in assessing for clinical improvement.

**Other Nutraceuticals**

Ginger extracts have been popular “natural” remedies for OA for some time. Most of the world’s ginger comes from China, and its “medicinal” use dates back more than 2000 years. Ginger actually contains small amounts of salicylate. In some animal models, ginger has been shown to have inhibitory effects on COX and lipoxygenase. One study with 247 evaluable patients revealed a small but statistically significant reduction in knee pain on standing (63% vs. 50%; \( P = .048 \)) after taking ginger. Reduction in knee pain after a 50-foot walk was also significant. Use of acetaminophen was reduced in the ginger-extract group, but the difference was not statistically significant. The extract was well tolerated, except for GI events such as dyspepsia, nausea, and eructation, which were increased over placebo. The question remains whether benefits observed represent a clinically relevant effect.

Some of the more unusual agents proposed as structure- or disease modifiers in OA are oral preparations of avocado and soy unsaponifiables (ASUs). These compounds are derived from unsaponifiable residues of avocado and soya oils mixed in a 1:2 ratio. In vitro studies on cultured chondrocytes showed partial reversal of IL-1\( \beta \) effects. The roles of IL-1\( \beta \) in OA are thought to include inhibition of prostaglandin synthesis by chondrocytes and stimulation of matrix metalloproteinases (MMPs) and nitric oxide production. MMPs and nitric oxide can degrade cartilage matrix and cause chondrocyte apoptosis. Use of ASUs also reportedly results in inhibited production of IL-6, IL-8, and MMPs and stimulation of collagen synthesis. Increased aggrecan synthesis has been reported as well. The mechanism of action is unknown, as is the active ingredient in ASUs. Symptomatic benefit in double-blind human trials in OA of the hip and knee has been reported. However, a double-blind, placebo-controlled trial in OA of the hip failed to show disease modification in the overall population, although a post hoc analysis reported benefit in those with more advanced OA at baseline. Some abstract presentations have suggested structure or disease modification in human hip OA.

Other nutritional supplements such as cat’s claw and shark cartilage have become entrenched in regional and international popular cultures. Many people take them, despite limited or no data to support their use. A small, placebo-controlled trial showed improvement of OA pain with activity in those taking cat’s claw extracts. Shark cartilage contains a small amount of chondroitin sulfate. S-adenosyl methionine (SAMe), a methyl group donor and oxygen radical scavenger, is often touted as a remedy for OA, although little evidence of its effectiveness has been published. In one double-blind, placebo-controlled study, two centers reported differing results. One center reported reductions in overall pain and rest pain, whereas the other showed no significant difference between the test group and placebo group. Another small, double-blind, placebo-controlled crossover study of 61 patients compared oral SAMe 1200 mg/day with oral celecoxib 200 mg/day for 16 weeks. After the first month of phase I, celecoxib provided superior pain relief that was statistically significant. By the end of the second month, however, there was no statistically significant difference between the groups. There is insufficient evidence to recommend the use of these products in the treatment of OA.

**Other Potential Structure- or Disease-Modifying Therapies**

The term *chondroprotective* has been used to describe structure- or disease-modifying agents. This is a misnomer, however, because the goal is to protect the entire joint (not only the cartilage) from the arthritic process. A workshop of the Osteoarthritis Research Society recommended that the term *structure-modifying drugs* be used for medications that previously would have been classified as chondroprotective. These drugs are intended to prevent, retard, stabilize, or even reverse the development of OA. Recently, the term disease-modifying osteoarthritis drug (DMOAD) has been used for any such agent (*Table 100-4*). Such a disease-modifying effect in OA would require prolonged observation, given the typically slow progression of OA. Therefore clinical trials in this area have been challenging, with most being designed for at least 2\( \frac{1}{2} \) to 3 years of follow-up. Progress in the methodology used to assess structure and disease modification may shorten the length of these trials. Radiographic assessments of joint space such as fluoroscopically positioned anteroposterior radiographs of the knee or magnetic resonance imaging may be useful in this regard.

Unfortunately, to date, no drug has been conclusively proved to be structure or disease modifying. Although this chapter focuses on medication-based therapies, other
Disease-modifying effect was demonstrated in the target knee, no such effect could be demonstrated in the contralateral knee. Thus the implications of these findings for clinical practice are uncertain. Other compounds with collagenase-inhibiting properties are being developed and investigated as structure- or disease-modifying agents not only in OA but also in rheumatoid arthritis.147 Glycosaminoglycan polysulfuric acid (GAGPS; Artesaron or Adequan) has been purported to work by reducing the activity of collagenase. It is a highly sulfated glycosaminoglycan, with a molecular weight ranging from 2000 to 16,000,148 derived from bovine tracheal cartilage. In a canine model of OA, GAGPS was administered intra-articularly twice weekly for 4 weeks.149 Four weeks after completion of the GAGPS treatment, medial femoral condylar lesions had developed to a lesser degree in the treated group than in saline-treated dogs. Swelling, an indicator of collagen network integrity, remained near control levels in the treatment group. In humans, OA of the knee was studied in a 5-year trial. There was improvement in multiple measured parameters including less time lost from work.150 Another double-blind, placebo-controlled trial evaluated GAGPS in 80 patients with OA of the knee; patients received two series of five intra-articular injections of 25 mg (0.5 mL) GAGPS at 1-week intervals. At 14 weeks, 31% of the GAGPS group had improvement as judged by the Lequesne index, compared with 15% in the placebo group.151 Potential allergy and heparin-like effects were observed. GAGPS is available in the United States for equine, but not human, use.

Another extract, a glycosaminoglycan-peptide complex (GP-C) known as Ramalol, has been investigated. It is a highly sulfated polysaccharide derived from bovine tracheal cartilage and bone marrow and is administered intramuscularly.152 It has been shown to increase the levels of tissue inhibitor of metalloproteinases (TIMPs).153 A randomized, placebo-controlled trial selected patients with hip or knee OA to receive 10 courses of injections of placebo or GP-C (2 mL) over 5 years (two courses per year). Each course consisted of 15 injections given twice weekly. GP-C failed to demonstrate a structure- or disease-modifying effect.154 In addition, there were no statistical differences favoring the active treatment group when measured by the Lequesne index, pain on passive motion, or consumption of NSAIDs. GP-C is available in parts of Europe and South America.

Pentosan polysulfate (Cartrofen) is a purified extract of beech hemicellulose administered intramuscularly or orally as a calcium salt. It can inhibit granulocyte elastase and has inhibited the catabolism of aggrecan in cartilage explants.155 It has been shown to increase the levels of tissue inhibitor of metalloproteinases (TIMPs).153 A randomized, placebo-controlled trial selected patients with hip or knee OA to receive 10 courses of injections of placebo or GP-C (2 mL) over 5 years (two courses per year). Each course consisted of 15 injections given twice weekly. GP-C failed to demonstrate a structure- or disease-modifying effect.154 In addition, there were no statistical differences favoring the active treatment group when measured by the Lequesne index, pain on passive motion, or consumption of NSAIDs. GP-C is available in parts of Europe and South America.

Table 100-4  Potential Structure- and Disease-Modifying Drugs in Osteoarthritis

<table>
<thead>
<tr>
<th>Tetracyclines</th>
<th>Metalloproteinase or collagenase inhibitors</th>
<th>Glucosamine</th>
<th>Diacerein</th>
<th>Growth factor and cytokine manipulation (IL-1Ra, TGF-β)</th>
<th>Gene therapy (IL-1Ra, IL-1RII)</th>
<th>Chondrocyte and stem cell transplantation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL-1Ra, interleukin-1 receptor antagonist; IL-1RII, interleukin-1 receptor type II; TGF-β, transforming growth factor-beta.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
comparable efficacy to NSAIDs but a slower onset of action. Diarrhea is the main potential side effect. On the strength of these prior trials, diacerein has been proposed as a slow-acting symptom-modifying and perhaps structure- or disease-modifying drug for OA.

A double-blind, randomized, placebo-controlled trial looking at the efficacy and safety of diacerein enrolled 484 patients with symptomatic knee OA. They were randomized to receive placebo, diacerein 25 mg twice a day, diacerein 50 mg twice a day, or diacerein 75 mg twice a day. Using intent-to-treat analysis, diacerein 100 mg/day was significantly superior to placebo \((P < .05)\) by the primary endpoint—patients’ assessment of pain on movement at week 24 \((-18.3 \pm 19.3 \text{ mm vs. } -10.9 \pm 19.3 \text{ mm})\). It was also superior on the basis of WOMAC and disability scores. However, no statistical difference was detected in the primary endpoint between placebo and 50 mg/day diacerein \((-15 \pm 21.0 \text{ mm})\) or 150 mg/day diacerein \((-14.3 \pm 23.7 \text{ mm})\).

There have also been investigations into the potential structure- or disease-modifying attributes of diacerein in OA. In one study, 507 patients with OA of the hip (according to American College of Rheumatology criteria) were randomized to receive either diacerein \((50 \text{ mg orally twice a day})\) or placebo for 3 years. Patients were followed with yearly pelvic radiographs to assess hip joint space. Using completer analysis, the diacerein patients showed a significantly lower rate of radiographic progression \((0.18 \text{ vs. } 0.23 \text{ mm/year})\). Using intent-to-treat analysis, a smaller proportion of those taking diacerein had significant joint space loss \((\text{defined as loss of } \geq 0.5 \text{ mm})\) during the study \((50.7\% \text{ vs. } 60.4\%)\). Unfortunately, almost 50% of the patients failed to complete the 3-year study. In the placebo group, the principal reason for discontinuation was lack of efficacy, whereas in the diacerein group, it was adverse effects such as diarrhea. Curiously, the symptom-relieving effect of diacerein observed in prior studies could not be confirmed in this one. A recent meta-analysis of clinical trials of diacerein in OA concluded that available clinical evidence supports pain relief in hip and knee OA. There was no analysis of a disease-modifying effect.

Potential methods of intervention in OA include growth factor and cytokine manipulation. Cytokines such as IL-1 and TNF are produced by the synovium and contribute to inflammation within osteoarthritic joints. Moreover, there may be deficient expression of naturally occurring anti-inflammatory compounds such as IL-1 receptor antagonist \((\text{IL-1Ra})\) by the chondrocytes of patients with OA. In some cases, increased nitric oxide production by OA articular chondrocytes may inhibit IL-1Ra synthesis. In a dog model of OA, IL-1Ra therapy reduced the expression of collagenase-1 in cartilage. The severity of cartilage lesions is also diminished. In a rabbit model of OA, transfer of the IL-1Ra gene to joints prevented OA progression. The effect of IL-1 blockade in humans with OA through the use of IL-1Ra is currently being investigated. Induction of repair in partial-thickness articular cartilage lesions by the timed release of transforming growth factor-\(\beta\) using liposomes has been attempted in an animal model. There was an increase in the cellularity of the defects, which were populated by cells of mesenchymal origin from the synovial membrane. The repaired cartilage resembled hyaline cartilage, and its integrity persisted up to 1 year after surgery. Combination therapy is another alternative. In a study of canine-induced OA, sodium pentosan polysulfate, when combined with insulin-like growth factor-I, reduced stromelysin activity and increased TIMP.

Initial attempts at gene therapy are intriguing. The control of genes such as TIMP and MMPs would, in theory, provide the opportunity to modulate the patient’s disease. As previously noted, gene expression of IL-1Ra has already been tried in rabbits and dogs, as well as in an equine model of OA using an adenovirus vector. Use of gene transfer-mediated overexpression of IL-1B decoy receptor has also been contemplated. Chondrocyte and stem cell transplants into articular cartilage defects have been tried as well. Chondrocytes transplanted \((\text{expressing a previously transfected } \beta\text{-galactosidase gene})\) into human cartilage explants survived up to 45 days in vitro in one trial. Transfection of chondrocytes with the galactosidase gene has been successful both before and after transplantation.

**Surgical Intervention**

Surgical interventions in OA usually consist of osteotomies or joint replacements. Osteotomies can be effective pain-relieving interventions and can delay the need for joint replacement surgery in selected patients. These tend to be younger subjects with OA.

Joint replacement surgery \((\text{joint arthroplasty})\) is effective in providing pain relief and restoring function in many patients with OA. Hip and knee joint replacements are most common. Indications for surgery include pain that is refractory to the previously discussed interventions and significant impairment of the patient’s daily life. Therefore patients should be the key decision makers because they are the ones who must weigh the severity of symptoms and impairment. Patients undergoing replacement surgery should be deemed able to undertake the rehabilitation necessary to regain reasonable use of the joint involved. Infections are rare but do occur. Joint replacements have a typical life span of between 10 and 15 years. Revision surgery may be necessary, particularly in a relatively young patient who outlives the useful life of the prosthesis.

Other potential rationales for surgical intervention in OA include removal of loose bodies, stabilization of joints, redistribution of joint forces \((\text{e.g., osteotomy}), \text{and relief of neural impingement (e.g., spinal stenosis, herniated disk).}\) The value of arthroscopic débridement or lavage in OA has been questioned. A recent randomized, blinded trial failed to demonstrate significant symptomatic benefit in OA of the knee.

**SUMMARY**

The treatment of OA includes a variety of possible non-pharmacologic and pharmacologic interventions. Treatment should be tailored to the individual and consists of a combination of modalities. These provide symptom relief but have no proven effect on the progression of disease. Structure and disease modification has yet to be achieved in OA. Trials that are under way could determine whether this is a realistic goal. Claims of structure or disease modification in OA should not be made for any drugs until
well-designed, double-blind, placebo-controlled trials demonstrate that this is so. It is hoped that with the eventual advent of disease-modifying OA drugs, treatment will eventually consist of a combination of symptom-relieving and disease-modifying interventions.

Selected References


Full references for this chapter can be found on www.expertconsult.com.
References


Osteoporosis is characterized by low bone density and a deterioration of bone microarchitecture, which reduces bone strength and increases fracture risk. The hallmark of osteoporosis is the loss of bone mineral and bone matrix that results in maintenance of a normal mineral-to-matrix ratio. Bone consists of an organic matrix (collagen and noncollagenous proteins) and an inorganic mineral component (calcium and phosphate in hydroxyapatite crystals; see Chapter 4). Normally, bone turnover is tightly coupled with osteoclast-mediated bone resorption followed by osteoblast-stimulated bone formation. This delicate balance in bone remodeling results in no net change in skeletal mass. Osteoblasts synthesize osteoid—bone matrix that subsequently undergoes mineralization and becomes mature bone matrix. The skeleton contains approximately 80% cortical bone, which is concentrated in the appendicular skeleton and femoral neck, and 20% more metabolically active trabecular bone, which is located in the spine, epiphyses, and pelvis.

Osteoporosis is characterized by reduced bone strength usually accompanied by a reduction in bone mass. Osteomalacia encompasses disorders in which there is decreased mineralization of bone matrix. Paget’s disease is a skeletal disorder characterized by increased rates of bone turnover usually accompanied by a reduction in bone mass. Osteomalacia encompasses disorders in which there is decreased mineralization of bone matrix. Paget’s disease is a skeletal disorder characterized by increased rates of bone turnover with the development of disorganized woven bone.

OSTEOPOROSIS

Epidemiology and Clinical Signs

Osteoporosis, the most common metabolic bone disease, affects 200 million individuals worldwide. Approximately 28 million Americans have osteoporosis or are at risk for it. Osteoporosis, or “porous bone,” is a “disease characterized by low bone mass and structural deterioration of bone tissue, leading to bone fragility and an increased susceptibility to fractures, especially of the hip, spine and wrist.” Although usually asymptomatic, osteoporosis can produce loss of height, pain, dowager’s hump, and increased risk of fracture. After 50 years of age, there is an exponential rise in fractures, such that 40% of women and 13% of men develop one or more osteoporotic fractures in their lifetimes. In the United States alone, there are more than 1.5 million osteoporotic fractures annually including 250,000 hip, 250,000 wrist, and 500,000 vertebral fractures. Hip fractures are associated with a 12% to 24% mortality rate in women and a 30% mortality rate in men within the first year of fracture, and 50% of patients are unable to ambulate independently and require long-term nursing home care. These numbers will continue to grow exponentially as the elderly population of industrialized nations increases.

KEY POINTS

Osteoporosis is a disease defined by low bone density and deterioration of microarchitecture, which reduces bone strength and increases fracture risk.

Major clinical risk factors for osteoporotic fractures include older age, low weight, family history of hip fracture, fracture occurring after age 50, glucocorticoid use, and inability to rise from a chair without assistance. More than 50% of osteoporosis in men results from secondary causes.

Postmenopausal and age-related bone loss results from an uncoupling of bone remodeling such that bone resorption is greater than bone formation, resulting in a net loss of bone.

Polymorphisms in antagonists of the wnt/B catenin signaling pathway that result in a gain of function (e.g., LRP5) are associated with a reduced risk of osteoporosis.

Biochemical markers measured in the serum including C and N telopeptide cross-links of type I collagen correlate with osteoclast activity on the bone surface.

Evaluation for risk of osteoporotic fractures with the Fracture Assessment Index (FRAX) for postmenopausal women and men is critical to determine individuals with high enough 10-year fracture risk (hip and major osteoporotic fracture sites) to warrant treatment.

Treatment of high-turnover osteoporosis from estrogen deficiency with antiresorptive agents (estrogen, raloxifene, and bisphosphonates—alendronate, risedronate, zoledronic acid, ibandronate, denosumab) and an anabolic agent (recombinant human parathyroid hormone 1-34) can reduce incident vertebral fractures.

Parathyroid hormone (PTH) increases osteoblast maturation and life span, increases trabecular bone mass and cortical thickness, improves bone strength, and decreases fractures. Antiresporptive therapy is necessary after a full course of PTH to maintain newly formed bone mass.

Glucocorticoid-induced bone loss results from increased osteoclast activity and reduced osteoblast activity and is most severe in the first 6 months of therapy. Bisphosphonate treatment can prevent fractures.

Aromatase inhibitors reduce serum estrogen and result in rapid bone loss in postmenopausal women on adjuvant breast cancer therapy.

Gonadotropin-releasing hormone agonists decrease testosterone and estrogen levels and cause bone loss in men being treated for prostate cancer.
Bone accretion occurs during adolescence, when there is a large increment in bone mass. Peak bone density is normally achieved after puberty and into the third decade of life. However, by age 22, most individuals have achieved nearly all of their peak bone mass. At menopause, an acceleration of bone loss typically occurs over approximately 5 to 8 years, with an annual 2% to 3% loss of trabecular bone and a 1% to 2% loss of cortical bone. Both men and women lose bone with age. Over a lifetime, women lose approximately 50% of trabecular and 30% of cortical bone; men generally lose two-thirds of these amounts. Osteoporosis was previously thought to be a silent disease that was part of the normal aging process. However, the advent of bone densitometry has made it possible to accurately and reproducibly identify patients at risk for osteoporosis so that prevention and treatment strategies can be instituted to reduce fractures. With a health care expenditure of $13.8 billion annually for osteoporosis-related fractures and a projected threefold rise in these costs over the next 40 years in the United States, the institution of effective prevention and treatment strategies to reduce fractures is of great importance.\textsuperscript{1,4}

**Pathophysiology of Menopausal and Age-Related Bone Loss**

Bone is constantly undergoing remodeling, whereby areas of bone resorption produced by osteoclastic action are replaced by bone laid down by osteoblasts. Osteoporosis results from an imbalance between bone resorption and formation. The initiation of bone remodeling is still being debated; however, the osteocytes, or terminally differentiated osteoclasts, located within the bone matrix and connected to one another and the bone surface may release chemical mediators that attract osteoclasts to the bone surface (Figure 101-1). Osteoclasts originate from the colony-forming unit granulocyte-monocytes, are attracted to the bone surface, attach to bone matrix, and resorb bone tissue. Generally, bone resorption is rapid, and a resorption pit is formed within 10 to 14 days. After resorption is complete, osteoblasts, derived from the bone marrow stromal cells, attach to the resorbed bone surface and produce osteoid, which is then mineralized. Bone formation can take up to 3 or 4 months. Therefore a normal bone remodeling cycle in adults can last 4 to 6 months (see Figure 101-1A). A number of metabolic changes such as estrogen deficiency, immobilization, metabolic acidosis, hyperparathyroidism, and systemic and local inflammatory diseases can increase osteoclast number and activity, uncoupling bone turnover. This results in greater bone resorption than bone formation and a net loss of bone tissue. New data show that a number of local factors in bone affect the regulation of bone formation and resorption and the coupling of these processes. These include insulin-like growth factors (IGFs), interleukins (IL-1, IL-6, and IL-11), tumor necrosis factor (TNF), receptor activator of nuclear factor κB ligand (RANKL), and transforming growth factor (TGF).\textsuperscript{5} Animal studies have shown that IL-1, IL-6, and TNF knockout mice do not lose bone with estrogen deficiency.\textsuperscript{5} In addition, inflammatory arthritis animal models find that TNF, IL-1, and IL-6 are all strong stimulators of osteoclastic bone resorption. This link between the immune system and the maintenance of bone mass is intriguing, but additional work is required before we can understand its significance.

A number of mechanisms underlie primary osteoporosis including a low peak bone mass as a young adult and rapid bone loss during menopause. Factors contributing to age-related bone loss include impaired calcium absorption with age, a compensatory rise in parathyroid hormone (PTH) levels, and greater resorption than formation of bone. Estrogen deficiency is associated with the release of cytokines including RANKL, IL-1, IL-6, and TNF, which leads to the recruitment and stimulation of osteoclasts in the marrow and increased production of bone-resorptive cytokines, which may contribute to menopause-related bone loss.\textsuperscript{3} Estrogen therapy, however, inhibits IL-1 release, and in oophorectomized rats and mice, an inhibitor of IL-1 (the IL-1 receptor antagonist) suppresses bone loss.\textsuperscript{6} IL-6 levels also increase with age in human marrow cultures\textsuperscript{7} and in peripheral monocytes. IL-1 and TNF induce the production of IL-6 from osteoblasts and stromal cells. Further evidence supporting a role of IL-6 in bone turnover includes data showing that oophorectomized IL-6 knockout transgenic mice do not lose bone. Two other proteins have been identified that influence osteoclast activity: osteoprotegerin (OPG) and RANKL, which are produced by osteoblasts.\textsuperscript{8} Estrogen deficiency increases osteoblast production of RANKL, which stimulates maturation and activity of osteoclasts by attaching to RANKL on the surface of immature and mature osteoclasts. Simultaneously, estrogen deficiency decreases osteoblast production of OPG, the decoy receptor that reduces RANKL production and activity. Adenoviral delivery of OPG ameliorates bone resorption in a mouse ovariectomy model of osteoporosis.\textsuperscript{9} Both preclinical animal models and clinical trials of women with low bone mass have been completed and demonstrate that inhibition of RANKL with a monoclonal antibody (RANKL inhibitor) prevents estrogen-deficiency bone loss.\textsuperscript{5}

In addition, a number of genetic, nutritional, and lifestyle risk factors predispose to the development of osteoporosis. Whites and Asians are at risk for low bone mass and osteoporosis, whereas African-Americans have a higher bone density and one-third to one-half the number of fractures.\textsuperscript{1,8,9} Some studies show that African-Americans have lower vitamin D and urinary calcium levels, higher PTH levels, and skeletal resistance to the effects of PTH on bone.\textsuperscript{10,11} Studies in twins and families show that up to 80% of the variance in bone mass is accounted for by genetic factors.\textsuperscript{13} A maternal history of hip fracture, for example, is associated with a twofold increased risk of a hip fracture.\textsuperscript{14} Data from Uitterlinden and colleagues\textsuperscript{15} show that the gene encoding collagen type IA1 is associated with low bone density. However, COLIA1 is associated with lower baseline bone density and not an increased rate of bone loss. Further, genetically determined architectural features of bone such as a long hip axis length may contribute to increased fracture risk; conversely, a short hip axis length confers some protective effect.\textsuperscript{16} Recently, a family has been described whose members have high bone mass but are otherwise phenotypically normal. This family has a
mutation (an amino acid change) in the low-density lipoprotein receptor-related protein 5 (LRP5). Using in situ hybridization to a rat tibia, expression of LRP5 was detected in areas of bone involved in remodeling. Additional studies have reported that this LRP5 mutation increases Wnt signaling, which may alter bone mass through a primary defect of bone formation. Individuals with this mutation demonstrate normal levels of bone resorption, but specific markers of bone formation are strikingly elevated. The observation that LRP5 is expressed at high levels in osteoblasts is consistent with its having a role in this area. Further work is now required to determine whether other mutations in the chromosome containing the LRP5 segment are associated with a variation in bone density in the general population.\(^{12,13}\)

Recently, genome-wide scans have been performed in a number of cohorts and have found a number of single nucleotide polymorphisms (SNPs) that are associated with osteoporosis (fracture or bone mineral density [BMD] associations). They include VDR, ESR1, ESR2, LRP5, LRP4, SOST, GRP177, OPG, RNAK, RNAKL, COL1A, SPP1, ITGA1, SP7, and SOX6, which can be reasonably assigned as confirmed or replicated, and another 30 or so genes as promising candidates. Of note, confirmed and promising genes are clustered in three biologic pathways: the estrogen endocrine pathway, the Wnt/beta-catenin signaling pathway,...
and the RANKL/RANK/OPG pathway. New biologic pathways will certainly emerge when more osteoporosis genes are identified and validated. \(^\text{18a}\) Other risk factors for osteoporosis, as enumerated in Table 101-1, include low body weight and reduced gonadal steroid levels. \(^\text{13}\) Lifestyle factors that may contribute to the development of osteoporosis include cigarette smoking, excessive alcohol intake, reduced physical activity, and inadequate calcium intake, according to some reports. Cigarette smokers have poorer health than nonsmokers, impaired calcium absorption, lower estrogen levels, earlier menopause, and more fractures, and they exercise less; smoking cessation reverses this risk of osteoporosis.

In a large, prospective study of 9516 women older than 65 years, the following lifestyle factors significantly increased the risk of hip fracture: no walking for exercise, intake of more than two cups of coffee daily, current use of long-acting benzodiazepines and anticonvulsant drugs, current weight less than weight at age 25 years, height greater than 5 feet 7 inches, age older than 80 years, fracture since age 50 years, inability to stand from a chair without using arms, poor depth perception, and self-evaluation of health as fair to poor. \(^\text{14}\) Low bone density in conjunction with a fall or trauma predisposes an individual to a fracture. Poor health and compromise of neuromuscular function increase the risk of osteoporosis and falls, which in turn increase the risk of hip fracture. \(^\text{14}\) Importantly, elderly white women with both a low bone mass and more than two risk factors have a nearly 20-fold increased risk for fracture.

Secondary causes of bone loss that can affect women and men of all ages and races are listed in Table 101-2. Glucocorticoid therapy is the most common secondary cause of bone loss. Osteoporotic fractures develop in an estimated 30% to 50% of glucocorticoid-treated patients. \(^\text{19}\) Glucocorticoid therapy causes bone loss through a number of different mechanisms such as producing a negative calcium balance through impaired intestinal calcium absorption, increasing urinary calcium excretion, decreasing bone formation, increasing bone resorption by stimulating osteoclast activity by macrophage colony-stimulating factor, and suppressing endogenous gonadal steroid production. \(^\text{19}\) Therapy with glucocorticoids leads to an early and, in some instances, dramatic loss of trabecular bone, with less effect on cortical bone. In hyperthyroidism (Graves’ disease or toxic nodule) or supraphysiologic therapy with thyroid hormone, the ensuing accelerated bone turnover may produce a reduction in bone mass when the thyroid-stimulating hormone level is suppressed, even when thyroid hormone levels are within the normal range. \(^\text{10}\) Athletic amenorrhea, anorexia nervosa, and other hypogonadal states including the use of gonadotropin-releasing hormone agonists \(^\text{17,22}\) may result in bone loss. In addition to estrogen deficiency, women with anorexia nervosa have low levels of IGF-I and reduced levels of adrenal androgen

---

**Table 101-1 Risk Factors for Osteoporosis**

<table>
<thead>
<tr>
<th>Primary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous fracture after age 30</td>
</tr>
<tr>
<td>Family history of hip fracture</td>
</tr>
<tr>
<td>Cigarette smoking</td>
</tr>
<tr>
<td>Weight &lt;127 lb</td>
</tr>
<tr>
<td>Low bone mineral density</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Secondary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonmodifiable</td>
</tr>
<tr>
<td>White race</td>
</tr>
<tr>
<td>Advanced age</td>
</tr>
<tr>
<td>Frailty or poor health</td>
</tr>
<tr>
<td>Dementia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modifiable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low calcium intake</td>
</tr>
<tr>
<td>Eating disorder</td>
</tr>
<tr>
<td>Low testosterone levels (men)</td>
</tr>
<tr>
<td>Premenopausal estrogen deficiency (amenorrhea &gt;1 yr or menopause at age &lt;45 yr)</td>
</tr>
<tr>
<td>Excessive alcohol intake</td>
</tr>
<tr>
<td>Physical inactivity</td>
</tr>
<tr>
<td>Impaired vision</td>
</tr>
<tr>
<td>Neurologic disorder</td>
</tr>
<tr>
<td>Lack of sunlight exposure</td>
</tr>
</tbody>
</table>

**Table 101-2 Medical Disorders and Medications Associated with Bone Loss and Osteoporosis**

<table>
<thead>
<tr>
<th>Primary osteoporosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juvenile osteoporosis</td>
</tr>
<tr>
<td>Postmenopausal osteoporosis</td>
</tr>
<tr>
<td>Involutional osteoporosis</td>
</tr>
<tr>
<td>Endocrine abnormalities</td>
</tr>
<tr>
<td>Glucocorticoid excess</td>
</tr>
<tr>
<td>Thyroid hormone excess (supraphysiologic)</td>
</tr>
<tr>
<td>Hypogonadism (including from prolactinoma or anorexia nervosa)</td>
</tr>
<tr>
<td>Hyperparathyroidism</td>
</tr>
<tr>
<td>Hypercalcemia</td>
</tr>
<tr>
<td>Processes affecting bone marrow</td>
</tr>
<tr>
<td>Multiple myeloma</td>
</tr>
<tr>
<td>Leukemia</td>
</tr>
<tr>
<td>Gaucher’s disease</td>
</tr>
<tr>
<td>Systemic mastocytosis</td>
</tr>
<tr>
<td>Immobilization</td>
</tr>
<tr>
<td>Space flight</td>
</tr>
<tr>
<td>Gastrointestinal diseases</td>
</tr>
<tr>
<td>Gastrectomy</td>
</tr>
<tr>
<td>Primary biliary cirrhosis</td>
</tr>
<tr>
<td>Celiac disease</td>
</tr>
<tr>
<td>Renal insufficiency</td>
</tr>
<tr>
<td>Chronic respiratory diseases</td>
</tr>
<tr>
<td>Connective tissue disorders</td>
</tr>
<tr>
<td>Osteogenesis imperfecta</td>
</tr>
<tr>
<td>Homocysteinuria</td>
</tr>
<tr>
<td>Ehlers-Danlos syndrome</td>
</tr>
<tr>
<td>Rheumatologic disorders</td>
</tr>
<tr>
<td>Ankylosing spondylitis</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Systemic lupus erythematosis</td>
</tr>
</tbody>
</table>

**Medications**

<table>
<thead>
<tr>
<th>Anticonvulsants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heparin</td>
</tr>
<tr>
<td>Methotrexate</td>
</tr>
<tr>
<td>Cyclophosphamide and gonadotropin-releasing hormone agonists (hypogonadism)</td>
</tr>
<tr>
<td>Lithium</td>
</tr>
<tr>
<td>Cyclosporine</td>
</tr>
<tr>
<td>Aluminum</td>
</tr>
<tr>
<td>Excessive alcohol</td>
</tr>
<tr>
<td>Premenopausal tamoxifen</td>
</tr>
<tr>
<td>Aromatase inhibitors</td>
</tr>
</tbody>
</table>

dehydroepiandrosterenedione, which may contribute to the development of osteoporosis.21

**Osteoporosis in Men**

Osteoporosis in men was not recognized 20 years ago but is now a major public health problem owing to men’s longer life spans. The epidemiology of osteoporosis in men is just now being evaluated. Fracture risk occurs in adolescence and young adulthood and then increases after the age of 70. Long bone fractures occur more commonly in young men, while hip and spine fractures are more prevalent in men older than 70 years. The increase in fractures in older men is just as significant as it is in women, but it occurs about 10 years later in life, with an age-adjusted incidence of hip fractures in men of about one-third to one-half that of women.24 Elderly men who sustain hip fractures have a greater risk of dying or being permanently disabled compared with women.24 Risk factors for osteoporosis in men include older age, low BMD, history of a low-trauma fracture as an adult, and a family history of osteoporotic fractures.

There are a number of secondary causes of osteoporosis in men; for instance, hypogonadism causes an increase in bone turnover and rapid bone loss as gonadal function declines with age. At this time, severe hypogonadism from androgen deprivation therapy for prostate cancer is common in elderly men. The exact role of estrogen and androgens in male skeletal health is not yet known. Although estrogen is necessary for the young male skeleton, serum estrogen levels are highly correlated with bone remodeling, BMD, and rate of BMD loss in older men; the associations are stronger than with testosterone. However, serum testosterone levels are also highly correlated with indices of bone resorption and formation. The roles of estrogen and testosterone in the male skeleton need additional investigation.25 Some of the other common causes of osteoporosis in men that are not as frequent in women are alcoholism, gastrointestinal disorders including hepatic disorders, and malabsorption.24

**Osteoporosis in Rheumatic Diseases and Other Conditions**

Recently, studies have reported significant bone loss in patients with systemic inflammatory diseases such as rheumatoid arthritis, systemic lupus erythematosus (SLE), and ankylosing spondylitis. Patients with rheumatoid arthritis experience periarticular and generalized bone loss, with an increased incidence of fractures compared with the general population.26 T lymphocytes, tissue macrophages, and synovial-like fibroblasts release inflammatory cytokines (IL-1, TNF, IL-6) and inhibitory wnt signaling proteins such as dkk-1 and RANKL, which stimulate osteoclasts in the bone marrow and synovium to actively resorb bone; in addition, osteoblast maturation is altered.27,28 In an animal model of inflammatory arthritis induced with collagen, animals pretreated with OPG did not have bone loss within the periarticular bone or the presence of erosions.29 Additional factors that may contribute to osteoporosis in patients with rheumatic diseases include decreased mobility, glucocorticoid therapy, and systemic inflammation.30 Some data, however, show that low-dose glucocorticoid therapy in women with rheumatoid arthritis does not have adverse skeletal effects, possibly because of a decrease in disease activity in association with the suppression of inflammatory cytokines and improved physical activity and function.31,32 Ankylosing spondylitis is also associated with fractures and reduced bone density in the spine and proximal femur, even early in the disease.33 Patients with SLE have a high rate of osteoporotic fractures in the presence of low to normal bone mass, suggesting that systemic inflammation alters bone turnover. Increased serum levels of TNF can reduce osteoblast maturation and increase osteoclast maturation and activity; in addition, other inflammatory factors such as oxidized low-density lipoproteins and inflammatory high-density lipoproteins can direct mesenchymal stem cells to differentiate into adipocytes instead of osteoblasts and impair bone mass.34 Infiltrative processes in the marrow such as multiple myeloma, mastocytosis, and Gaucher’s disease may produce osteoporosis. Patients with Gaucher’s disease show an accumulation of glucocerebrosides in macrophages in the spleen, liver, and bone marrow, which causes hepatosplenomegaly, anemia, thrombocytopenia, bone infarcts and infections, fractures, and aseptic necrosis.35

The immunosuppressant drug cyclophosphamide (Cytoxan) induces amenorrhea and hypogonadism, which may increase the risk of bone loss. Women who undergo premature menopause from cyclophosphamide therapy can have estrogen-deficiency bone loss in their 30s. Young women with SLE who try to preserve ovarian function while undergoing cyclophosphamide therapy by taking gonadotropin-releasing hormone agonists may also experience estrogen-deficiency bone loss. In rodent models, the immunosuppressive drug cyclosporine produces a time- and dose-dependent bone loss;36 in contrast, azathioprine (Imuran) and rapamycin (sirolimus) do not appear to adversely affect skeletal homeostasis.37 Therapy with both cyclosporine and prednisone in transplant recipients is associated with early accelerated bone loss after the initiation of treatment and the development of osteoporosis and fractures with continued exposure.38

Vitamin D deficiency may also manifest as osteopenia and fractures, but this condition is both preventable and treatable.39 Vitamin D insufficiency is common in older patients and in those with SLE who do not get an adequate amount of sunlight or use potent sunscreens. Also, patients with malabsorption syndromes and liver disease can be vitamin D deficient. Unlike the situation in osteoporosis, very low vitamin D levels are often characterized by a mineralization defect and osteomalacia. Vitamin D deficiency is reported to be present in up to 50% of women with hip fractures.39

**Assessment of Bone Density and Osteoporotic Risk**

Osteoporosis may first be diagnosed when a radiograph shows signs of demineralization or a spinal film shows evidence of compression fractures of vertebral bodies. Because an estimated 25% to 50% of bone mass must be lost to show osteopenia on radiographs, conventional radiography is an insensitive technique for diagnosing bone loss. Radiographs may demonstrate signs of secondary causes of osteoporosis.
such as the presence of subperiosteal resorption in hyperparathyroidism, characteristic lytic changes or bone infarcts in Gaucher’s disease, local sites of lytic destruction in malignancy, and pseudofractures in osteomalacia. Bone densitometry makes it possible to measure the amount of bone in the relevant fracture sites of the spine, forearm, and proximal femur, as well as the total body.

Techniques for evaluating bone mass include dual-energy x-ray absorptiometry (DEXA) and quantitative computed tomography (CT) scanning of the spine.\(^{1,2}\) Bone density evaluations using DEXA incorporate the attenuation of soft tissue and bone by radiographs to calculate the BMD. DEXA is both precise and safe, with a low radiation exposure. With reproducibility errors of approximately 0.6% to 1.5%, this technique can detect small changes over time.\(^{2,40,41}\) Further, newer DEXA techniques measure bone density rapidly, in 0.5 to 2.5 minutes. Quantitative CT scanning allows for the direct measurements of trabecular bone in the central region of the spine, but the procedure entails a comparatively high radiation exposure and time and precision errors are usually higher than those associated with DEXA.

Figure 101-1B and C show the BMD in a postmenopausal patient compared with that of young, healthy controls to determine whether there is reduction in BMD compared with peak bone mass (percentage of young healthy controls expressed as a t score) and with age-matched controls to assess whether BMD is diminished relative to an age-matched cohort (percentage of age-matched controls expressed as a z score). There is an inverse relationship between bone density and the gradient of risk for fracture.\(^{42}\) Prospective studies show that bone densitometry identifies patients with an increased gradient of risk for fracture. In 8134 women, a 1–standard deviation (SD) decrement in the bone density of the spine and the femoral neck compared with age-adjusted controls was associated with a 1.6- and 2.6-fold increased risk of hip fracture, respectively.\(^{43}\) Measurement of bone density in the hip is more predictive of hip fracture than is measurement at another site. Studies show that in women older than 65 years, hip bone density is predictive of spine and hip fracture and that conventional spine density does not add to the diagnostic utility of a single hip bone density test in assessing the risk of fracture.

Although bone densitometry provides a quantitative measure of bone mass, in vitro studies using ultrasonography indicate that this technique also provides information about the mechanical properties of bone including both density and elasticity. These qualities are strong predictors of bone strength. Ultrasound techniques include speed-of-sound and broadband ultrasound attenuation methods; the speed-of-sound technique reflects bone density and elasticity, and broadband ultrasound attenuation is an indicator of bone density, bone structure, and composition. Approved for diagnostic use by the U.S. Food and Drug Administration (FDA), both ultrasound techniques have been shown to discriminate between normal and osteoporotic patients at increased fracture risk.\(^{44}\) The t score parameters used by some ultrasound machines do not correspond to t score levels as measured by DEXA. Although ultrasonography is a radiation-free technique that may provide information about the risk of fracture and bone quality, the reproducibility of this technique and the measurement sites of mainly cortical bone or low-weight-bearing locations may make it unsuitable for monitoring small changes in bone over time. Therefore ultrasound measurements cannot reliably be used to monitor response to osteoporosis therapies. Further data are necessary to validate the clinical utility of ultrasonography.

On the basis of the guidelines of the Scientific Advisory Board of the National Osteoporosis Foundation, bone densitometry is useful in determining which patients might benefit from therapy to protect the skeleton including patients who have a deficiency of gonadal hormones (postmenopausal women younger than 65 years with one or more risk factors or older than 65 years regardless of risk factors), postmenopausal fracture, evidence of osteopenia or a vertebral abnormality on radiographs, hyperparathyroidism, or exposure to supraphysiologic doses of glucocorticoids (Table 101-3). Bone densitometry is also used to decide when to commence therapy for osteoporosis and to assess the clinical response to therapeutic interventions.\(^{40}\) Screening normal premenopausal women is not cost-effective.

The World Health Organization (WHO) has published criteria for osteoporosis on the basis of bone density\(^{45}\):

1. Normal bone density if the t score is greater than −1.
2. Osteopenia (low bone mass) is defined as a bone density measurement between 1 and 2.5 SD below the young-adult mean (t score between −1 and −2.5).
3. Osteoporosis is defined as a bone density measurement less than 2.5 SD below that of young, healthy controls (t score < −2.5).

The National Osteoporosis Foundation recommends treatment for all individuals who have a lumbar spine, hip, or femoral neck t score of −2.5 or lower. However, for individuals who have a bone density between −1 and −2.5 the National Osteoporosis Foundation recommends performing a Fracture Risk Assessment or FRAX using a computer program that incorporates clinical risk factors for osteoporosis with and without femoral neck BMD. FRAX also provides a 10-year risk of a hip fracture or a major osteoporotic fracture (hip, proximal humerus, and wrist). The clinical risk factors included in the FRAX program include age, weight, height, history of a fracture as an adult, parental history of a hip fracture, current glucocorticoid use, secondary cause of osteoporosis, alcohol intake of more than two drinks a day, and current smoker.\(^{45,46}\) These risk factors are added to femoral neck t score for the 10-year fracture risk. In the

### Table 101-3: Indications for Bone Densitometry

<table>
<thead>
<tr>
<th>Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>All postmenopausal women &lt; 65 yr who have one or more additional risk factors for osteoporosis (besides menopause)</td>
</tr>
<tr>
<td>All women &gt; 65 yr regardless of additional risk factors</td>
</tr>
<tr>
<td>To document reduced bone density in patients with vertebral abnormalities or osteopenia on radiographs</td>
</tr>
<tr>
<td>Estrogen-deficient women at risk for low bone density who are considering use of estrogen or an alternative therapy, if bone density would influence the decision</td>
</tr>
<tr>
<td>Women who have been on estrogen replacement therapy for prolonged periods or to monitor the efficacy of a therapeutic intervention or interventions for osteoporosis</td>
</tr>
<tr>
<td>To diagnose low bone mass in glucocorticoid-treated individuals</td>
</tr>
<tr>
<td>To document low bone density in patients with asymptomatic primary or secondary hyperparathyroidism</td>
</tr>
</tbody>
</table>
Markers of Bone Turnover

The development of sensitive biochemical markers of bone turnover makes it possible to analyze changes in bone formation and resorption at a given point in time and obtain additional information about a patient's risk of bone loss and fracture. Only three bone formation markers are currently available. Osteocalcin, a noncollagenous matrix protein in bone, is produced exclusively by osteoblasts; it correlates with histomorphometric bone measurements. In most conditions, bone resorption and formation are tightly coupled and osteocalcin levels reflect bone turnover. The other markers of bone formation are bone-specific alkaline phosphatase (BSAP), an enzyme that is activated as osteoblasts mature, and amino-terminal propeptide of type I procollagen, a protein whose synthesis is high in maturing osteoblasts.

Sensitive indicators of bone resorption derived from the degradation of mature collagen include the urine and serum markers of type I collagen cross-links including amino-terminal telopeptide of type I collagen (N telopeptides, or NTX) or carboxy-terminal telopeptide of type I collagen (C telopeptides, or CTX). Urinary pyridinoline cross-link, NTX, and CTX levels correlate with histomorphometric determinations of bone resorption; these biomarkers increase with menopause and are high in patients with a variety of disorders characterized by accelerated bone turnover including Paget's disease, osteoporosis, and rheumatoid arthritis. Urinary excretion of N telopeptides is inversely related to total hip and spinal bone density and, according to some studies, may be a more specific index of bone resorption than urinary pyridinoline levels. Small prospective studies have found high turnover markers, and low BMD has been associated with incident fracture risk but these studies have not been replicated in larger studies. In clinical studies, antiresorptive agents such as estrogen, bisphosphonates, and inhibitor of RANKL induce a significant decrease (30% to 80%) first in markers of resorption and then in bone formation markers, often within 3 to 6 months. Resorption markers decrease before formation markers and correlate with either maintenance of or increase in BMD. A significant change in bone markers can be observed within months of antiresorptive therapy, before there are changes in BMD. Both bone formation and resorption marker changes over 6 to 12 months have been found to predict future fracture risk. In a study of alendronate to reduce osteoporotic fractures, patients who had more than a 30% reduction in bone alkaline phosphatase had the greatest reduction in risk for new vertebral and nonvertebral fractures. Interestingly, studies have found that reductions in markers of bone turnover, either resorption or formation markers, are associated with a reduction in fracture risk. However, long-term, prospective studies of large numbers of women are necessary to determine whether selective biochemical markers of bone turnover can predict changes in BMD or fracture risk and whether these tests should be used in standard clinical practice.

Evaluation for Secondary Bone Loss

The workup for osteoporosis is directed toward excluding secondary causes of bone loss and includes a determination of serum calcium, phosphorus, supersensitive thyroid-stimulating hormone, 25-hydroxyvitamin D (25-OHD), and intact PTH, as well as urine calcium and creatinine levels. Also, a complete blood cell count, alkaline phosphatase and liver function tests, erythrocyte sedimentation rate (in some cases), and serum and urine protein electrophoresis for patients older than 50 years may be necessary (Table 101-4). In men, additional testing for secondary causes of osteoporosis includes serum testosterone and luteinizing hormone. Further tests to rule out neoplastic or endocrinologic disorders and a bone biopsy (a decalcified bone specimen is obtained after a double tetracycline label with two different fluorescent labels) should be considered in certain patients with progressive bone loss and in those in whom osteoporosis is unlikely. Identification and appropriate therapy for underlying secondary causes of osteoporosis are important. For example, treatment of vitamin D deficiency is best accomplished with vitamin D supplements. Parathyroidectomy in patients with hyperparathyroidism characterized by hypercalcemia, hypercalciuria, nephrolithiasis, age younger than 50 years, or low cortical BMD (z score ≤2) was associated with a large (4% to 12.8%) increase in bone density over 4 years. Bone density was, however, stable for up to
6 years in patients with mild hyperparathyroidism. In addition, treatment of hyperthyroidism, hypercortisolism, and a variety of other disorders that may cause osteoporosis can produce increments in bone mass. Reduction in the systemic inflammation associated with rheumatic diseases such as TNF blocking agents for rheumatoid arthritis or ankylosing spondylitis or glucocorticoid-sparing agents for SLE (e.g., azathioprine [Imuran], mycophenolate mofetil [CellCept]) can also produce increments in bone mass.

**Treatment**

**Calcium**

The goals of therapy for osteoporosis are to reduce bone resorption and enhance bone formation, if possible. Bone loss occurs when the calcium intake and absorption are insufficient to balance the daily calcium losses. Prospective data show that calcium stabilizes bone mass.

Table 101-5 shows the current recommendations for optimal calcium intake for women and men from the 1997 report of the Institute of Medicine to the National Academy of Sciences. In the absence of kidney stones or an underlying disorder of calcium metabolism, these calcium intakes are safe. To prevent negative calcium balance, premenopausal women require 1000 mg and postmenopausal women 1200 mg of total elemental calcium daily. Children have increasing calcium requirements during adolescence, and data show increased bone accretion with increased calcium intake in prepubescent and pubertal children. Calcium carbonate contains 40% elemental calcium and should be taken with meals because of poor absorption in achlorhydric patients in the absence of food. Calcium citrate, which contains 24% elemental calcium, has better bioavailability and is more readily absorbed. It is also absorbed well on an empty stomach in patients with achlorhydria. Recent studies underscore the fact that only recommended daily allowances of calcium and vitamin D are useful for bone health in postmenopausal women and elderly men. Additional supplementation over the recommended daily allowance has resulted in reports of increased cardiovascular disease.

**Estrogen**

Hormone replacement therapy (HRT) was once the mainstay of treatment in osteoporosis because estrogen inhibits bone resorption, produces a small rise in bone density, and reduces the risk of fracture by approximately 50% in retrospective observational studies. Cardiovascular disease is the leading cause of death in postmenopausal women. Previous data from longitudinal observational studies suggested that estrogen replacement had a beneficial effect on reducing primary and secondary cardiac events in postmenopausal women. However, in 1998, data from the 4-year Heart and Estrogen Progestin Replacement Study were published. In this study, 2763 postmenopausal women with a previous history of heart disease were randomized to receive estrogen (0.625 mg) plus progestin (2.5 mg) or placebo alone. Results showed no reduction in the overall rate of coronary heart disease or cardiac events in the treatment group; in fact, an early increase in risk for cardiac events was noted, possibly related to increased coagulability.

In addition, a large, multicenter, longitudinal study by the Women’s Health Initiative (WHI)—in which 162,000 women aged 50 to 79 years were randomized into a placebo group, an HRT group (if the uterus was intact), or an estrogen-only group (if the uterus was absent)—was terminated early due to an increased risk of breast and cardiovascular events. The research goals for the WHI study were to determine the effects of HRT, diet modification, and calcium and vitamin D supplements on heart disease, osteoporosis, and colorectal cancer risk. After a mean follow-up of 5.3 years in an 8.5-year study, the HRT group had an increased risk of seven more cardiac events per 10,000 women taking the drug for a year, eight more invasive breast cancers, eight more strokes, and eight more pulmonary emboli, but six fewer colorectal cancers and five fewer hip fractures.

At this time, the general recommendation is that HRT should be used only for vasomotor symptoms that occur at the time of menopause. When these symptoms abate, it is recommended that estrogen replacement (combined estrogen and progestin for women with an intact uterus) be stopped because the perceived cardiovascular benefits have

![Table 101-4 Evaluation of Osteoporosis](image)

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Optimal Daily Calcium Intake (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infants</td>
<td></td>
</tr>
<tr>
<td>Birth-6 mo</td>
<td>400</td>
</tr>
<tr>
<td>6 mo-1 yr</td>
<td>600</td>
</tr>
<tr>
<td>Children 1-8 yr</td>
<td>500-800</td>
</tr>
<tr>
<td>Adolescents 9-18 yr</td>
<td>800-1200</td>
</tr>
<tr>
<td>9-10 yr</td>
<td>1200-1500</td>
</tr>
<tr>
<td>11-18 yr</td>
<td>1500</td>
</tr>
<tr>
<td>Pregnant and nursing females</td>
<td>1300</td>
</tr>
<tr>
<td>Men and women</td>
<td>19-50 yr</td>
</tr>
<tr>
<td>&gt;50 yr (± hormone replacement therapy)</td>
<td>1200-1500</td>
</tr>
</tbody>
</table>

not been substantiated, and the cardiovascular disease and breast cancer risk make the benefit-to-risk ratio unacceptable for most women. It is important to acknowledge that the estrogen-only arm of the WHI study in women without a uterus did not show an increased risk of heart disease or breast cancer.

If a woman and her physician decide that she is going to take HRT or estrogen alone for vasomotor symptoms, in those with an increased risk of coagulability, transdermal estrogen replacement should be used.

**Selective Estrogen Receptor Modulators**

The ideal estrogen replacement therapy would confer the beneficial effects of estrogen on bone and cardiovascular disease without increasing the risk of breast or uterine cancer. Selective estrogen receptor modulators (SERMs) are a nonsteroidal class of drugs that bind to the estrogen receptor and differ from one another in their actions on estrogen-responsive tissues, acting selectively as agonists or antagonists. Tamoxifen, the first available SERM, is an estrogen antagonist that binds to the estrogen receptor and also has estrogen-agonist effects on bone, lipids, clotting factors, and endometrium. Tamoxifen therapy in women with breast cancer produced a small increase in bone density of the spine over 2 years, with no effect on radial bone density, in association with reductions in both low-density lipoprotein and total cholesterol. The Breast Cancer Prevention Trial studied 13,388 women at increased risk for breast cancer, comparing treatment with tamoxifen (20 mg daily) to placebo for 5 years. Tamoxifen reduced the risk of invasive and noninvasive breast cancer by 50%, and a decreased risk of fracture was observed as well: 45% reduction at the hip and 29% at the spine. An increased incidence of low-grade endometrial cancer was noted, but there was no change in the risk of ischemic heart disease.

Raloxifene, now FDA approved for the prevention and treatment of osteoporosis, is a SERM that acts as an estrogen agonist on bone, with antagonist effects on the breast and uterus. Raloxifene (60 mg/day over a 2-year study period) increased BMD in the lumbar spine by 2.4%, in the total hip by 2.4%, and in the total body by 2%, with a reduction in fracture risk at 2 years similar to that seen with estrogen or alendronate (5 mg) treatment. Over the 2-year study period, raloxifene produced a significant reduction in vertebral fractures: Fractures were present in 1.6% of raloxifene-treated women, compared with 2.9% of those in the placebo group; fractures recurred in 7.6% of treated women with a previous fracture, compared with 14.3% of those in the placebo group. Endometrial thickness is not increased by raloxifene, but menopausal symptoms may be made worse. Raloxifene has been shown to decrease low-density lipoprotein cholesterol by 12%, with a nonsignificant increase in high-density lipoprotein cholesterol; cardiovascular protection has not yet been determined. However, raloxifene, unlike estrogen, does not affect C-reactive protein, which is associated with a risk of cardiovascular disease. Raloxifene also decreased the incidence of breast cancer by 76% in patients enrolled in a clinical study of osteoporosis, with breast cancer incidence studied as a secondary endpoint. A study that evaluated the effects of raloxifene on cardiovascular disease found no effect. One study compared tamoxifen and raloxifene, and another study evaluated raloxifene versus placebo, in the prevention of breast cancer. The first study reported that both tamoxifen and raloxifene reduced the risk of developing breast cancer, and the second found that raloxifene reduced the risk of estrogen receptor–positive breast cancer compared with placebo in postmenopausal women. At this time, there is little information on the use of raloxifene in men, so it is not recommended for male patients.

**Testosterone**

Men with osteoporosis, hypogonadism, and symptoms of low libido may benefit from testosterone replacement therapy. This can be administered as testosterone cypionate or enanthate (50 to 400 mg intramuscularly every 2 to 4 weeks) or as a transdermal testosterone replacement patch that is applied to the scrotal area (Testoderm, 4 to 6 mg/day) or elsewhere (Androderm, 2.5 or 5 mg/day). Most studies find that bone mass increases with testosterone replacement when levels of testosterone were low at the initiation of therapy.

**Calcitonin**

Calcitonin, a 32–amino acid peptide synthesized by the C cells of the thyroid gland, is a potent inhibitor of osteoclast-mediated bone resorption. Although human and salmon calcitonin are commercially available, salmon calcitonin is most commonly used because of its greater potency. On the basis of data showing an increase in total body calcium, parenteral calcitonin was approved by the FDA for the treatment of osteoporosis in 1984, and calcitonin in a nasal spray was approved for the treatment of postmenopausal osteoporosis in 1995. Parenteral calcitonin (100 IU subcutaneously or intramuscularly three times a week or daily) can maintain bone density or produce a small increase in bone mass in the spine and, in some instances, the forearm, particularly in patients with a high bone turnover. Nasal spray calcitonin is absorbed through the nasal mucosa and is approximately 40% as potent as the parenterally administered drug (e.g., 50 to 100 IU of injectable calcitonin is comparable with 200 IU of nasal spray calcitonin). In osteoporotic women more than 5 years past menopause, nasal calcitonin (200 IU/day) increases spinal bone density 2% to 3% compared with placebo, with no effect on proximal femur bone mass; higher doses are necessary in the early menopausal period. Nasal spray calcitonin therapy in patients with osteoporosis is associated with a 36% reduction in vertebral fractures over 5 years.

The adverse effects of parenteral calcitonin include nausea, flushing, and local irritation at the injection site. Calcitonin given intranasally is well tolerated, with rhinitis and nasal symptoms such as dryness and crusting being potential side effects. Patients treated with parenteral or intranasal calcitonin may also obtain a beneficial analgesic response in the presence of osteoporotic fractures.

**Bisphosphonates**

Bisphosphonates are analogues of pyrophosphate, with a P-C-P rather than a P-O-P core; they are absorbed by the
Table 101-6  Inhibition of Metaphyseal Bone Resorption in Vivo by Bisphosphonates

<table>
<thead>
<tr>
<th>Chemical Modification</th>
<th>Examples</th>
<th>Antiresorptive Potency</th>
</tr>
</thead>
<tbody>
<tr>
<td>First generation: short alkyl or halide side chain</td>
<td>Etidronate</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Clodronate</td>
<td>10</td>
</tr>
<tr>
<td>Second generation: NH3-terminal group</td>
<td>Tiludronate*</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Pamidronate</td>
<td>100</td>
</tr>
<tr>
<td>Third generation: cyclic side chain</td>
<td>Alendronate</td>
<td>100-1000</td>
</tr>
<tr>
<td></td>
<td>Risedronate</td>
<td>1000-10,000</td>
</tr>
<tr>
<td></td>
<td>Ibandronate</td>
<td>1000-10,000</td>
</tr>
<tr>
<td></td>
<td>Zoledronate</td>
<td>10,000</td>
</tr>
</tbody>
</table>

*Tiludronate has a cyclic side chain, not an NH3-terminal group, but it is generally classified as a second-generation compound on the basis of its time of development and potency.


hydroxyapatite of bone and suppress bone resorption. Modification of the side chains can result in the development of a variety of compounds with differing abilities to inhibit bone resorption (Table 101-6). Some bisphosphonates are administered intermittently because of a long skeletal half-life and prolonged retention in bone. These compounds must be taken on an empty stomach because gastrointestinal absorption is less than 10%.

Bisphosphonates have been used for the treatment of patients with Paget’s disease of bone, hypercalcemia of malignancy, and osteoporosis and for the prevention and treatment of glucocorticoid-induced osteoporosis. Etidronate (Didronel) administered intermittently (400 mg/day for 2 weeks in 3-month cycles) produced approximately 5% increase in bone density of the spine and a 50% reduction in vertebral fractures at 2 years. Longer follow-up did not reveal a significant reduction in vertebral fractures compared with baseline, except in a post hoc analysis of patients with three or more fractures and low bone density. Etidronate is not approved by the FDA for the treatment of osteoporosis.

Alendronate (Fosamax) is FDA approved for the prevention and treatment of osteoporosis. Data in postmenopausal women with bone density at least 2.5 SD below peak bone mass show that alendronate (10 mg/day) compared with placebo produces an 8.8% and 7.8% increase in bone density in the spine and femoral trochanter, respectively, and a 5.9% increase in the femoral neck after 3 years of therapy; there are smaller rises (2.3% to 4.4%) in bone density in the spine and proximal femur in women within 0.5 to 3 years of menopause. Fosamax treatment in women with osteoporosis (T score < 2.5) yields a significant reduction in spine and hip fractures compared with the placebo-treated patients.

Treatment with alendronate (5 mg/day over 2 years) increased BMD in the lumbar spine by 2.9% and in the hip by 1.3%; in contrast, estrogen-progestin therapy increased BMD in these locations by 4% and 1.8%, respectively. Alendronate did not reduce the incidence of clinical fractures in women who had low bone mass but not osteoporosis, although longer studies may be necessary. Alendronate treatment is also effective in increasing bone mass in the spine, hip, and total body and helps prevent vertebral fractures and height loss in men with osteoporosis. Adverse effects of bisphosphonates include gastrointestinal symptoms such as stomach pain and esophagitis (caution is advised in patients with active symptoms or a history of ulcer disease), myalgias and arthralgias, and, rarely, osteonecrosis of the jaw, and subtrochanteric fractures with long-term use. A once-a-week preparation of alendronate (70 mg) is the most commonly used dose for the treatment of osteoporosis. This preparation increased spinal and hip bone mass similarly to alendronate 10 mg/day over a 2-year study period.

Risedronate, another oral bisphosphonate, administered at a dose of 5 mg/day increased bone mass and reduced the risk of new vertebral fractures 50% better than placebo. Another study performed to assess the effect of risedronate on hip fractures found that women with osteoporosis (defined by a femoral neck t score of ≤ −4.0) had a significant reduction in the risk of hip fracture. Risedronate has been approved for the prevention and treatment of osteoporosis (35 mg once a week) and for the treatment of Paget’s disease (30 mg/day for 2 months, with retreatment if relapse occurs after 2 months). Studies show that risedronate may be well tolerated even in patients with mild gastrointestinal symptoms. Bisphosphonates may also reduce bone pain.

Ibandronate (Boniva), another aminobisphosphonate, is approved for the treatment and prevention of postmenopausal osteoporosis. In phase III studies of ibandronate (2.5 mg/day) versus placebo in postmenopausal women with osteoporosis, incident vertebral fractures were reduced about 50%. Another study compared ibandronate 150 mg once a month to the daily 2.5-mg dose and found similar gains in lumbar spine and hip BMD. The FDA approved ibandronate 150 mg/month for the treatment of osteoporosis on the basis of this bridging study. Recently, intravenous ibandronate in a dose of 3 mg every 3 months was found to be similar to ibandronate 2.5 mg/day in terms of increasing lumbar spine and hip BMD, and the FDA has approved intravenous ibandronate for this indication. There are no data on hip fractures for this compound.

Zoledronic acid (Reclast) is approved for the treatment and prevention of postmenopausal osteoporosis. In phase III studies in postmenopausal women with osteoporosis treated with zoledronic acid, 5 mg once a year by intravenous infusion, the risk of incident vertebral fractures was reduced by 68%, hip fractures by 40%, and other major osteoporotic fractures by 20% compared with the placebo-treated subjects. Another phase III study of patients who had suffered a hip fracture were randomized to either zoledronic acid 5 mg once a year of placebo and after 3 years found that incident osteoporotic fractures were significantly less in zoledronic-treated patients and the mortality was lowered compared with the placebo-treated patients. Side effects of zoledronic acid include arthralgias and myalgias; however, these adverse events tend to be less frequent with subsequent infusions. Patients undergoing zoledronic acid treatment should have both serum calcium and 25-OHD levels monitored and replaced to normal levels before treatment. Zoledronic acid has been approved for the treatment and prevention of postmenopausal osteoporosis (5 mg once a year), osteoporosis in men (5 mg a year), prevention and treatment of glucocorticoid-induced osteoporosis (5 mg a year), and Paget’s disease. Zoledronic acid at a dose of 4 mg intravenously every 4 weeks is currently approved for the
prevention and treatment of bone metastases in patients with breast cancer and multiple myeloma.

**RANK Ligand Inhibitor**

Denosumab (Prolia) is a monoclonal antibody that is directed against RANKL and is approved for the treatment of postmenopausal osteoporosis. In phase III studies in postmenopausal women with osteoporosis treated with denosumab (60 mg subcutaneously every 6 months) for 36 months or 3 years, the incidence of vertebral fractures was reduced by 68%, hip fractures by 40%, and major osteoporotic fractures by 20%.\(^{97}\) The medication is well tolerated, but adverse events of skin infection requiring hospitalization were higher in denosumab-treated subjects than placebo-treated subjects. Denosumab is a potent inhibitor of bone resorption with suppression of a marker of osteoclast activity.

CTX-1 is suppressed to nearly 90% a few weeks after each injection; however, the suppression of bone turnover is transient and bone turnover and bone density return to baseline levels rapidly if the injections of denosumab are discontinued. Postmenopausal women with low bone mass who were treated with alendronate 70 mg once a week and then switched to denosumab (60 mg every 6 months) had a significant gain of BMD compared with subjects continued on alendronate.\(^{98}\) Before treatment with denosumab, serum calcium and 25-OHD should be checked and replaced if needed up to normal levels.

**Parathyroid Hormone**

Small randomized studies have determined that the 1-34 fragment of the PTH protein can significantly increase bone mass in the spine, with small losses or no gain at the skeletal sites rich in cortical bone.\(^{99,99a}\) In 2001 a recombinant human PTH (rhPTH) composed of the 34 amino acids from the amino-terminal end of the hormone, known as Forteo, was approved for the treatment of postmenopausal osteoporosis. In a large international, multicenter study, osteoporotic women with fracture were randomized to receive rhPTH 20 µg/day, 40 µg/day, or placebo for an average of 21 months. Lumbar spine bone mass increased between 9% and 13% in the rhPTH-treated subjects compared with the placebo-treated ones; hip bone mass also increased slightly. Most important, the risk of new vertebral fractures was reduced nearly 70% in both sets of rhPTH-treated subjects, and nonvertebral low-trauma fractures were reduced nearly 50% compared with placebo-treated patients.\(^{103}\) This study was initially supposed to continue for 3 years; however, it was stopped at approximately 21 months because of preclinical evidence of malignant bone tumors in animal models. Additional studies of osteoporosis in men treated with rhPTH 1-34 have reported significant gains in bone mass.\(^{103}\) Forteo is given as a daily injection. Individuals using this medication may experience headache, nausea, and flushing with initiation of treatment, but these side effects generally become less severe after a few weeks.

A number of recent studies have evaluated whether rhPTH 1-34 or rhPTH 1-84 is more effective in combination with an antiresorptive agent (either bisphosphonates or raloxifene) than rhPTH alone in increasing BMD and reducing fractures.\(^{101,102}\) Interestingly, two studies found that the combination of PTH and alendronate was less effective in stimulating bone gain at the lumbar spine in both osteopenic women and men over 1 to 1.5 years.\(^{101,102}\) PTH stimulates new bone formation, increases bone mass, and reduces new vertebral and nonvertebral fractures, but when the medication is discontinued, the bone gained is rapidly lost. Black and colleagues\(^{102}\) performed a study in which patients were given PTH for 1 year, followed by 1 year of alendronate treatment. Interestingly, the BMD gain after 1 year of PTH was about 6%; when followed by alendronate, nearly 6% BMD was gained at the spine. These data suggest that although PTH is an effective monotherapy for increasing bone mass, especially in the spine, the gain in BMD should be maintained with a potent antiresorptive agent for a number of years. Recent data from Deal and colleagues\(^{104}\) indicate that lumbar spine BMD gained after PTH therapy is maintained with raloxifene treatment.

PTH is the first bone anabolic agent approved for the treatment of osteoporosis. Patients give themselves a subcutaneous injection daily for 18 to 24 months. Other routes of administration are now being studied including an intranasal route and a skin patch.

**Vitamin D**

Physiologic doses of vitamin D are important to ensure normal bone mineralization. Individuals 50 years of age and older should take at least 600 to 1000 IU of vitamin D daily as a multivitamin or combined with a calcium supplement. Hypovitaminosis D is common in the elderly population, with one study demonstrating that 57% of patients in a general medical ward were vitamin D deficient.\(^{105}\) Low vitamin D levels increase the risk of bone loss and fracture. LeBoff and colleagues\(^{39}\) found that 50% of patients admitted with acute femur fractures had vitamin D deficiency (25-OHD level <12 ng/mL), and 36.7% had secondary hyperparathyroidism. Data show seasonal variations in vitamin D levels; low 25-OHD levels during the winter and spring are associated with decreases in bone density. The importance of vitamin D to skeletal health was shown when elderly women in a nursing home treated with only 800 IU/day of vitamin D had a 40% reduction in incident hip fractures over 18 months, compared with placebo-controlled subjects.\(^{106}\) Although this dramatic effect on fractures in an elderly population may represent a correction of vitamin D insufficiency, this study underscores that adequate vitamin D replacement can effectively diminish fractures in older individuals. Insufficient calcium and low 25-OHD levels are common in ambulatory patients and should be identified and treated before antiresorptive or other therapies for osteoporosis are initiated.

Although 200 IU of vitamin D prevents bone loss in the spine, data show that a higher daily intake (800 IU) is necessary to diminish bone loss in the hip during the winter and spring. Daily treatment with 700 IU of cholecalciferol and 500 mg of calcium carbonate reduced the rate of bone loss significantly in the femoral neck, spine, and total body and decreased the incidence of nonvertebral fractures by 50%.\(^{107}\) Therefore to maintain skeletal health, patients require a vitamin D intake that results in a serum 25-OHD level of at least 30 ng/mL. To achieve this level
in patients who do not receive regular sunlight, the daily intake of vitamin D needs to be higher. Replacement with 1,25-dihydroxyvitamin D (1,25(OH)2D) is not recommended because hypercalcemia and hypercalciuria are common and require regular and costly monitoring.

**Preventive Measures**

Because bone loss is not completely reversible with existing therapies, prevention is essential for optimizing skeletal health. Strategies directed at increasing peak bone mass, reducing risk factors for bone loss (e.g., hypogonadism, decreased body fat, cigarette smoking, inactivity, excessive alcohol intake), and reversing the secondary causes of osteoporosis may prevent bone loss. Patients should be advised to consume adequate vitamin D and calcium and participate in a regular weight-bearing exercise program. Weight-bearing exercise increases muscle strength and may stabilize or modestly increase bone density. Emerging data show that increased calcium intake and exercise can add to bone accretion during adolescence and that interventions such as vitamin D and calcium supplementation can reduce fractures in older patients. Thus it is highly recommended that preventive strategies or therapies be instituted at any age to diminish the risk of fractures, which rises exponentially with age.

### GLUCOCORTICOID-INDUCED OSTEOPOROSIS

Bone loss is a common sequela of therapy with glucocorticoids, and glucocorticoid use increases the risk of fractures in patients with rheumatic diseases. The severity of the bone loss in glucocorticoid-treated patients varies, with an approximately 3% to 20% decrease in bone density over 1 to 2 years. Glucocorticoid therapy is associated with increased fractures of the ribs and vertebrae, sites that contain predominantly trabecular bone, and it triples the risk of hip fracture in one-third of patients after 5 to 10 years of treatment. In adults, alternate-day glucocorticoid therapy does not prevent bone loss. In patients with rheumatic diseases, concerns over the development of glucocorticoid-induced osteoporosis often limit the dose and duration of glucocorticoid therapy.

The lowest possible glucocorticoid dose should be used, along with general preventive strategies such as a regular weight-bearing exercise program, adequate calcium and vitamin D intake, and reduction of other risk factors that might contribute to the development of osteoporosis. However, data show that even patients on prednisone 5 mg/day have accelerated bone loss compared with controls. General prophylactic measures to prevent glucocorticoid-induced osteoporosis are shown in Table 101-7. Intestinal calcium absorption is impaired in glucocorticoid-treated patients, and early studies showed that this could be offset with vitamin D (40 to 100 µg/day two to three times a week) or 25-OHD, which produced an increase in bone density of the forearm. However, the administration of supraphysiologic doses of vitamin D requires careful monitoring of the serum and urinary calcium concentrations in patients at high risk for bone loss (with a normal urinary calcium level and no history of nephrolithiasis). An alternative approach in patients at risk for osteoporosis and receiving long-term glucocorticoid therapy is to raise the 25-OHD level into the upper-normal range (>30 ng/mL) to ensure adequate intestinal calcium absorption. This can usually be accomplished by administering vitamin D at 800 IU/day.

Because of the enhanced bone resorption in patients treated with glucocorticoids, investigators have examined the effects of inhibitors of bone resorption. The use of bisphosphonates in patients receiving chronic glucocorticoid therapy is beneficial for both the prevention and treatment of osteoporosis. The use of alendronate 5 or 10 mg/day for 1 year in patients receiving glucocorticoid therapy increased lumbar spine BMD by 2.1% and 2.9%, respectively, and increased femoral neck BMD by 1.2% and 1%, respectively (P < .001). After 1 year of treatment, there was an insignificant reduction in new vertebral fractures, but after 2 years of treatment, there was a nearly 40% reduction in new vertebral fracture risk. Risedronate 5 mg/day was also effective in the prevention and treatment of glucocorticoid-induced bone loss. Recent studies of zoledronic acid (5 mg IV/year) versus risedronate (5 mg/day) and teriparatide (20 µg/day) versus alendronate (10 mg/day) for 18 months indicated that both zoledronic acid and teriparatide improved bone mass more than the comparator, and teriparatide-treated subjects had a significant vertebral fracture risk reduction in glucocorticoid-treated patients. Combination studies of rhPTH 1-34 with HRT were also found to be more effective than HRT alone in glucocorticoid-induced osteoporosis. No difference between treatments in fracture data was observed; however, quantitative CT of the lumbar spine, a measure of only

<table>
<thead>
<tr>
<th>Table 101-7</th>
<th>Recommendations for Prevention and Treatment of Glucocorticoid-Induced Osteoporosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevention</strong></td>
<td>Patients starting GC therapy at a dose equivalent to prednisone ≥ 5 mg/day for 3 mo or longer should:</td>
</tr>
<tr>
<td></td>
<td>- Modify risk factors for osteoporosis (stop smoking, decrease excessive alcohol consumption)</td>
</tr>
<tr>
<td></td>
<td>- Start regular weight-bearing physical exercise</td>
</tr>
<tr>
<td></td>
<td>- Initiate intake of calcium (total 1500 mg/day) and vitamin D (400-800 IU/day)</td>
</tr>
<tr>
<td></td>
<td>- Consider MBD testing to predict risk of fracture and bone loss</td>
</tr>
<tr>
<td></td>
<td>- Initiate bisphosphonate therapy (alendronate 5 mg/day or 35 mg/wk, or risedronate 5 mg/day or 35 mg/wk)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td>Patients on long-term GC therapy should be tested for osteoporosis using BMD measurement. If the T-score is &lt; −1, consider:</td>
</tr>
<tr>
<td></td>
<td>- Risk factor modification including reducing risk of falls</td>
</tr>
<tr>
<td></td>
<td>- Regular weight-bearing physical exercises</td>
</tr>
<tr>
<td></td>
<td>- Calcium and vitamin D supplementation</td>
</tr>
<tr>
<td></td>
<td>- Replacement of gonadal steroids, if deficient</td>
</tr>
<tr>
<td></td>
<td>- Bisphosphonate therapy (alendronate 10 mg/day or 70 mg/wk, or risedronate 5 mg/day or 35 mg/wk); if bisphosphonates are contraindicated or not tolerated, consider calcitonin as second-line agent, intravenous bisphosphonate (pamidronate or zolendronate), or parathyroid hormone 1-34</td>
</tr>
<tr>
<td></td>
<td>- Repeat BMD measurement annually or biannually</td>
</tr>
</tbody>
</table>

trabecular bone, found a nearly 35% increase in PTH-treated patients compared with the estrogen-only group after 12 months of therapy.115

Men on glucocorticoids can have a lowering of testosterone levels.116 They are generally asymptomatic, but if men on glucocorticoids have evidence of low serum testosterone levels and symptoms of low libido, they can be safely treated with testosterone. Bone mass changes have been observed in men with low testosterone levels on glucocorticoids who were treated with testosterone.117 However, because of the risks associated with testosterone treatment, it is more prudent to treat these patients with a bisphosphonate medication.

To prevent bone loss in patients with pulmonary diseases requiring glucocorticoid treatment, therapy with inhaled glucocorticoids has been studied.118 Inhaled glucocorticoids appear to uncouple bone turnover and increase bone loss; however, this is dose dependent. Less than 800 µg/day of inhaled budesonide dipropionate does not increase the risk of osteoporosis, but more than 800 µg/day does. New inhaled steroids are more potent than older ones; for instance, Advair 200 µg/day is equivalent to nearly 5 mg/day of prednisone.119 Therefore patients on chronic steroid inhalers should be screened for bone loss. Because patients receiving glucocorticoids may lose a dramatic amount of bone, it is important to monitor the efficacy of a treatment intervention, assess the need for further diagnostic evaluation for other causes of bone loss, and consider alternative treatment strategies if a given therapy is ineffective in preventing bone loss or fractures. In patients at increased risk of fracture, therapy with a potent bisphosphonate is highly recommended to slow bone loss and the rate of new fractures. Alendronate, risedronate, and zoledronic acid are available for use, and future studies of more potent bisphosphonates and PTH are expected.120-122

OSTEOMALACIA

Osteomalacia is characterized by impaired mineralization of bone matrix. Calcium, phosphate, and vitamin D are necessary for the mineralization of bone. Normally, there is a steep inverse relationship between the serum calcium and PTH concentrations. A small decrease in the serum calcium concentration leads to a rise in PTH release, which promotes distal renal calcium reabsorption, proximal tubular renal phosphorus excretion, and resorption of calcium from bone. Vitamin D is produced in the skin in the presence of ultraviolet light or absorbed in the intestine from dietary or nutritional sources. Activation of vitamin D to 25-OHD occurs in the liver and to 1,25(OH)2D in the proximal tubules of the kidney. PTH, hypocalcemia, and hypophosphatemia stimulate the renal 1-hydroxylase enzyme that converts 25-OHD to 1,25(OH)2D, which in turn indirectly enhances intestinal calcium absorption.123

Osteomalacia results from reduced availability of calcium or phosphate for incorporation into the hydroxyapatite of bone or from deficient absorption or activation of vitamin D.124 The term rickets applies to the defective mineralization of bone and the cartilaginous growth plate in growing children. As shown in Table 101-8, osteomalacia or rickets may result from decreased availability of vitamin D as a consequence of insufficient ultraviolet light exposure, insufficient vitamin intake, or malabsorption in patients with gastrointestinal or biliary disorders. Reduced levels of 25-OHD are caused by severe liver disease, increased renal excretion of vitamin D metabolites due to nephrotic syndrome, or accelerated metabolism of 25-OHD caused by anticonvulsant drugs. Decreased activation of 1,25(OH)2D is seen in patients with renal insufficiency due to increased phosphate levels; the resultant lower ionized calcium levels lead to secondary or tertiary hyperparathyroidism.

A careful history is important in the diagnosis of osteomalacia or vitamin D insufficiency. For example, a history of a malabsorptive process such as gastrectomy, intestinal resection, sprue, primary biliary cirrhosis, or pancreatic deficiency may lead to the identification of vitamin D deficiency and osteomalacia.125 Patients with osteomalacia may present with generalized pain involving the pelvis, spine, ribs, or lower extremities or with skeletal deformities such as bowing of the long bones, kyphoscoliosis, or pelvic abnormalities. Another clinical sign of osteomalacia in adults is proximal muscle weakness, which may result in an antalgic or waddling gait and difficulty ambulating.

Pain may be elicited by deep palpation of the tibia, ribs, or pubic ramus.126 One of the radiographic signs of osteomalacia is the presence of pseudofractures, or Looser’s zones, which are transverse lines of rarefaction through the cortices, with incomplete healing in the ribs, scapulae, long bones (Figure 101-2), or pubic rami. Pseudofractures,

---

**Table 101-8 Causes of Osteomalacia and Rickets**

<table>
<thead>
<tr>
<th>Vitamin D Deficiency or Dysfunction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reduced Availability</strong></td>
</tr>
<tr>
<td>Nutritional deficit</td>
</tr>
<tr>
<td>Reduced exposure to ultraviolet light</td>
</tr>
<tr>
<td>Malabsorption (gastrointestinal or biliary disease, surgical resection)</td>
</tr>
<tr>
<td><strong>Alteration in Metabolism</strong></td>
</tr>
<tr>
<td>Reduced 25-hydroxyvitamin D from liver or gastrointestinal disease, nephrotic syndrome, anticonvulsant drugs</td>
</tr>
<tr>
<td>Reduced 1,25-dihydroxyvitamin D from renal disease, vitamin D–dependent rickets type I</td>
</tr>
<tr>
<td><strong>Alteration in Action on Target Tissues</strong></td>
</tr>
<tr>
<td>Vitamin D–dependent rickets type II</td>
</tr>
</tbody>
</table>

**Phosphate Deficiency**

| Decreased availability—dietary deficiency, phosphate-binding antacids |
| Decreased renotubular phosphate reabsorption |
| Familial—X-linked hypophosphatemic rickets, adult-onset vitamin D–resistant osteomalacia |
| Acquired—hypophosphatemic osteomalacia (phosphate diabetes), oncogenic osteomalacia |
| Generalized renotubular disorders |

**Acidosis**

| Renotubular acidosis |
| Ureterosigmoidostomy |
| Carbonic anhydrase inhibitors (acetazolamide) |

**Miscellaneous Mineralization Defects**

| Inhibitors of mineralization—fluoride, bisphosphonates (e.g., etidronate), chronic renal failure (aluminum) |
| Hypophosphatasia |

Osteomalacia may be associated with a deficiency of phosphate, principally in patients with decreased renal tubular reabsorption of phosphate. Familial hypophosphatemic vitamin D–resistant rickets in children or osteomalacia in adults usually presents with renal phosphate leak, hypophosphatemia, rachitic or osteomalacic changes, and inappropriately normal or low-normal 1,25(OH)₂D levels for the degree of hypophosphatemia. This X-linked dominant disorder may present in young children with the inability to walk, followed by progressive bowing and skeletal deformities, without signs of proximal myopathy. The genetic locus for X-linked hypophosphatemic rickets has been mapped to Xp22.1, and the gene is named PHEX (phosphate-regulating gene with homology to endopeptidases on the X chromosome).128

Oncogenic osteomalacia or rickets is a vitamin D–resistant process associated with certain neoplasias, principally small, benign mesenchymal or endodermal tumors and, infrequently, certain malignant tumors (e.g., multiple myeloma; prostatic, oat cell, breast carcinomas).129,130 Such patients typically present with decreased renal tubular phosphate reabsorption, hypophosphatemia, muscle weakness, diminished 1,25(OH)₂D levels, and normocalcemia. The benign tumors tend to be small and difficult to identify on physical examination or radiographs. Surgical removal of these tumors results in a rise in the phosphate and 1,25(OH)₂D levels and resolution of the skeletal process. Osteomalacia may also be associated with generalized renal tubular disorders and the use of certain drugs that contain inhibitors of mineralization (e.g., fluoride, etidronate, aluminum). The evaluation of a patient suspected of having osteomalacia is outlined in Table 101-9.

Osteomalacia is often a treatable disease, but the diagnosis may be overlooked. Vitamin D deficiency can be treated with physiologic doses of vitamin D, but higher doses (1000 to 2000 IU/day) may hasten the healing of bone. In the presence of intestinal malabsorption, and until the underlying malabsorptive process is corrected, large doses of vitamin D (50,000 IU once a week to three or more times a week) are often required. Careful monitoring of the serum and urinary calcium levels and 25-OHD concentrations is necessary to prevent vitamin D intoxication. Use of the active metabolite of 25-OHD (Calderol) may occasionally be necessary in resistant patients or in those with severe liver disease who cannot achieve activation of this metabolite. The potential advantages of using 25-OHD are more stable bioavailability, shorter half-life, and greater potency than the parent compound,129 although the cost is greater.

In patients with hypophosphatemia and disorders of renal tubular phosphate reabsorption, mineralization of bone occurs with phosphate therapy and moderately high doses of 1,25(OH)₂D, the latter being necessary to prevent the secondary hyperparathyroidism associated with phosphate therapy. In patients with renal insufficiency or failure, a phosphate binder (calcium acetate or calcium carbonate) should be used after meals to decrease intestinal phosphate absorption. Calcium citrate therapy should not be used because it augments aluminum absorption. In those with

Table 101-9 Evaluation of Osteomalacia

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcium, phosphorus, alkaline phosphatase, urinary calcium levels; 25-hydroxyvitamin D and intact parathyroid hormone levels</td>
<td>In selected patients: 1.25-Dihydroxyvitamin D levels (e.g., renal insufficiency, vitamin D–resistant osteomalacia or rickets) Vitamin D absorption test: obtain 25-hydroxyvitamin D levels at 0, 4, and 8 hr (e.g., some cases of malabsorption) Tubular reabsorption of phosphate (e.g., vitamin D–resistant osteomalacia or rickets) Bone biopsy with double tetracycline labels</td>
</tr>
</tbody>
</table>
renal failure, 1,25(OH)₂D therapy administered orally, (or intravenously in some dialysis patients) suppresses parathyroid cell secretion and proliferation; a threefold elevation of the PTH level is advocated by some investigators to prevent adynamic bone disease. Analogues of 1,25(OH)₂D that do not produce hypercalcemia but decrease levels of PTH are now available and may be useful in patients with renal insufficiency.

**PAGET’S DISEASE OF BONE**

Paget’s disease of bone affects approximately 2% to 3% of the population older than 50 years and is uncommon in individuals younger than 40 years. Paget’s disease is characterized by enhanced resorption of bone by giant, multinucleated osteoclasts, followed by the formation of disorganized woven bone by osteoblasts. The resultant bone is expanded, weak, and vascular, so affected bones may become enlarged and deformed and the overlying skin may feel warm to the touch.

**Cause**

The cause of Paget’s disease is uncertain, although data showing the presence of viral inclusion particles in giant pagetic osteoclasts support a viral cause, possibly associated with measles, respiratory syncytial, or canine distemper virus. Paget’s disease tends to aggregate in families in an autosomal dominant pattern, and 40% of patients have at least one other family member affected. Recently, in a family with juvenile Paget’s disease, the disease was found to be associated with a polymorphism in the OPG allele. Studies of additional families with this mutation may lead to the discovery of the cause of Paget’s disease.

**Clinical Features**

Many patients with Paget’s disease are asymptomatic, and the disease is detected by the incidental finding of an elevated alkaline phosphatase level or characteristic radiographic abnormality. Other patients present with a range of symptoms that include bone pain; skeletal deformities (bowing of long bones, enlarged skull, pelvic alterations); pathologic fractures; increased cardiac output (with extensive disease); and nerve compression. Paget’s disease typically includes a lytic phase, a combined lytic and blastic phase, and a sclerotic or “burned out” phase occurring late in the disease process.

Radiographic signs of the three stages of Paget’s disease may be present at different sites in the same patient. The skeletal sites commonly involved with Paget’s disease include the skull (Figure 101-3), vertebrae, pelvis, sacrum, and lower extremities. Degenerative joint disease may develop adjacent to the bones and cause pain that may obscure the symptoms associated with Paget’s disease. Ten percent to 30% of patients with Paget’s disease may experience fractures that present initially as asymptomatic or painful short fissure fractures traversing the bony cortex (Figure 101-4). Complete fractures of the bones such as the “chalk stick” fracture also occur; fractures of the long bones may be a serious complication because the increased vascularity of pagetic bone may lead to excessive blood loss.
Neurologic symptoms generally result from compression of the nerves by pagetic bone. Hearing loss is common and is caused by sensory loss and conduction abnormalities due to pagetic involvement of the bones of the inner ear. Paget’s disease of the skull may also produce ocular and other cranial nerve palsies. Compression of the base of the skull may lead to basilar invagination, cerebellar dysfunction, or obstructive hydrocephalus, with symptoms of nausea, ataxy, incontinence, gait disturbances, and dementia. Neurologic compromise of the thoracic or lumbar spine may lead to spinal cord compression or, in the latter instance, cauda equina syndrome.

**Laboratory Findings**

Biochemical indices in patients with Paget’s disease usually show normal serum calcium and phosphate levels, although hypercalcemia may develop with immobilization when there is an uncoupling of bone resorption and formation. Patients with Paget’s disease may also be hypercalcemic if they coincidentally acquire primary hyperparathyroidism, which can further increase bone remodeling and worsen the disease process. Secondary hyperparathyroidism may develop in approximately 10% to 15% of patients with Paget’s disease, presumably because of inadequate calcium intake to meet the skeletal demands of the heightened bone remodeling.\(^ {132} \)

Alkaline phosphatase levels of bone origin (BSAP) are commonly elevated in patients with significant Paget’s disease because of the increased osteoblastic activity combined with bone breakdown. In the absence of liver disease, the alkaline phosphatase level typically correlates with the extent of the pagetic involvement of bone, although it may be more elevated in Paget’s disease of the skull (see Figure 101-3).

Unexpectedly, circulating osteocalcin levels do not reflect disease activity in patients with Paget’s disease as well as bone-specific alkaline phosphatase levels reflect disease activity. Markers of bone resorption such as urinary collagen cross-links like N telopeptides and C telopeptides are also elevated in active Paget’s disease. Other laboratory abnormalities in patients with Paget’s disease include hypercalcemia, hyperuricemia, and hyperuricemia, possibly related to the increased turnover of osteoclasts. Serum uric acid levels should be measured periodically because of the association of Paget’s disease with gouty arthritis.

**Diagnosis**

Bone scans are valuable tools for assessing the extent of Paget’s disease and are therefore useful as part of the initial evaluation.\(^ {133} \) As diagnostic tests, however, bone scans in general are sensitive but not specific for a number of skeletal processes. Radiographs show characteristic radiologic findings such as transverse lucent areas, osteoporosis circumscripta, enlargement of the bones, expanding lytic changes, the “blade-of-grass” lesion shown in Figure 101-4, thickened cortices, a coarse trabecular pattern, or sclerotic changes.

In patients with involvement of the skull and changes in mental status, a skull radiograph, magnetic resonance imaging (MRI), or quantitative computed tomography (QCT) may be useful to diagnose platybasia and flattening of the base of the skull, basilar invagination, or the infrequent complication of hydrocephalus. Audiologic evaluation may reveal hearing loss in patients with pagetic involvement of the skull.

**Treatment**

The indications for treatment of Paget’s disease (Table 101-10) include pain, hypercalcemia, fractures, high-output cardiac failure (rare), and neurologic compromise. Therapy can also be used to prevent the progression of deformity or risk of nerve compression when there is pagetic involvement of the skull, a vertebral body, or a weight-bearing bone (femur) or when disease is present adjacent to a major articular joint.

Treatment of symptomatic Paget’s disease is usually directed at suppression of the enhanced bone resorption and skeletal turnover with calcitonin or bisphosphonates.\(^ {132,135} \) Response to therapy is monitored by the reduction of symptoms and maintenance of the alkaline phosphatase level in a mid-normal range, with retreatment once values rise 25% above normal.

**Calcitonin**

Salmon calcitonin and human calcitonin inhibit the function of osteoclasts, which are active in the pagetic process; both types of calcitonin preparations come in an injectable form and are FDA approved for patients with Paget’s disease.\(^ {136} \) Salmon calcitonin therapy is usually initiated at a low dose to ensure patient tolerance and then increased to a daily dose of 100 Medical Research Council units (intramuscularly or subcutaneously). After 6 months of therapy, the patient may be maintained on 50 to 100 Medical Research Council units daily.\(^ {132} \) Approximately two-thirds of patients show a decrease in alkaline phosphatase levels of 50% or more in 2 to 6 months. Some patients experience resistance to the effects of calcitonin, which can be reversed in some instances by switching from salmon calcitonin to human calcitonin. Calcitonin is a safe drug.

Calcitonin is useful in patients with expanding lytic lesions, particularly of a weight-bearing bone, or for preoperative therapy before elective orthopedic procedures. The use of calcitonin nasal spray has few systemic side effects but is not FDA approved for the treatment of Paget’s disease.

**Bisphosphonates**

Several bisphosphonates are currently approved by the FDA for the treatment of Paget’s disease; they include…

---

**Table 101-10  Indications for Treatment of Paget’s Disease of Bone**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td>Hypercalcemia</td>
</tr>
<tr>
<td>Fractures</td>
<td>High-output cardiac failure (rare)</td>
</tr>
<tr>
<td></td>
<td>Skull involvement</td>
</tr>
<tr>
<td></td>
<td>Neurologic compromise</td>
</tr>
<tr>
<td></td>
<td>Periarticular disease</td>
</tr>
<tr>
<td></td>
<td>Prevention of progression of Paget’s disease</td>
</tr>
</tbody>
</table>
etidronate disodium (Didronel), pamidronate (Aredia), alendronate (Fosamax), tiludronate disodium, risedronate (Actonel), and zoledronic acid (Reclast).

Etidronate disodium is an orally administered drug that produces a clinical and biochemical response similar to that of calcitonin. The therapeutic dose for Paget's disease is 5 mg/kg per day (400 mg/day, or a minimum of 200 mg/day for smaller patients) for 6 months; the drug is then stopped for 6 months before being reintroduced in 6-month cycles of therapy. As mentioned previously, higher doses of etidronate are associated with defective mineralization of bone and osteomalacia, with symptoms of pain and fractures. Measurements of alkaline phosphatase levels at 3- to 6-month intervals are useful to ensure the suppression of bone turnover; an estimated 25% of patients may become resistant to etidronate. However, newer, more potent bisphosphonates are effective for the treatment of Paget's disease.

Parenteral pamidronate is also approved for the treatment of symptomatic Paget's disease of bone in patients with a threefold or greater elevation of alkaline phosphatase concentrations. This more potent bisphosphonate is useful in patients who become resistant to etidronate and in those with more severe disease. An advantage of this therapy is that alkaline phosphatase levels may be reduced to the normal range, with a sustained response for a prolonged period (up to a year or more). The FDA-recommended dose of pamidronate for patients with Paget's disease is 30 mg/day as a 4-hour infusion on 3 sequential days (total dose 90 mg), with retreatment possible if necessary. Other regimens for the treatment of Paget's disease include 60 mg of pamidronate daily (infused over 3 hours) once a week for 1 or 2 weeks in patients with alkaline phosphatase levels between 300 and 400 U/L. For more extensive disease, three or four infusions of pamidronate every 1 to 2 weeks may be necessary. To assess the efficacy of these regimens for Paget's disease, clinical symptoms should be reviewed and alkaline phosphatase levels measured 2 to 3 months later. Some patients treated with pamidronate may experience a transient fever, musculoskeletal and flulike symptoms, and hypocalcemia; calcium supplementation (500 mg twice daily in vitamin D–replete patients) can offset the hypocalcemia that results from the suppression of bone resorption. Pamidronate is available only as a parenteral drug, which restricts its use.

Oral alendronate has been approved for the treatment of Paget's disease at a dose of 40 mg/day for 6 months. Therapy with alendronate is recommended for patients with at least a twofold elevation in alkaline phosphatase levels or for those with specific indications for therapy (see Table 101-10). The use of alendronate produces a normalization of, or a 60% or greater reduction in, the alkaline phosphatase level in approximately 85% of patients. Studies indicate that this therapy is more effective than etidronate. (All bisphosphonates must be taken correctly to minimize gastrointestinal side effects). Risedronate is another potent bisphosphonate (see Table 101-6) that is FDA approved for the treatment of Paget's disease. Siris and colleagues treated 162 patients with moderate to severe Paget's disease with oral risedronate (30 mg/day for 84 days, followed by 112 days without treatment). This cycle was repeated if the serum alkaline phosphatase level did not normalize or increased more than 25% from its nadir value. After the first and second cycles, the serum alkaline phosphatase level decreased 65% and 69%, respectively, and urine markers decreased 50% and 66.9%, respectively. The serum alkaline phosphatase level normalized in 53.8% of patients, and a significant decrease in bone pain was noted. Risedronate is well tolerated and has few adverse effects including a flulike syndrome, gastrointestinal symptoms, and, rarely, iritis. Other groups have shown a decrease in serum alkaline phosphatase levels of 79% and 86%, with an 85% and 100% decrease in urine markers. There is no evidence of osteomalacia in bone biopsies from patients treated with 30 mg of risedronate. Patients who have become resistant to etidronate appear to respond to risedronate. Patients must be instructed to have an adequate intake of calcium and vitamin D while taking risedronate. Recently, zoledronic acid was also found to be effective in the treatment of Paget's disease, and approval is pending from the FDA.

In addition to antiresorptive therapies, nonsteroidal anti-inflammatory drugs (NSAIDs) and aspirin are useful modalities to alleviate the joint pain and other symptoms that result from degenerative joint disease. Finally, surgical intervention is sometimes warranted in patients with Paget's disease and bony deformities, pathologic fractures, nerve compression, or degenerative arthritis. Orthopedic procedures such as total joint replacement and osteotomies are associated with a reduced risk of intraoperative bleeding or other complications if patients are treated medically (e.g., with calcitonin or other bisphosphonates) to reduce the disease activity and vascularity for at least 6 weeks before the procedure.

**OTHER MEDICATION-INDUCED OSTEOPOROSIS**

Aromatase inhibitors in women undergoing breast cancer treatment are associated with bone loss. Postmenopausal women maintain a low level of circulating estrogen because of aromatization of androgens to estrogen in tissues such as fat and muscle by cytochrome P-450 enzyme. Inhibition of this enzyme is now used in postmenopausal women with breast cancer. There are two classes of aromatase inhibitors: nonsteroidal reversible inhibitors (anastrozole and letrozole) and steroidal reversible inhibitors (exemestane). Because these agents prevent the conversion of androgen to estrogen, this results in low serum estrogen levels and increased bone remodeling. Fracture rates in clinical trials of aromatase inhibitors compared with either tamoxifen or placebo ranged from 3% to 7%. In a 2-year study, significantly higher markers of bone turnover and nearly twice the lumbar spine bone loss and fracture rates were observed with anastrozole compared with tamoxifen. Although the data are just beginning to be collected regarding skeletal health in women treated for breast cancer with aromatase inhibitors, it is important to obtain a history of clinical risk factors for osteoporosis and a BMD measurement of the hip and spine. Preventive treatment should be initiated in women with normal or low bone mass and no history of fractures. Treatment of women with low bone mass (t score ≤ −2)
should be initiated with potent antiresorptive agents, and BMD should be monitored at least every 2 years. Recent studies with zoledronic acid and denosumab have been reported to be quite effective in this patient population. 143,144 If a woman continues to lose bone mass on aromatase inhibitors despite compliance with potent antiresorptive agents, and if the patient has not had radiation to the skeleton as part of the breast cancer protocol, rhPTH 1-34 treatment can be used to build up bone mass.

Gonadotropin-releasing hormone antagonists are used to treat women with endometriosis and men with prostate cancer. These compounds induce bone loss by lowering estrogen levels, resulting in accelerated bone turnover. A study of more than 50,000 men with prostate cancer treated with androgen-deprivation therapy consisting of either gonadotropin-releasing hormone agonists or orchiectomy found an increased risk of fracture or hospitalization due to fracture. Although the overall risk of fracture associated with androgen-deprivation therapy was only modestly increased overall, the risk of fracture was significantly associated with the number of doses of gonadotropin-releasing hormone agonists. Other studies have reported that androgen-deprivation therapy increases bone loss at all sites, with a 2% to 8% annual loss in the lumbar spine and a 2% to 6% loss in the hip after 1 year. 145 Given the high incidence of prostate cancer and the increasing use of this treatment, assessment of bone mass and the prevention of additional bone mass loss are probably appropriate. Oncologists currently recommend that BMD be measured at the time of initiation of androgen-deprivation therapy and that clinical risk factors for osteoporosis be reviewed including history of fracture after age 30, family history of hip fracture, smoking history, use of glucocorticoids, low testosterone level, and rheumatoid arthritis. If the patient has a low BMD (t score < −2.5) or a t score between −1 and −2.5 and other risk factors, treatment with calcium and vitamin D supplementation and a bisphosphonate (zoledronic acid, alendronate, risedronate, or denosumab) should be initiated. BMD of the lumbar spine and hip should be measured at least once a year while patients are maintained on androgen-deprivation therapy. 146–149

Selected References

115. Behman Q, Lang T, Lane NE: Daily treatment with parathyroid hormone is associated with an increase in vertebral cross-sectional area in postmenopausal women with glucocorticoid-induced osteoporosis, Osteoporos Int 41:374–382, 2003.

Full references for this chapter can be found on www.expertconsult.com.
References


141. Reid IR, Miller P, Lyles K, et al: Comparison of a single infusion of
tozoledronic acid with risendronate for Paget's disease, N Engl J Med
tor on BMD and bone turnover markers: 2-year results of the
Anastrozole, Tamoxifen, Alone or in Combination (ATAC) trial
effectively prevents aromatase inhibitor-associated bone loss in post-
menopausal women with early breast cancer receiving adjuvant letro-
zole: Z-fast study 36 month follow-up, Clin Breast Cancer
144. Ellis GK, Bone HG, Chiebowski R, et al: Randomized trial of deno-

GNRH agonist therapy for prostate cancer, J Urol 175:136–139,
2006.
with prostate carcinoma receiving androgen-deprivation therapy:
147. Shahinian VB, Kuo YF, Freeman JL, Goodwin JS: Risk of fracture
164, 2005.
receiving androgen-deprivation therapy for prostate cancer, N Engl J
149. Campbell SC, Bhoopalam N, Moritz TE: The use of zoledronic acid
in men receiving androgen deprivation therapy for prostate cancer
with severe osteopenia or osteoporosis, Urology 75:1138–1143, 2010.
Proliferative bone diseases encompass a variety of conditions characterized by exuberant bone and enthesal ossifications and calcifications. New bone formation is the main feature in diffuse idiopathic skeletal hyperostosis (DISH) and hypertrophic osteoarthropathy (HOA) and is a common finding in osteoarthritis. New bone formation may also accompany some spondyloarthropathies such as ankylosing spondylitis, psoriatic arthritis, and sternoclavicular syndrome, also known as SAPHO (synovitis, acne, pustulosis, hyperostosis, and osteomyelitis). New bone formation has also been described in endocrine diseases, however, such as thyroid disorders, acromegaly, and hypoparathyroidism (Table 102-1). Osteoarthritis, spondyloarthropathies, and endocrine disorders are discussed elsewhere in this book.

**DIFFUSE IDIOPATHIC SKELETAL HYPEROSTOSIS**

DISH is a condition characterized by calcification and ossification of soft tissues, mainly ligaments and entheses. This condition was described by Forestier and Rotes-Querol in 1950 and was termed *senile ankylosing hyperostosis*. There is a marked predilection to the axial skeleton, particularly the thoracic spine. Recognition that the condition is not limited to the spine and may involve peripheral joints led researchers to coin the name DISH, a term now widely used.3

DISH is characterized by the production of coarse, flowing osteophytes involving, in particular, the right side of the thoracic spine with preservation of the intervertebral disk space and by ossification of the anterior longitudinal ligament. Calcification and ossification of the posterior longitudinal ligament seem to be additional skeletal manifestations of DISH. Other enthesal regions might be affected such as the peripatellar ligaments, Achilles tendon insertion, planter fascia, olecranon, and others.6-8

The diagnosis is usually based on the definition suggested by Resnick and Niwayama.5 This radiographic approach requires the presence of flowing, coarse osteophytes on the right side of the thoracic spine, connecting at least four contiguous vertebrae, or ossification of the anterior longitudinal ligament, preserved intervertebral disk height in the involved segment, and the absence of apophyseal joint ankylosis and sacroiliac joint involvement (Table 102-2).8 Another set of criteria, for epidemiologic purposes, was suggested by Utsinger.1 These criteria consider also peripheral enthesopathies. A definite diagnosis of DISH is established by criteria similar to those suggested by Resnick and Niwayama. A probable diagnosis of DISH is possible, however, with continuous ossification or calcification, or both, of the anterolateral aspect of at least two contiguous...
The association of DISH with excess body weight has been well known since the early descriptions by Forestier and others. This association was reiterated in a study in which patients with DISH were compared with healthy individuals and patients with spondylosis. More often, DISH was reported to be associated with more complex metabolic and endocrine derangements, with or without overt type II diabetes mellitus, comprising glucose intolerance, hyperinsulinemia, dyslipidemia, hyperuricemia, and elevated growth hormone levels, elevated insulin-like growth factor (IGF)-1, hyperuricemia, use of retinoids, and genetic factors (Table 102-3). The association of DISH with excess body weight has been well known since the early descriptions by Forestier and others. This association was reiterated in a study in which patients with DISH were compared with healthy individuals and patients with spondylosis. More often, DISH was reported to be associated with more complex metabolic and endocrine derangements, with or without overt type II diabetes mellitus, comprising glucose intolerance, hyperinsulinemia, dyslipidemia, hyperuricemia, and elevated growth hormone levels, elevated insulin-like growth factor (IGF)-1, hyperuricemia, use of retinoids, and genetic factors (Table 102-3). The association of DISH with excess body weight has been well known since the early descriptions by Forestier and others. This association was reiterated in a study in which patients with DISH were compared with healthy individuals and patients with spondylosis. More often, DISH was reported to be associated with more complex metabolic and endocrine derangements, with or without overt type II diabetes mellitus, comprising glucose intolerance, hyperinsulinemia, dyslipidemia, hyperuricemia, and elevated growth hormone levels, elevated insulin-like growth factor (IGF)-1, hyperuricemia, use of retinoids, and genetic factors (Table 102-3). The association of DISH with excess body weight has been well known since the early descriptions by Forestier and others. This association was reiterated in a study in which patients with DISH were compared with healthy individuals and patients with spondylosis. More often, DISH was reported to be associated with more complex metabolic and endocrine derangements, with or without overt type II diabetes mellitus, comprising glucose intolerance, hyperinsulinemia, dyslipidemia, hyperuricemia, and elevated growth hormone levels, elevated insulin-like growth factor (IGF)-1, hyperuricemia, use of retinoids, and genetic factors (Table 102-3).
osteoblasts.\textsuperscript{19,26,32,33} ALP promotes the calcification process during bone formation and is considered a good indicator for the maturation stages of osteoblasts.\textsuperscript{14,35} There is no explanation yet as to why the new bone formation is localized mainly at the ligamentous and enthesal sites. In male DISH patients, growth hormone serum levels were not elevated in the serum but were much higher in the synovial fluid.\textsuperscript{36} In the spine, vertebral blood supply could be a factor in the onset or progression of DISH.\textsuperscript{37} Intraerythrocyte growth hormone serum levels may exceed serum growth hormone levels and could be transported to the vertebral site by the mechanism described by Denko and colleagues.\textsuperscript{36}

The expression of various genes involved in cell division and growth is regulated by nuclear factor κB (NFκB), which is capable of regulating the differentiation of multipotential cells. It was shown that activation of environmental factors such as platelet-derived growth factor (PDGF)-BB and transforming growth factor (TGF)-β1 in ligament cells stimulates the activation of NFκB, which influences the osteoblastic differentiation of mesenchymal cells. This event is accompanied by elevation of ALP activity in cells of patients with DISH and serves as an indicator of maturation stages of the osteoblast.\textsuperscript{34} Inflammatory cytokines such as PDGF-BB, TGF-β1, and others may be related to the onset of non–insulin-dependent diabetes mellitus and may be the link between this condition and the occurrence of DISH.\textsuperscript{36,39}

Vitamin A and its derivatives have been implicated in the pathogenesis of DISH owing to their ability to promote new bone formation. Levels of vitamin A were reported to be higher in patients with DISH compared with controls, and some reports showed DISH-like manifestations in young patients treated with vitamin A or its derivatives.\textsuperscript{20,40} The role of vitamin A is unclear, however, because more recent studies did not show an increased prevalence of DISH among patients treated with vitamin A in various dosages and for various lengths of time.\textsuperscript{22,25,28} Larger prospective studies are necessary to elucidate the role played by this vitamin in the pathogenesis of DISH.

Familial clustering of DISH or families with early presentation of DISH suggest a genetic background for this disorder.\textsuperscript{41,42} Ossification of the posterior longitudinal ligament is closely related to DISH, and the two conditions can coexist. COL6A1, which is the candidate gene for ossification of the posterior longitudinal ligament, was reported to be significantly associated with DISH in Japanese, but not in Czech, patients.\textsuperscript{43,44} This finding would suggest that other factors might play an important role in the genetic predisposition for the development of DISH.

There are no convincing explanations for the predilection of the hyperostotic process to affect the anterolateral aspect of the thoracic spine. The limited range of motion of the thoracic spine has been cited as a possible cause for the predilection to this site. This assumption cannot explain the involvement of the extremely mobile cervical spine or the lumbar spine, however. The less frequent involvement of the left side of the thoracic spine was ascribed to the pulsation of the aorta. This assumption was based on a few reports that described left-sided bridging osteophytes in cases with a right-sided aorta, suggesting that the aortic pulsations interfere with the production of the osteophytes.\textsuperscript{45} Calcifications, ossification, and subsequent stiffening of ligaments and joint capsules have important pathogenetic implications.

Osteoarthritis may have pathogenetic features in common with the peripheral joint manifestations of DISH. It was suggested that in the small non–weight-bearing joints in osteoarthritis, the process is caused by an increased intra-articular pressure and subsequent development of “crash” forces.\textsuperscript{46} This development was attributed to thickening of the collateral ligaments of these joints that enforce a constraint movement, not to primary damage to the cartilage. It seems reasonable that the joints affected by DISH may develop the same “crash” forces operating in small osteoarthritis joints, as a result of this mechanism. This mechanism might explain the involvement of “atypical” joints, not commonly affected by osteoarthritis, and the hypertrophic osteoarthritic changes in the commonly affected joints.

**Clinical Manifestations**

The lack of specific symptoms and signs of DISH, as well as the radiographic diagnostic criteria, have raised doubts about DISH as a separate entity.\textsuperscript{47} Although the disease may be asymptomatic, it was reported to be associated with morning stiffness, dorsolumbar pain, and reduced range of motion in most patients.\textsuperscript{25} Patients with DISH may have extremity pain involving peripheral large and small joints and peripheral entheses such as the heel, Achilles tendon, shoulder, patella, and olecranon. Pain in the axial skeleton may involve all three segments of the spine and the costosternal and sternoclavicular joints. The level of pain and disability is significantly higher compared with healthy subjects but is not different from patients with spondylitis.\textsuperscript{28} Complaints of pain referable to the thoracic spine are common and are accompanied by a reduced chest expansion.

Although similar in some aspects to osteoarthritis of the spine, DISH is a distinct clinical entity with different characteristics.\textsuperscript{48} Classically, the portions of the spine that are involved in osteoarthritis are the lower portions of the cervical spine and the lumbar spine. Thoracic spine involvement is uncommon in osteoarthritis or occurs in late stages of the disease, as opposed to the common involvement of the thoracic spine in DISH. Thoracic spine involvement in DISH is characterized by preserved intervertebral height, whereas in spinal osteoarthritis, reduced intervertebral disk height is common. These differences in the radiologic appearance and anatomic spinal distribution probably have to do with the different pathogenetic mechanisms described earlier. It is presumed that the primary target for the osteoarthritic process is the cartilage represented in the spine by the intervertebral disks and the cartilage of the facet joints.

The wear-and-tear forces operating in the extremely mobile lower cervical and lumbar portions of the spine might explain the frequent involvement of these segments in osteoarthritis, whereas the thoracic spine is the least mobile of the spinal segments. The main targets of the disease in DISH are the spinal ligaments and entheses (Figure 102-1).\textsuperscript{1,49} These abnormalities are not limited to the thoracic spine and may involve the lumbar spine and the cervical spine (Figure 102-2). In the lumbar spine, the large bridging osteophytes are not uniformly one-sided.\textsuperscript{50}
spine is affected including dysphagia, hoarseness, stridor, ossification of the posterior longitudinal ligament, myelopathy, aspiration pneumonia, sleep apnea, atlantoaxial complications, thoracic outlet syndrome, esophageal obstruction, endoscopic and intubation difficulties, and fractures (Table 102-4).53 The high prevalence of coexisting intervertebral disk damage in young patients with DISH suggests an important role for DISH in the pathogenesis of spondylitis in this group of patients.54 At times, patients with DISH may assume the typical postural abnormalities of advanced ankylosing spondylitis such as accentuated kyphosis and reduced mobility of the spine. These two entities, although they may coexist, can usually be distinguished by the different age of onset, clinical presentation, radiographic appearance of the spine and sacroiliac joints, and HLA-B27 associations.55,56

Clinical manifestations similar or identical to those of osteoarthritis are prominent features of DISH in the peripheral joints. The peripheral joints affected by DISH have features that distinguish them from primary osteoarthritis, however. One is the more frequent involvement of joints that are not usually affected in osteoarthritis such as the metacarpophalangeal joints, elbows, and shoulders (Figure 102-3).55-66 Other features are a more severe hypertrophic disease that may result in a reduced range of motion in the affected joints and calcified and/or ossified prominent enthesopathies.59,61,62

As described previously, the primary event in DISH is thickening, calcification, or ossification of ligaments and entheses. In particular, enthesopathy affecting the peripheral joints has been described.54 The radiographic appearance of peripatellar, cruciate ligament insertion, and pericapsular osseous enthesopathies are some examples of the contribution of DISH to stiffening of the soft tissues surrounding a joint (Figure 102-4).64 Enthesal ossification at various sites other than joints such as the heel, ribs, and pelvis is a common finding in DISH. These enthesopathies may become symptomatic exhibiting pain and swelling in the affected region. A high probability for the presence of spinal DISH was noted for ossification of the iliolumbar and sacrotuberous ligaments and with bony overgrowth of the inferior acetabular rim.63,65 The tendency for new bone for-

### Table 102-4  Clinical Manifestations of Diffuse Idiopathic Skeletal Hyperostosis in the Cervical Spine

<table>
<thead>
<tr>
<th>Spontaneous</th>
<th>Induced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dysphagia</td>
<td>Endoscopic problems</td>
</tr>
<tr>
<td>Hoarseness</td>
<td>Intubation difficulties</td>
</tr>
<tr>
<td>Stridor</td>
<td>Fractures</td>
</tr>
<tr>
<td>Ossification of posterior longitudinal ligament</td>
<td></td>
</tr>
<tr>
<td>Myelopathy</td>
<td></td>
</tr>
<tr>
<td>Aspiration pneumonia</td>
<td></td>
</tr>
<tr>
<td>Sleep apnea</td>
<td></td>
</tr>
<tr>
<td>Atlantoaxial complications (pseudarthrosis, subluxation)</td>
<td></td>
</tr>
<tr>
<td>Thoracic outlet syndrome</td>
<td></td>
</tr>
</tbody>
</table>

Table 102-5  Therapeutic Targets in Diffuse Idiopathic Skeletal Hyperostosis

<table>
<thead>
<tr>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptomatic relief of pain and stiffness</td>
</tr>
<tr>
<td>Prevent, retard, or arrest progression</td>
</tr>
<tr>
<td>Treatment of associated metabolic disorder</td>
</tr>
<tr>
<td>Prevent spontaneous complications</td>
</tr>
<tr>
<td>Prevent traumatic complications</td>
</tr>
<tr>
<td>Prevent complications that might emerge during diagnostic or therapeutic procedures</td>
</tr>
</tbody>
</table>


Described earlier put patients at an increased risk for cardiovascular diseases. The diagnosis of DISH should be suspected in patients with osteoarthritis in atypical locations (e.g., elbow), in patients with hypertrophic osteoarthritis, and in patients with large enthesopathies and entrapment neuropathies of uncertain origin. This is particularly true for patients with the associated diseases and metabolic abnormalities discussed earlier. It has been shown that chest radiographs might serve as a screening tool for the diagnosis of DISH with a sensitivity of 77% and specificity of 97%.

### Treatment

Treatment of DISH should address several issues. Treatment is expected to alleviate pain and stiffness; prevent, retard, or arrest progression; correct the associated metabolic disorders; and prevent spontaneous or induced complications (Table 102-5).

Specific therapeutic interventions in DISH have not been systematically explored; this is probably related to described earlier put patients at an increased risk for cardiovascular diseases. The diagnosis of DISH should be suspected in patients with osteoarthritis in atypical locations (e.g., elbow), in patients with hypertrophic osteoarthritis, and in patients with large enthesopathies and entrapment neuropathies of uncertain origin. This is particularly true for patients with the associated diseases and metabolic abnormalities discussed earlier. It has been shown that chest radiographs might serve as a screening tool for the diagnosis of DISH with a sensitivity of 77% and specificity of 97%.

### Treatment

Treatment of DISH should address several issues. Treatment is expected to alleviate pain and stiffness; prevent, retard, or arrest progression; correct the associated metabolic disorders; and prevent spontaneous or induced complications (Table 102-5).

Specific therapeutic interventions in DISH have not been systematically explored; this is probably related to...
the inclusion of DISH in the spectrum of osteoarthritis and the assumption that the same therapeutic interventions for osteoarthritis are suitable for DISH. It was suggested more recently that serum levels of growth hormone and IGF-I might be a useful surrogate marker for assessing DISH progression and remission. It was estimated that a period of at least 10 years is necessary for the pathologic process to evolve completely. This notion implies that a long observation period is necessary to show that a therapeutic intervention prevents the development of the disease, arrest its progression, or, it is hoped, reverses the pathologic changes.

There are few reports about remedies to alleviate the symptoms of the disease, but some investigators have reported on the beneficial effects of light exercise, heat, analgesics, and nonsteroidal anti-inflammatory drugs (NSAIDs). Exercise therapy failed to show a significant improvement in the spinal range of motion except for lumbar sacral flexion. More recently, the use of locally acting NSAIDs for the treatment of osteoarthritis was shown to be as effective as the same product by oral route, suggesting that locally acting NSAIDs also might be successfully employed for the symptomatic relief of pain and stiffness in the peripheral joints of patients with DISH. Treatment of symptomatic enthesopathies might be necessary to alleviate local pain and swelling; this can be achieved by local soft tissue applications such as insoles for plantar spurs or protective bandages at other sites. Infiltration of local anesthetic with long-acting corticosteroids might offer at least temporary relief in severely symptomatic cases. When multiple sites are involved, the same therapeutic modalities mentioned for osteoarthritis may be used.

The coexistence of many cardiovascular risk factors places patients with DISH at a higher risk for cardiovascular complications. It seems appropriate to screen these patients for known cardiovascular risk factors and to treat when appropriate. General measures such as weight reduction, adequate physical activity, and a diet low in saturated fat and carbohydrates all might be important in preventing or arresting the progression of DISH. Some of these factors may have pathogenetic implications and may become therapeutic targets. On the basis of present understanding, therapeutic interventions should aim at a reduction of insulin secretion and insulin resistance. In patients with non–insulin-dependent diabetes mellitus, the use of biguanides, which decrease insulin resistance, may offer an advantage over the use of sulfonylureas, which increase insulin secretion. When coexisting hypertension should be treated, medications that might improve insulin resistance such as angiotensin-converting enzyme inhibitors, calcium channel blockers, and α-blockers should be preferred to medications that might worsen insulin resistance such as thiazide diuretics and β-blockers. Some growth factors that might have a role in the development of DISH such as NFκB, PDGF-BB, TGF-β1, growth hormone, and IGF-I may become targets someday for specific therapeutic interventions.

Some complications can be avoided if taken into consideration. Aspiration pneumonia can be partially avoided if instructions in proper deglutition and preservation of an upright position after meals are carefully explained to the patient. Physicians familiar with DISH can avoid or minimize damage to the cervical spine or to soft tissues in patients who might need certain diagnostic or therapeutic interventions such as upper gastrointestinal endoscopy or endotracheal intubation. It is reasonable to adopt the common measures to prevent falls and trauma, especially in elderly patients. Heterotopic ossification after orthopedic surgeries, in particular hip arthroplasty, is common in patients with DISH. Several therapeutic interventions aimed at abolishing heterotopic ossification such as administration of NSAIDs, anti–vitamin K, and irradiation, have been reported with variable success. Patients in the high-risk group to develop this complication, such as patients with DISH, should be considered candidates for one of these regimens. The therapeutic options are summarized in Figure 102-5.

Many tasks lay ahead to define better, understand the pathogenesis, and delineate effective interventions for this disorder (Table 102-6).

### HYPERTROPHIC OSTEOARTHROPATHY

HOA is a well-known entity characterized by skin and bone proliferation. The hallmark and main visual manifestation is a bulbous deformity of the distal end of the digits, also known as clubbing or drumsticks. Periostosis is a progressive process with predilection for the tubular bones, principally the tibia and fibula. Periostosis is bilateral; is symmetric; and spares the medullary cavity, the axial skeleton, and the skull. The prevalence of the condition is unknown, but it was found in skeletal human remains dating thousands of years ago. It may be primary or secondary to many other diseases.

#### Etiology

Primary HOA is an autosomal dominant disorder characterized by periostosis, clubbing, thickening of the skin of the face and scalp, seborrhea, and hyperhidrosis. It is also termed pachydermoperiostosis. There is a male predominance with a male-to-female ratio of 9:1 with one peak of presentation in the first year of life and the other in adolescence.

Secondary forms of HOA may manifest as isolated clubbing or with the full spectrum of the disease. Clubbing may be unilateral or bilateral and has been reported to occur in a variety of diseases including pulmonary (most commonly cancers), cardiac, gastrointestinal, neurologic, infectious, vascular, and other diseases (Table 102-7).

#### Pathogenesis

HOA is characterized by excessive collagen deposition, endothelial hyperplasia, edema, and new bone formation involving mainly the distal extremity and eventually

<table>
<thead>
<tr>
<th>Table 102-6 Future Considerations in Diffuse Idiopathic Skeletal Hyperostosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Establish and validate diagnostic criteria that consider also the peripheral manifestations of the disease</td>
</tr>
<tr>
<td>Clarify the natural course and prognosis</td>
</tr>
<tr>
<td>Study the systemic nature and the impact on quality of life and life expectancy</td>
</tr>
<tr>
<td>Seek a better understanding of the pathogenetic basis for the disease</td>
</tr>
<tr>
<td>Offer a disease-modifying therapeutic approach</td>
</tr>
</tbody>
</table>

---

<sup>1</sup> 1675

<sup>2</sup> 1685

<sup>3</sup> 1676

<sup>4</sup> 1677

<sup>5</sup> 1678

<sup>6</sup> 1679

<sup>7</sup> 1680

<sup>8</sup> 1681

<sup>9</sup> 1682

<sup>10</sup> 1683

<sup>11</sup> 1684

<sup>12</sup> 1685

<sup>13</sup> 1686

<sup>14</sup> 1687

<sup>15</sup> 1688

<sup>16</sup> 1689

<sup>17</sup> 1690

<sup>18</sup> 1691

<sup>19</sup> 1692

<sup>20</sup> 1693

<sup>21</sup> 1694

<sup>22</sup> 1695

<sup>23</sup> 1696

<sup>24</sup> 1697

<sup>25</sup> 1698

<sup>26</sup> 1699

<sup>27</sup> 1700

<sup>28</sup> 1701

<sup>29</sup> 1702

<sup>30</sup> 1703

<sup>31</sup> 1704

<sup>32</sup> 1705

<sup>33</sup> 1706

<sup>34</sup> 1707

<sup>35</sup> 1708

<sup>36</sup> 1709

<sup>37</sup> 1710

<sup>38</sup> 1711

<sup>39</sup> 1712

<sup>40</sup> 1713

<sup>41</sup> 1714

<sup>42</sup> 1715

<sup>43</sup> 1716

<sup>44</sup> 1717

<sup>45</sup> 1718

<sup>46</sup> 1719

<sup>47</sup> 1720

<sup>48</sup> 1721

<sup>49</sup> 1722

<sup>50</sup> 1723

<sup>51</sup> 1724

<sup>52</sup> 1725

<sup>53</sup> 1726

<sup>54</sup> 1727

<sup>55</sup> 1728

<sup>56</sup> 1729

<sup>57</sup> 1730

<sup>58</sup> 1731

<sup>59</sup> 1732

<sup>60</sup> 1733

<sup>61</sup> 1734

<sup>62</sup> 1735

<sup>63</sup> 1736

<sup>64</sup> 1737

<sup>65</sup> 1738

<sup>66</sup> 1739

<sup>67</sup> 1740

<sup>68</sup> 1741

<sup>69</sup> 1742

<sup>70</sup> 1743

<sup>71</sup> 1744

<sup>72</sup> 1745

<sup>73</sup> 1746

<sup>74</sup> 1747
progressing proximally. Various hypotheses have been generated in an attempt to explain the development of HOA. Most cases with secondary HOA have severe lung or cyanotic heart diseases. It was suggested that megakaryocytes are fragmented into platelets during their passage in the lung capillaries. In severe lung diseases or right-to-left shunts, megakaryocytes or platelet aggregates bypass the pulmonary capillary bed, however, and lodge in the peripheral vasculature of the digits.

It was shown that locally released growth factors such as vascular endothelial growth factor (VEGF) and PDGF were remarkably increased in tissue samples obtained from digits of patients with HOA. It is feasible that these substances might be responsible for the distal overgrowth of collagen and bone. VEGF was found to be produced by a lung tumor in a patient with HOA. In this case, serum levels of VEGF were high and resection of the tumor reversed the digital clubbing and reduced the serum VEGF levels.

Activation of platelets and endothelial cells was supported by an increase in circulating von Willebrand factor antigen. Other growth factors have been associated with digital clubbing including hepatocyte growth factor, which was found to be increased in the serum of patients with lung cancer and HOA compared with patients with lung cancer without HOA.

**Clinical Manifestations**

Often HOA is asymptomatic, and sometimes it is the patient who notes the changes in the shape of the fingers. Symptomatic patients complain about a deep-seated pain in the lower extremity and over the long tubular bones, which is exacerbated by palpation. Large joint effusions are
common, and the synovial fluid is thick with few white blood cells.\textsuperscript{90} Skin hypertrophy may be confined to the nail beds or involve the face or larger areas overlying the tubular bones or joints.

The most common and apparent clinical manifestation is digital clubbing. The bulbous deformity of the fingertips is accompanied by a convex nail (watch-crystal nail). The skin around the nail bed becomes shiny and thin with disappearance of the creases (Figure 102-6). Palpation of the base of the nail bed yields the sensation of a “floating” nail within the soft tissue. Cases of advanced clubbing can be identified easily. Several methods were developed, however, to diagnose early phases of the condition. Among those techniques, the digital index and the phalangeal depth ratio have been most widely used.\textsuperscript{91,92} The digital index measures the ratio between the circumference at the level of the nail bed and the circumference at the distal interphalangeal joint of the 10 fingers. The sum of ratios greater than 10 suggests clubbing. The phalangeal depth ratio measures the ratio between the depth of the distal phalanx and the depth of the distal interphalangeal joint of the index finger. A ratio greater than 1 is considered abnormal.

There are no specific laboratory tests to diagnose HOA. Radiographs of the fingers and toes may show acro-osteolysis, and periostitis manifest by cortical thickening of long bones is often observed. The process may involve few or multiple sites and can be regular or irregular in appearance. Characteristically, there is no reduction in joint space or erosions. Radioisotope bone scanning can be useful for diagnosis and for evaluating the extent of the process. Increased uptake can be seen in the cortex of long bones sometimes in the form of splints (Figure 102-7).

\textbf{Treatment Considerations}

Asymptomatic cases need no treatment. NSAIDs are sometimes useful in symptomatic patients. Case reports have suggested that significant pain relief was observed after treatment with octreotide or pamidronate.\textsuperscript{93,94} In secondary cases of HOA, all features and symptoms promptly regress with successful treatment of the primary disease such as correction of a heart malformation, removal of tumors, and therapy of infective endocarditis or inflammatory bowel disease. The role played by VEGF in the pathogenesis of HOA suggests that treatment with VEGF inhibitors might improve the symptoms of this condition.\textsuperscript{95}

\textbf{SAPHO SYNDROME}

SAPHO syndrome is a chronic inflammatory disorder, often relapsing, of unknown etiology. The term SAPHO was coined in order to include several associated clinical manifestations: synovitis, acne, pustulosis, hyperostosis, and osteitis.\textsuperscript{95} The most common sites of bone involvement are in the anterior chest wall, mainly the clavicles, sternum, and sternoclavicular joint. The syndrome has been associated with a variety of skin lesions such as acne conglobata, acne fulminans, palmoplantar pustulosis, and psoriasis. The condition is uncommon with an estimated prevalence of 1 in 10,000, but the exact prevalence is difficult to establish in the absence of validated diagnostic or classification criteria. Furthermore, it is presumed that the prevalence of SAPHO syndrome is underestimated, in particular with mild or absent skin manifestations.\textsuperscript{95}

\textbf{Etiology and Pathogenesis}

The etiology of SAPHO syndrome is unknown. Infectious etiology has been proposed on the basis of the isolation of \textit{Propionibacterium acnes}, a slowly growing anaerobic microorganism often found in acne, from sternal osteosclerotic lesions.\textsuperscript{96} It has been postulated that \textit{P. acnes} might act directly inducing bony erosions by chronic indolent inflammation. It has also been suggested that \textit{P. acnes} may trigger complement activation and promote production of...
 interleukin (IL)-1, IL-8, and tumor necrosis factor (TNF), which contribute to both humoral and cellular proinflammatory responses. An imbalance between proinflammatory and anti-inflammatory mediators has been also suggested to contribute to the inflammatory response and the subsequent damage.\(^{100,101}\)

Genetic susceptibility has not yet been found. No associations with HLA-B27, other alleles often associated with psoriatic arthritis, or other candidate genes have been identified.\(^ {97,102,103}\)

Inconsistencies in the reported prevalence of common autoantibodies encountered in autoimmune diseases, in patients with SAPHO syndrome, preclude at present any firm evidence for their role in the development of this condition.\(^ {104}\) It has been suggested that the increased association of SAPHO syndrome with inflammatory bowel disease might link SAPHO to the seronegative spondyloarthropathies, but this view has not yet been confirmed.\(^ {97,105}\)

Clinical Manifestations and Imaging Findings

SAPHO syndrome is characterized by the combination of skin and osteoarticular manifestations. The skin manifestations include palmpoplantar pustulosis, severe acne, suppurrative hidradenitis, and at times psoriasis. The skin manifestations can anticipate or follow the osteoarticular manifestations, at times by many years.\(^ {97,106,107}\)

The osteoarticular features of SAPHO syndrome are hyperostosis, aseptic osteitis occasionally involving the adjacent joints, and synovitis. The early histologic findings of the bone lesions are similar to those of osteomyelitis with periosteal bone formation. Subsequently the lesions evolve into chronic inflammation with preponderance of mononuclear infiltrate, and only in the late phases prominent marrow fibrosis, sclerosis, and enlarged bone trabeculae ensue.\(^ {108}\)

The most characteristic clinical manifestation is pain in the anterior chest wall due to common involvement of the clavicles, sternum, and the sternoclavicular joints.\(^ {97,107}\) Other common sites of involvement in the axial skeleton are the vertebral bodies with vertebral sclerosis, hyperostosis, spondylodiskitis, nonmarginal syndesmophytes, at times with anterior bridging; the bone adjacent to the sacroiliac joint; and the pubic symphysis.\(^ {105,109}\) Extraspinal involvement is infrequent but may be observed in the form of osteitis involving the tibia, femur, or mandible and with arthritis involving knees, hips, ankles, or small joints of hands and feet.\(^ {97,105,110}\) The diagnosis is highly suspected in cases with usually sterile, multifocal, recurrent osteomyelitis, with or without cutaneous lesions; arthritis associated with palmpoplantar pustulosis, pustular psoriasis, or severe acne; and osteitis associated with severe acne, palmpoplantar pustulosis, or pustular psoriasis.\(^ {107}\)

Laboratory investigations may reveal moderately elevated erythrocyte sedimentation rate (ESR) or acute phase reactants such as CRP and C3 and C4. These findings may reflect the inflammatory nature of the condition but are less reliable than in other inflammatory rheumatic disorders.\(^ {97,105}\)

The diagnosis is usually established on the correlation between the clinical presentation and imaging findings. Bone scintigraphy is sensitive to detect the anterior chest wall involvement demonstrating the characteristic “bull’s head” sign.\(^ {111}\) Plain radiographs may detect advanced lesions, but computed tomography (CT) scan enhances the ability to detect these lesions in particular in flat bones. Radiographic studies of the anterior chest lesions show osteitis in the form of osteosclerosis with homogeneous fibillary pattern, hyperostosis in the form of periosteal reaction, and cortical thickening leading to bone hypertrophy. These lesions with eventual erosive arthritis (due to either primary arthritis or extension of the adjacent osteitis), classically involve the sternoclavicular joint, the upper costosternal and manubriosternal junctions. The spine, being involved in about a third of the patients, presents with similar radiographic appearance with vertebral sclerosis, hyperostosis, spondylodiskitis, nonmarginal asymmetric syndesmophytes, and at times hyperostosis in the form of osseous bridging along the anterior aspect of the spine.\(^ {106}\) In a recent magnetic resonance imaging (MRI) study, corner vertebral erosions were a consistent finding, frequently with involvement of the vertebral end plates and at times adjacent vertebra.\(^ {112}\)

The meaning of these findings has not yet been fully elucidated, and further studies are necessary. Enthesial ossifications, osteitis, osteosclerosis, and periosteal new bone formation have all been reported with variable frequency in long bones.

Treatment

Nonsteroidal anti-inflammatory drugs (NSAIDs) are usually the first line of treatment and may improve the symptoms. However, they often fail to control the disease and additional therapy with corticosteroids and other modalities is required.\(^ {97,113}\) Several empiric therapeutic approaches have been reported in SAPHO including NSAIDs, corticosteroids, bisphosphonates, sulfasalazine, methotrexate, antibiotics, and anti-TNF agents.\(^ {113}\) Second-line drugs including methotrexate, cyclosporine, sulfasalazine, and leflunomide have been tried with mixed results.\(^ {105,113,114}\) Treatment with antibiotics, usually azithromycin, has been reported to be beneficial in patients with SAPHO syndrome. However, many patients were also treated with other disease-modifying antirheumatic drugs and the response faded after discontinuation of the antibiotic. It is not clear whether the response to the antibiotic, though transitory, is due to its antibacterial or anti-inflammatory effect.\(^ {115}\) In recent years, bisphosphonates, in particular pamidronate have been reported to elicit a good or partial response in the majority of patients affected by SAPHO syndrome.\(^ {116}\) It seems that beyond their ability to prevent bone resorption, bisphosphonates are capable of suppressing the production of proinflammatory cytokines such as IL-1, TNF, and IL-6.\(^ {117}\) Treatment with TNF blockers has emerged as an effective therapy in SAPHO syndrome. The results appear to be less spectacular than in other seronegative spondyloarthropathies and may be hampered by relapse of the skin manifestations.\(^ {118,119}\)

References


The references for this chapter can also be found on www.expertconsult.com.
Osteonecrosis literally means “bone death” (ossa [Latin] = bone; necrosis = killing or causing to die). Other synonyms include avascular necrosis, ischemic necrosis of bone, aseptic necrosis, and subchondral avascular necrosis. The term osteonecrosis dissecans is sometimes used synonymously with osteonecrosis, although, strictly speaking, it is a consequence of osteonecrosis involving dessication of bone leading to fracturing or cracking of bone. The concept of bone death was first described by Hippocrates,1 but the first clinical description of osteonecrosis was a case of sepsis-induced bone death described by Russell in 1794.2 It was almost a century later that bone death was described to occur in the absence of infection.3 The first report of osteonecrosis in a deep sea diver appeared in 1936.4 The pathogenesis of osteonecrosis is complex, but whatever the mechanism, bone death ultimately occurs as a result of complete or partial disruption of the delivery of oxygen and/or nutrients to the bone and surrounding tissues. It is likely that multiple molecular mechanisms may be simultaneously in play in order for osteonecrosis to occur.1,6

**EPIDEMIOLOGY**

The prevalence of osteonecrosis is unknown, but it is estimated that there are 10,000 to 20,000 new patients diagnosed per year in the United States. Osteonecrosis occurs in 15% to 80% of patients with femoral neck fractures.7 Ten percent of the 500,000 hip replacements done in the United States each year are thought to be for osteonecrosis.8 The disease primarily affects men, with a notable exception for osteonecrosis associated with systemic lupus erythematosus, which has a significant female predominance. Osteonecrosis primarily occurs in the third to fifth decade of life.9 As a result of this age distribution, long-term morbidity can be significant because most hip replacements have a finite period of viability.

**ETIOLOGY**

Osteonecrosis has been linked to numerous conditions (Table 103-1). The strength of a causal relationship varies greatly, and in some cases only case reports have been published. The most common cause of nontraumatic osteonecrosis is corticosteroid use, which was first described in 1957.10 Although other adverse effects of corticosteroids are perhaps better known, osteonecrosis of the femoral head is one of the serious complications.

In a 1998 study, in which the investigators reviewed associations in 2500 to 3300 cases of nontraumatic osteonecrosis, corticosteroid use was present in 34.7% of cases. Alcohol use was found in 21.7% of the cases, and the remainder was idiopathic. Although the risk of developing osteonecrosis with corticosteroid use is small, the severity of the adverse event and the high morbidity associated with osteonecrosis make this an important complication to consider when starting a patient on corticosteroids.

Studies have attempted to determine the duration of use and the dosages of corticosteroids necessary to precipitate osteonecrosis. There are several forms of corticosteroids of differing potency and half-life, and dosages and duration of use vary between studies, so any conclusions about a “safe” dose of corticosteroids are wrought with potential confounding variables and errors. In one study of 20 patients diagnosed with stage 1 osteonecrosis by magnetic resonance imaging (MRI), the interval between the use of steroids and diagnosis ranged from 1 to 16 months.11 The cumulative dose of steroids in this study ranged from 1800 to 15,505 mg (mean, 5928 mg) of prednisolone or the equivalent. In other studies cumulative doses of steroids associated with osteonecrosis ranged from 48012 to 432013 mg of dexamethasone dose equivalence. A recent paper by Powell and colleagues14 attempted to collectively analyze the available
Table 103-1  Conditions Associated with Osteonecrosis

<table>
<thead>
<tr>
<th>Dietary, Drugs, and Environmental Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corticosteroids15, 157</td>
</tr>
<tr>
<td>Bisphosphonates158, 159</td>
</tr>
<tr>
<td>Alcoholism25, 160</td>
</tr>
<tr>
<td>Cigarette smoking27</td>
</tr>
<tr>
<td>Dysbaric osteonecrosis4, 161</td>
</tr>
<tr>
<td>Lead poisoning154, 165</td>
</tr>
<tr>
<td>Electric shock156</td>
</tr>
<tr>
<td><strong>Musculoskeletal Conditions: Compromise in Structural Integrity</strong></td>
</tr>
<tr>
<td>Trauma156</td>
</tr>
<tr>
<td>Legg-Calvé-Perthes disease116, 167</td>
</tr>
<tr>
<td>Congenital hip dislocation154, 169</td>
</tr>
<tr>
<td>Slipped femoral capital epiphysis170, 171</td>
</tr>
<tr>
<td><strong>Metabolic Diseases: Abnormality in Fat or Other Metabolic Component</strong></td>
</tr>
<tr>
<td>Fat embolism172, 173</td>
</tr>
<tr>
<td>Pancreatitis174</td>
</tr>
<tr>
<td>Chronic liver disease176</td>
</tr>
<tr>
<td>Pregnancy42, 177</td>
</tr>
<tr>
<td>Fabry’s disease178, 179</td>
</tr>
<tr>
<td>Gaucher’s disease45, 180</td>
</tr>
<tr>
<td>Gout181</td>
</tr>
<tr>
<td>Hyperparathyroidism182</td>
</tr>
<tr>
<td>Hyperlipidemia47, 173</td>
</tr>
<tr>
<td>Hypercholesterolemia81</td>
</tr>
<tr>
<td>Diabetes183</td>
</tr>
<tr>
<td><strong>Hematologic Conditions: Abnormalities in Blood Components</strong></td>
</tr>
<tr>
<td>Sickle cell anemia178, 184, 185</td>
</tr>
<tr>
<td>Hemophilia45, 46</td>
</tr>
<tr>
<td>Hemoglobinopathies27, 176</td>
</tr>
<tr>
<td>Thalassemia180</td>
</tr>
<tr>
<td>Disseminated intravascular coagulation103, 187, 189</td>
</tr>
<tr>
<td>Thrombophilia190</td>
</tr>
<tr>
<td>Hypofibrinolysis190, 191</td>
</tr>
<tr>
<td>Narrow infiltrative disorders</td>
</tr>
<tr>
<td>Thrombophileitis/venous thrombosis192</td>
</tr>
<tr>
<td><strong>Rheumatologic Conditions</strong></td>
</tr>
<tr>
<td>Antiphospholipid antibody syndrome193</td>
</tr>
<tr>
<td>Rheumatoid arthritis194</td>
</tr>
<tr>
<td>Inflammatory bowel disease195, 196</td>
</tr>
<tr>
<td>Necrotizing artery197</td>
</tr>
<tr>
<td>Mucocutaneous lymph node syndrome198</td>
</tr>
<tr>
<td>Polymyositis199</td>
</tr>
<tr>
<td>Sarcoidosis200</td>
</tr>
<tr>
<td>Mixed connective tissue disease</td>
</tr>
<tr>
<td><strong>Infectious Diseases</strong></td>
</tr>
<tr>
<td>Human immunodeficiency virus infection200, 201</td>
</tr>
<tr>
<td>Osteomyelitis202</td>
</tr>
<tr>
<td>Meningococcemia203, 204</td>
</tr>
<tr>
<td>Severe acute respiratory syndrome (SARS)105, 106, 205</td>
</tr>
<tr>
<td><strong>Oncologic Disorders, Transplantation, and Their Treatment</strong></td>
</tr>
<tr>
<td>Organ transplantation (with or without corticosteroid exposure)210</td>
</tr>
<tr>
<td>Radiation exposure212-217</td>
</tr>
<tr>
<td>Regional deep hyperthermia118</td>
</tr>
<tr>
<td>Acute lymphoblastic leukemia219, 220</td>
</tr>
</tbody>
</table>

Additional host-inherent risk factors also play a role in susceptibility. The incidence of osteonecrosis in a group of patients receiving glucocorticoid replacement therapy for primary or secondary adrenal insufficiency was 2.4%. In a study of renal transplantation patients, the 26 patients who developed osteonecrosis had a higher cumulative oral dose of prednisone after 1 and 3 months compared with 28 control transplant patients who did not develop osteonecrosis. A separate study estimated the incidence of osteonecrosis in renal transplant patients to be 5%. There is no evidence to consistently link the use of topical, inhaled, or nasal corticosteroids to osteonecrosis. The evidence for an association between osteonecrosis and intramuscular or intra-articular corticosteroids is limited to case reports. Parenteral use poses a higher risk because of rapid absorption and longer half-life of the drugs used.

Bisphosphonate-induced osteonecrosis of the jaw is particularly interesting because of the intended use of bisphosphonates on bone diseases. There has been a link between cigarette smoking and osteonecrosis, with smokers having a threefold higher relative risk for developing osteonecrosis, independent of all other factors.

The association between osteonecrosis and alcohol consumption was first described in 1922. A study of patients with idiopathic osteonecrosis revealed that the risk of osteonecrosis increased with increasing daily consumption of alcohol. The subjects were divided into three groups on the basis of their alcohol consumption of less than 400 mL/week, 400 to 1000 mL/week, and greater than 1000 mL/week, and the relative risk of osteonecrosis, independent of corticosteroid use or smoking, was 3-fold, 10-fold, and 18-fold, respectively, when compared with hospital controls. Liver damage was also found unnecessary for the development of osteonecrosis in alcohol-consuming patients, although elevated liver enzymes may be present. The incidence of osteonecrosis in patients who received treatment for alcoholism was 5.3%. The femoral head was again the most common site (82 of 92 lesions), with the other 10 sites involving the humeral head.

Musculoskeletal conditions can lead to osteonecrosis in children. Legg-Calvé-Perthes disease was first described in children between 3 and 12 years of age in 1910. Femoral head osteonecrosis is a feature of this disease and has been linked to trauma, congenital hip dislocation, and transient synovitis. Bilateral involvement is common, and associated clinical manifestations include abnormal growth and stature, delayed skeletal maturation, disproportionate skeletal growth, congenital anomalies, and abnormal hormone levels. Children with acute lymphoblastic leukemia can develop osteonecrosis as well, but this may be a result of steroid use. An additional risk factor for this cohort of patients is high body mass index.

Osteonecrosis has also been associated with metabolic disorders and in pregnancy. Diagnosis is often delayed until months after delivery. Women who develop osteonecrosis in pregnancy tended to have a small body frame and a large weight gain.

Hematologic conditions have been associated with osteonecrosis. The long-term morbidity of osteonecrosis in patients with sickle cell anemia is dismal.
deformities include decreased mobility, abnormal gait, and leg-length discrepancy.44 Osteonecrosis in hemophilia patients has been reported, but no statistically reliable causal link can be established.45–50

Dysbaric osteonecrosis was first described in construction workers in the Elbe tunnel exposed to high-pressure environments.51 The prevalence of dysbaric osteonecrosis is 4.2% in divers and 17% in compressed air workers.52 Patients with dysbaric osteonecrosis may have more than one lesion, and common sites besides the femoral head include the tibia and the humeral head and shaft. The condition is not related to decompression sickness, and although proper decompression procedures can reduce “the bends,” they do not have any effect on the development of osteonecrosis, which can occur months or years after the last exposure to high-pressure environments.

Osteonecrosis has also been associated with a number of infectious diseases including severe acute respiratory syndrome (SARS). Many patients who contracted SARS in the early 2000s received treatment with corticosteroids, and some subsequently developed osteonecrosis.53 The incidence of osteonecrosis appears higher in this group of patients compared with patients with other conditions who were treated with corticosteroids.54 Chan and colleagues55 reported five children with SARS treated with corticosteroids who developed osteonecrosis.

CLINICAL FEATURES

The primary presenting symptom in osteonecrosis is pain. In osteonecrosis of the femoral hip, the pain is located in the hip joint but may radiate to the groin, anterior thigh, or knee. The severity of the pain can vary, depending on the size of the infarct and whether the onset of disease is insidious or sudden. In trauma, where there is sudden and severe disruption of blood flow, and in Gaucher’s disease, dysbarism, or hemoglobinopathy, where the infarcts are large, pain can be intense and sudden. In other conditions where the onset is more insidious, the pain can follow a gradual and slow incremental progression. The pain of osteonecrosis is usually increased with use of the joint, but in advanced disease the pain can be persistent at rest. Limitation of range of motion is progressive and is usually a late symptom, except when resulting from accompanying pain. The risk of developing osteonecrosis of the contralateral hip when one side is affected ranges from 31% to 55%.

In addition to the femoral head, osteonecrosis can affect other sites including the humeral head,56–59 femoral condyles,60–62 and proximal tibiae,63,64,66 wrists and ankles,67 bones of the hands and feet,68 the vertebrae,69,71 jaw,72,73 and bony structures of the face.56 Osteonecrosis of the humeral head is the second most commonly seen location, and pain is usually in the shoulder and associated with reduced range of motion and weakness. Pain in the ankle is the main presenting symptom in nontraumatic osteonecrosis of the talus, and in some cases, the disease had already progressed to Ficat and Arlet stage 3 by the time of presentation of pain.67 Kienböck’s disease involves osteonecrosis of the lunate. Patients present with pain in the radiolunate joint, along with weakness and limitation of motion. Kienböck’s disease appears to be related to manual labor. Soccer players have been reported to develop osteonecrosis of the foot,77 and football players may be prone to developing osteonecrosis of the hip.78

The Ficat and Arlet method of staging osteonecrosis consists of four stages. Stages 1 and 2 are reversible, whereas stage 3 (subchondral collapse) and stage 4 (joint space narrowing and destruction of cartilage) are irreversible. The Marcus staging system consists of six stages, in which the first two are reversible and the subsequent four are irreversible. The modified Steinberg staging system is based on the Marcus system and also consists of six stages. Each stage is further divided into three subclasses on the basis of the extent of femoral head involvement. Subclass A involves less than 25%; B involves 26% to 50%, and C involves greater than 50%.

Table 103-2 shows the Modified Steinberg system for staging osteonecrosis. The Association of Research Circulation Osseous (ARCO) has proposed a modification to the Ficat and Arlet system, adding a stage 0 or patients with negative imaging studies but who are at risk for developing osteonecrosis. In addition, stages 1 and 3 are further stratified to take into account lesion size, location, and extent of collapse.79 In 2001 the Japanese Ministry of Health, Labor and Welfare proposed revising criteria for the diagnosis and staging of osteonecrosis of the femoral head.80 Diagnostic criteria included the following: (1) collapse of the femoral head without joint space narrowing or acetabular abnormality on plain radiograph, (2) demarcating sclerosis in the femoral head without joint space narrowing or acetabular abnormality, (3) “cold in hot” on bone scans, (4) low-intensity band on T1-weighted MRI, and (5) trabecular and narrow necrosis on histology. If a patient fulfills two of the five criteria, the diagnosis is established. The working group also proposed four types of lesions on the basis of extensiveness and defined stages of disease on the basis of diagnostic imaging.

Bone Marrow Edema

Bone marrow edema is a common observation in osteonecrosis and is frequently accompanied by vascular congestion. Bone marrow edema is not specific for osteonecrosis and may be seen in many musculoskeletal disorders including osteomyelitis, osteoarthritis, occult intraosseous fracture, stress fracture, osteoporosis, and sickle cell crisis.

A specific syndrome known as bone marrow edema syndrome has been described and was initially thought to be a

---

**Table 103-2** Modified Steinberg Staging Systems for Osteonecrosis

<table>
<thead>
<tr>
<th>Stage</th>
<th>Radiographic Appearance</th>
<th>Reversible</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Normal radiographs, but abnormal bone scan or magnetic resonance image</td>
<td>Yes</td>
</tr>
<tr>
<td>II</td>
<td>Lucent and sclerotic changes</td>
<td>Yes</td>
</tr>
<tr>
<td>III</td>
<td>Subchondral fracture without flattening</td>
<td>No</td>
</tr>
<tr>
<td>IV</td>
<td>Subchondral fracture with flattening or segmental depression of femoral head</td>
<td>No</td>
</tr>
<tr>
<td>V</td>
<td>Joint space narrowing or acetabular changes</td>
<td>No</td>
</tr>
<tr>
<td>VI</td>
<td>Advanced degenerative changes</td>
<td>No</td>
</tr>
</tbody>
</table>
precursor to osteonecrosis, but it is now believed to be a separate entity. Bone marrow edema is a transitory, self-limiting condition typically seen in middle-aged men and in women in their third trimester of pregnancy. Patients complain of pain, limited range of motion, and an abnormal gait. Osteopenia is detected on conventional radiographs, and MRI confirms this with a low signal on T1-weighted images and a high signal on T2-weighted images. The three phases of bone marrow edema syndrome include an initial phase lasting about 1 month, followed by a plateau phase lasting 1 to 2 months, and finally a regression phase lasting for an additional 4 to 6 months. Subchondral fractures do not occur. Biopsy specimens obtained in the initial phase show diffuse interstitial edema, fragmentation of fatty marrow cells, and increased new bone formation.

A study of 24 cases of bone marrow edema syndrome of the knee showed that although migrating bone marrow edema occurred in a third of patients at a 5-year follow-up, the patients were asymptomatic and MRI signal alterations had resolved. Biopsy specimens of affected bone were obtained using arthroscopic surgery and core decompression, and histology revealed areas of bone marrow edema and vital trabeculae covered by osteblasts and osteoid seams. None of the cases progressed to osteonecrosis.

Bisphosphonates and Osteonecrosis of the Jaw

Bisphosphonate is a class of drug used to treat osteoporosis and diseases where bone is not formed adequately. Bisphosphonates are composed of two forms, and osteonecrosis appears to occur in association with nitrogen-containing bisphosphonates. The mechanism of action of bisphosphonate-induced osteonecrosis of the jaw appears to parallel that of glucocorticoids, with derangement in lipid metabolism, bone homeostasis, and apoptosis of bone cells. It is interesting that the jawbone seems to be the most vulnerable bone in bisphosphonate-induced disease, as opposed to the femoral head in most other associations or causes of osteonecrosis. This may be because of the high bone turnover rate in the jaw or because bisphosphonates exert their action on not only bone but also many elements of the surrounding tissue including fibroblasts and blood vessels.

PATHOGENESIS

Anatomic Considerations in Trauma-Related Osteonecrosis

The femoral head is the most common site of osteonecrosis. An understanding of the anatomy of the femoral head may help to explain why that is the case. Three arterial networks supply the femoral head and neck. The extracapsular arterial ring consists of the lateral femoral circumflex artery and the medial femoral circumflex artery, which arise from the profunda femoris. The medial femoral circumflex artery and its branches supply most of the blood to the head and neck of the femur. The lateral femoral artery winds anterolaterally, and the medial femoral artery winds posteromedially around the neck of the femur, ultimately anastomosing with each other at the superolateral aspect of the femoral head. The lateral femoral circumflex artery and the medial femoral circumflex artery further anastomose with the superior and inferior gluteal branches of the internal iliac artery, providing collateral circulation between the femoral artery and the internal iliac artery. Small vessels known as retinacular arteries, ascending cervical branches of the extracapsular ring, form an intra-articular ring at the level of the cartilage. Eiphyseal arterial branches arise from this ring and penetrate the head and neck of the femur including the epiphyses. The artery of the ligament of the head of the femur is a branch of the obturator artery and may be the sole supplier of blood to the proximal fragment of the head.

Some of these anatomic features may render the femoral head particularly vulnerable to ischemia. The retinacular arteries are believed to supply 80% of the femoral epiphysis. Compromising this critical vascular system may lead to osteonecrosis originating in the anterosuperior aspect of the femoral head, as indicated by angiographic studies in early osteonecrosis in which these arteries are not visualized. A schematic of the blood supply to the femoral head is shown in Figure 103-1.

Histologically, after an infarct, a rim of bony thickening or sclerosis begins to form around the margins of the infarcted area. If the necrotic lesion is within the weight-bearing region of the femoral head, subchondral fractures follow. With repeated microfractures and continued weight bearing, the original fracture cannot heal completely and new fractures appear. The secondary fracture propagates along the junction between subchondral bone and the necrotic segment. As time goes on, the femoral head becomes flattened and eventually collapses. A nonspherical head articulating with the acetabulum produces friction and erosion and loss of cartilage. The cycle repeats itself, and the structure of the joint deteriorates, leading to degenerative changes and eventual total joint destruction.

Nontraumatic Osteonecrosis

Disruption of the blood supply to the femoral head can occur through a number of different mechanisms. In
traumatic osteonecrosis of the femoral head, the cause of this disruption is often viewed as completely mechanical and appears to be easily understood. But there may be an additional component to the disruption that is related to the immunologic and inflammatory changes that occur in damaged bone tissue and surrounding soft tissues.

The immunologic changes occurring in nontraumatic osteonecrosis may help explain why corticosteroids are particularly dangerous to the integrity of the blood supply of the femoral hip. Some have likened osteonecrosis to “coronary disease” of the hip and propose that the same mechanisms that cause ischemia of the myocardium may also cause ischemia of the femoral head (Table 103-3).

### Mechanical and Vascular Considerations

In Legg-Calvé-Perthes disease, obstruction to venous drainage elevates intraosseous pressure and consequently elevates intra-articular pressures. In a study of patients with Legg-Calvé-Perthes disease, bone scintigraphy using Tc99m methylene diphosphonate (Tc99m MDP) was employed to measure arterial and venous flow in the diseased hip. Although arterial flow was normal, there was significant disruption in venous drainage. This disturbance was reproduced in a dog model in which injection of silicone was used to obstruct venous flow distal to the hip. Ischemia resulted from the obstruction to venous drainage, leading to a cessation of endochondral ossification in the preossous epiphyseal cartilage and the physis plate. Widening of the joint space ensued, followed by revascularization of the epiphysis and deposition of new immature bone. A weakened or unstable femoral epiphyseal plate resulted, and the subchondral bone became prone to segmental collapse and fracture.

The pathologic mechanism of dysbaric osteonecrosis is unclear. The most intuitive explanation is that formation of gas bubbles causes arterial occlusion and ischemia. However, the true mechanism may not be quite so simple. Multiple other factors might contribute to the disease including thromboembolic events such as platelet aggregation, erythrocyte clumping, lipid coalescence, intraosseous vessel compression as a result of extravascular gas bubbles, formation of fibrin thrombi, and narrowing of arterial lumina owing to myointimal thickening caused by gas bubbles. The interaction between gas and blood can lead to the formation of vessel-occluding substances. All of these events can lead to redistribution of blood flow.

The increased vulnerability of bone to compression disorders has been explained by several factors including the relative rigidity of bone and inability to absorb increased gas pressure, inherent poor vascularization, and gas supersaturation of fatty marrow. A sheep model of dysbaric osteonecrosis has been developed. Exposure to compressed air at pressures of 2.6 to 2.9 atmospheres for 24 hours results in extensive bone and marrow necrosis. The authors proposed that the initial event involving elevated intramedullary pressures leads to the formation of nitrogen gas bubbles in the fatty marrow of the long bones. Radiography shows medullary opacities and endosteal thickening. Later, neovascularization of previously ischemic fatty marrow occurs, followed by new bone formation. Osteonecrosis occurs in subchondral cortical bone with marrow fibrosis and osteocyte loss.

Changes in the vasculature, through injury or inflammation from other diseases, may in turn lead to a compromise in blood flow. Examples include structural damage to arteriolar walls, degeneration of the tunica media, smooth muscle cell necrosis, and disruption of the internal elastic lamina. These changes can lead to eventual hemorrhagic infarction, which was observed in a study of 24 core biopsy specimens from osteonecrotic femoral heads. The changes did not occur in 11 femoral heads with osteoarthrosis.

### Osteoimmunology

Although bone marrow is a critical component of the immune system, bone matrix is often perceived to be static scaffolding that functions primarily to support the musculoskeletal system. It is now known that, in fact, bone matrix is a dynamic tissue that is constantly replacing itself. It is estimated that about 10% of a person’s bone is replaced every year. Diseases such as osteopetrosis and osteoporosis are a result of a dysfunction in the balance between bone deposition and bone resorption. The factors that regulate this homeostasis include cells of the bone matrix, immune cells, signaling molecules, cytokines and chemokines, and vitamins. Some of these regulatory factors may be present on both bone cells and immune cells, often serving different functions, thereby providing a link between the immune system and bone. Osteonecrosis, in fact, may be linked to such an imbalance in bone homeostasis. Immune factors may affect surrounding soft tissue as well, contributing to the development of osteonecrosis. The study of immune regulation of bone in osteonecrosis may

---

**Table 103-3  Proposed Mechanism of Disease of Common Conditions Associated with Osteonecrosis**

<table>
<thead>
<tr>
<th>Associated Condition</th>
<th>Apoptosis</th>
<th>Osteoblast/Osteoclast Abnormalities</th>
<th>Lipid Abnormalities</th>
<th>Coagulation Abnormalities</th>
<th>Oxidative Stress</th>
<th>Parathyroid/Calcium Imbalance</th>
<th>Vascular Plugging</th>
<th>Vasoactive Substances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corticosteroids</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bisphosphonates</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol abuse</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trauma</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renal transplantation</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dialysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Sickle cell disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

---

---

---
encompasses many of the previously proposed mechanisms of osteonecrosis including apoptosis, oxidative stress, and genetic predisposition.

Immune factors involved in bone homeostasis include receptor activator of NFκB (RANK) and its ligand (RANKL), IL-1, IL-6, IL-10, TFG-β, TNF, CD80, CD86, CD40, macrophage colony-stimulating factor (M-CSF), NFATc, and vitamin D. (See Table 103-4 for roles and function.) Many of these factors can be categorized into one of two categories, those with the overall effect of inducing osteoclastogenesis and those that inhibit osteoclastogenesis. In addition, factors involved in cell survival and apoptosis such as Blimp-1 and Bcl6 may also play a role. RANKL is expressed on osteoblasts and is critical for the differentiation and proliferation of osteoclasts. Because transcription of factors involved in the regulation of bone homeostasis is often influenced by glucocorticoids, this may begin to explain why steroids may be associated with osteonecrosis.

The action of glucocorticoids is mediated by the glucocorticoid receptor, which is present on many cell types including osteoclasts, osteoblasts, osteocytes, and cartilage. Binding of glucocorticoids to its receptor leads to the anti-inflammatory activity known to be a function of steroids. One mechanism by which this anti-inflammatory effect is mediated is by transcription of genes that inhibit the synthesis of inflammatory mediators.

Osteoblast/Osteoclast Balance

Any disturbance in the normal homeostasis between bone deposition and bone resorption can lead to bone disease. Moreover, defective bone deposition or bone resorption in which new bone is formed in an aberrant manner can lead to disease. Alcohol can affect the ability of mesenchymal stem cells to differentiate into osteogenic lineages. The bone marrow in the proximal head of femurs was isolated during hip replacement surgery from 33 patients with either femoral neck fractures or alcohol-induced osteonecrosis. The cells from femurs of patients with alcohol-induced osteonecrosis showed a reduced ability to differentiate into

<table>
<thead>
<tr>
<th>Immune Factor</th>
<th>Ligand</th>
<th>Cellular Source</th>
<th>Function in Bone Homeostasis</th>
<th>OC</th>
<th>Immune Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANK</td>
<td>RANKL</td>
<td>Osteocasts, dendritic cells</td>
<td>Upon binding to RANKL, signals differentiation into osteoclast</td>
<td>↑</td>
<td>RANKL-RANK binding leads to dendritic cell activation</td>
</tr>
<tr>
<td>RANKL</td>
<td>RANK</td>
<td>Osteoblasts, T helper cells</td>
<td>Activation of osteoclasts. Overproduction can result in RA or PA</td>
<td>↑</td>
<td>Dendritic cell maturation</td>
</tr>
<tr>
<td>OPG</td>
<td>RANKL</td>
<td>Osteoblasts, macrophages, bone fibroblasts, stromal cells</td>
<td>Decoy receptor for RANKL Stimulates osteoclastogenesis</td>
<td>↓</td>
<td>Influences hematopoietic stem cells to differentiate into macrophages</td>
</tr>
<tr>
<td>M-CSF</td>
<td>CSF-1 receptor</td>
<td>Osteoblasts, fibroblasts, stromal cells</td>
<td>Stimulates osteoclastogenesis</td>
<td>↑</td>
<td>Influences multiple signaling pathways, including NFκB, death signaling and MAP kinase pathway</td>
</tr>
<tr>
<td>TNF</td>
<td>TNF receptor</td>
<td>Macrophages, lymphocytes, mast cells, and many others</td>
<td>Stimulates osteoclastogenesis</td>
<td>↑</td>
<td>Proinflammatory cytokine, endogenous pyrogen</td>
</tr>
<tr>
<td>TGF-β</td>
<td>TGF-β receptor</td>
<td>Multiple cell lines</td>
<td>Induction of apoptosis</td>
<td>↑</td>
<td>Regulatory role, blocks activation of lymphocyte- and monocyte-derived phagocytosis</td>
</tr>
<tr>
<td>Blimp-1</td>
<td>Bcl6 promoter</td>
<td>Plasmablasts, plasma cells</td>
<td>Binds to Bcl6 promoter, suppression expression Inhibits osteoclastogenesis</td>
<td>↑</td>
<td>Inhibits Tfh cell differentiation in mice</td>
</tr>
<tr>
<td>Bcl-6</td>
<td>?</td>
<td>Germinal center B cells</td>
<td>Stimulation of Tfh cell differentiation in mice</td>
<td>↓</td>
<td>Stimulation of Tfh cell differentiation in mice</td>
</tr>
<tr>
<td>IL-1</td>
<td>IL-1R</td>
<td>Macrophages, monocytes, fibroblasts, dendritic cells</td>
<td>Directly activates RANK signaling to promote osteoclastogenesis</td>
<td>↑</td>
<td>Proinflammatory cytokine</td>
</tr>
<tr>
<td>IL-6</td>
<td>IL-6R</td>
<td>Osteoblasts</td>
<td>Activation of osteoclastogenesis</td>
<td>↑</td>
<td>Proinflammatory cytokine</td>
</tr>
<tr>
<td>IL-10</td>
<td>IL-10Rα</td>
<td>Monocytes, lymphocytes</td>
<td>Supress bone resorption</td>
<td>↓</td>
<td>Anti-inflammatory cytokine, blocks NFκB activity, regulatory cytokine</td>
</tr>
<tr>
<td>Vitamin D</td>
<td>VDR</td>
<td>Osteoblast, monocyte/macrophage</td>
<td>Facilitate adhesion of osteoclast precursor to osteoblast</td>
<td>↑</td>
<td>Cell proliferation and differentiation</td>
</tr>
<tr>
<td>Estrogens</td>
<td>Estrogen receptor</td>
<td>Ovarian follicle cells</td>
<td>Reduces osteoclast IL-1 responsiveness and cell survival; stimulates osteoprotegrin</td>
<td>↓</td>
<td>Angiogenesis, endothelial healing</td>
</tr>
<tr>
<td>IL-17</td>
<td>IL-17R</td>
<td>T cells</td>
<td>May have opposing roles of bone protection and bone loss</td>
<td>↑↓</td>
<td>Proinflammatory cytokine</td>
</tr>
<tr>
<td>IL-18</td>
<td>IL-18R</td>
<td>Macrophages</td>
<td>Inhibits TNF-mediated osteoclastogenesis in a T cell–independent manner</td>
<td>↓</td>
<td>Proinflammatory cytokine, works in synergy with IL-12</td>
</tr>
</tbody>
</table>

Examples of some of the factors involved in bone metabolism. In addition to the factors listed, there are many others that play a role, either by themselves or in conjunction with other factors. The factors listed may have many other functions. Only select functions are listed.

Bcl6, B cell lymphoma 6 protein; Blimp, B lymphocyte–induced maturation protein 1; CSF-1, colony-stimulating factor 1; OC, osteoclastogenic; OPG, osteoprotegrin; PA, psoriatic arthritis; RA, rheumatoid arthritis; RANK, receptor activator for NFκB; RANKL, receptor activator for NFκB ligand; Tfh, T follicular helper cell; TGF-β, transforming growth factor beta; TNF, tumor necrosis factor; VDR, vitamin D receptor.
osteoblasts. A subsequent study compared the mesenchymal stem cells from patients with hip osteoarthritis, idiopathic osteonecrosis, and nontraumatic osteonecrosis associated with steroid or alcohol use. In idiopathic and alcohol-induced osteonecrosis, the ability of mesenchymal stem cells to differentiate into osteoblasts was decreased, but in steroid-induced osteonecrosis, it was elevated, although not to a statistically significant level. The apoptotic differentiation ability was similar in all four groups.

In rats fed a diet of alcohol and glucose, lower bone mineral content and density were detected compared with controls. In hamsters, alcohol led to thinning of the trabeculae of the distal part of the femur. Cytologic effects included mitochondrial swelling in osteoblasts and osteocytes. Partial osteonecrosis of the femoral head was detected in Merino sheep that were injected with ethanol. In humans, alcohol causes increased plasma calcium levels, decreased osteocalcin and circulating parathyroid hormone levels, reduced serum calcitriol, reduced bone volume, and increased osteoclast number.

Alterations in osteoblast function may also contribute to the pathogenesis of osteonecrosis. In one study, osteoblastic cells were obtained from bone biopsy specimens from the intertrochanteric region of the femur and of the iliac crest of 13 patients with osteonecrosis and 8 patients with hip osteoarthritis. Cell replication was measured on the basis of proliferation rate in secondary culture. Levels of alkaline phosphatase activity, collagen synthesis, and the sensitivity to 1,25-dihydroxyvitamin D were measured. The results indicated that although differentiation was not affected, the proliferation rate of osteoblastic cells was reduced in samples obtained from the patients with osteonecrosis compared with patients with osteoarthritic hips.

Apoptosis and Osteonecrosis

Glucocorticoids can also act via its action on apoptosis of immune and bone cells. When mice were administered prednisolone for 27 days, increased metaphysial apoptotic activity of both osteoblasts and osteoclasts were noted. The result was decreased bone turnover, density, and formation; increased formation of cancellous bone; and decreased trabecular width. The decreased bone turnover can be explained by the reduced osteoclast survival, and the reduction in trabecular width can be explained by a decrease in osteoblasts. An accumulation of apoptotic elements was also found in the region of the “fracture crescent” in the femurs of glucocorticoid-treated patients. On the other hand, glucocorticoids may also increase osteoclast survival, leading to increased bone loss. Clearly, the effect of osteoclast survival on bone disease is more complicated than at first glance, and it involves the interaction of the osteoclast with the osteoblast. Because osteoblasts are also responsible for osteoclast differentiation under the right circumstances, there exists a significant feedback system that maintains bone homeostasis.

Osteocyte death is also a feature of osteonecrosis. In a rat model, ischemia caused an induction in the expression of stress proteins, oxygen-regulated protein (ORP150) and hemeoxygenase 1 (HO1). Induction of ischemia in these rates caused DNA fragmentation and the presence of apoptotic bodies in chondrocytes, bone marrow cells, and osteocytes. Both alcohol and corticosteroids can induce osteocyte apoptosis, possibly via lipid abnormalities.

Lipids and Osteonecrosis

The bone marrow of rabbits that were fed alcohol showed fatty infiltration of the liver and adipogenesis in the bone marrow. Increases in fat cell hypertrophy and proliferation, as well as a decrease in hematopoiesis in the subchondral head, were observed. Osteocytes contained triglyceride deposits, and there was an increase in empty osteocyte lacunae. Alcohol also primarily triggered differentiation of bone marrow stromal cells into adipocytes in a dose-dependent manner. Intracellular lipid deposits led to the death of osteocytes.

In corticosteroid-induced osteonecrosis, the alteration in lipid metabolism parallels that of alcohol-induced osteonecrosis. In both cases, fatty infiltration of osteocytes has been postulated to occur. Table 103-5 lists lipid-altering effects of corticosteroids and alcohol. In addition, interosseous venous stasis affects the interosseous microcirculation, which can lead to hemodynamic and structural changes in the femoral head. The resulting decrease in blood flow leads to osteonecrosis. In chickens treated with steroids, fatty infiltration of the liver and fat cell hypertrophy and proliferation in the femoral head occurred concurrently 1 week after the initiation of steroids. As in the case of alcohol-induced osteonecrosis, adipocytes contained triglyceride vesicles. In rabbits treated with steroids, it was found that interosseous pressure was increased and the size of bone marrow fat cells was larger than in control rabbits. A histologic study of acetabular and proximal femoral bone in osteonecrosis of the femoral head revealed that osteonecrosis is more extensive in corticosteroid-induced compared with alcohol-induced or idiopathic osteonecrosis. The reason for this is unknown.

In osteonecrosis of the jaw, bisphosphonates inhibit protein prenylation via inhibition of the enzyme farnesyl diphosphate synthase. The normal lipid metabolism of pathways that regulate cytoskeletal integrity and osteoclastogenesis such as Rho, Rac, and Ras is disrupted. This is one of the mechanisms by which bisphosphonates exert their intended action, but their ability to disrupt normal regulation of bone metabolism may instead lead to osteonecrosis.

Coagulation and Osteonecrosis

The hyperlipidemia, increased serum free fatty acids, and increased prostaglandins that are associated with alcohol-induced osteonecrosis may potentially trigger vascular

<table>
<thead>
<tr>
<th>Fat emboli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty infiltration of bone marrow</td>
</tr>
<tr>
<td>Adipogenesis of marrow stromal cells</td>
</tr>
</tbody>
</table>
inflammation and coagulation. Other triggers for intravascular coagulation include atherosclerosis and arteriolar fibroid degeneration. Jones proposed that the progression of osteonecrosis from stage 1A to 1B is linked to an inability to clear procoagulants from blood or tissue. He proposed that decreased clearance of procoagulants leads to persistent levels of tissue thromboplastin, leading to arteriolar thrombosis, vascular stasis, free fatty acid–induced endothelial damage, and hypercoagulability. Studies have shown that patients with osteonecrosis had a much higher frequency of having at least one and at least two abnormal coagulant levels compared with normal controls. Of patients with osteonecrosis, 82% had at least one abnormal procoagulant level, and 47% had at least two. In normal controls, only 30% had one abnormal procoagulant level and only 2.5% had two or more. The procoagulants measured included free protein S, protein C, lipoprotein A, homocysteine, plasminogen activator inhibitor, stimulated tissue plasminogen activator, antcardiolipin antibodies (IgM and IgG), and resistance to activated protein C.

In addition, both thrombophilia and hypofibrinolysis have been associated with osteonecrosis. Hypofibrinolysis leads to an increased likelihood of clot formation, and thrombophilia results in a decreased ability to lyse clots. This is yet another mechanism by which corticosteroids lead to osteonecrosis—high-dose steroids lead to increased plasma plasminogen activator inhibitor, decreased tissue plasminogen activator activity, and inhibition of the fibrinolytic pathway, thus leading to a higher risk for clot formation. There is an early indication that coagulation abnormalities may play a significant role in corticosteroid-induced osteonecrosis in SARS patients.

### Oxidative Stress and Osteonecrosis

Alcohol consumption is associated with reduced superoxide dismutase activity. Alcohol has deleterious effects on muscle including increased oxygen free radical–related damage, reduced myocardial contractility, defective mitochondrial function, and increased tissue enzymes. When rabbits were injected with methylprednisolone, elevations in 8-hydroxy-2′-deoxyguanosine, a marker of DNA oxidative injury, were observed. This coincided with the development of osteonecrosis. A polymorphism in the nitric oxide synthase, described later, was also associated with the development of osteonecrosis. This relationship between osteonecrosis and oxidative injury leads one to wonder if corticosteroid-induced osteonecrosis can be prevented or lessened in severity by simultaneous or prophylactic administration of antioxidants.

### Nitric Oxide Synthase and Osteonecrosis

Glucocorticoids can cause derangements in vascular responsiveness to vasoactive substances such as nitric oxide. Endothelial nitric oxide synthase (eNOS) stimulates the production of nitric oxide. Nitric oxide regulates vascular “tension” by acting as a vasodilator, inhibiting mononuclear adhesion to endothelial cells and preventing platelet aggregation. A defect in this activity can lead to increased vascular resistance and disruption to downstream blood flow, resulting in osteonecrosis.

### Multihit Hypothesis

Other proposed mechanisms involve endothelial cell injury, abnormal angiogenesis and repair mechanisms, the effects of vasoactive substances, activity of hepatic cytochrome P450 3A4, and intramedullary hemorrhage. Multiple mechanisms may be simultaneously occurring. Kenzora was the first to introduce the concept of cumulative stress. Corticosteroid-induced osteonecrosis seems to occur with greater frequency in patients who have significant underlying illness such as systemic lupus erythematosus or transplantation and less frequently or never in patients who are not chronically ill but are on steroids for an acute event such as head injury. Recent observations that corticosteroids induce osteonecrosis in SARS patients further support the notion that more than one insult to the bone or surrounding tissue may be necessary to precipitate osteonecrosis. For each of the known associations of osteonecrosis, different mechanisms may predominate such as lipid anomalies and apoptosis of osteoblasts in steroid-induced osteonecrosis, as well as elevated intraosseous pressures and coagulation abnormalities in dysbaric osteonecrosis, but additional factors may be necessary to precipitate osteonecrosis. The accumulated cell stress theory suggests that when the damaging effects of multiple events are added together, the involved bone is unable to recover from the chronic stress and osteonecrosis ensues.

### Genetic Considerations

The degree to which genetics and the environment play in the pathogenesis of osteonecrosis is the subject of an ongoing investigation. Certainly, single nucleotide polymorphisms have been noted in a number of genes that may be associated with osteonecrosis. It has been argued that endothelial nitric oxide synthase is an important player in the development of osteonecrosis. Nitric oxide may have beneficial effects on three systems involved in osteonecrosis, namely skeletal, vascular, and thrombotic. Each of these may be targets for proposed mechanisms of pathogenesis of osteonecrosis. A comparative analysis of the 26-base pair repeat polymorphism in intron 4 and the Glu298Asp polymorphism in exon 7 of the eNOS gene in patients with idiopathic, steroid-induced, alcohol-induced, and normal control subjects was performed. The frequency of the homozygous 4a allele was found to be higher in patients with idiopathic osteonecrosis compared with control subjects. The frequency of the 4a/b allele was found to be higher in all types of osteonecrosis when compared with control subjects. The 4a allele is known to be associated with reduced synthesis of endothelial nitric oxide synthase, suggesting that nitric oxide may play a protective role against the development of osteonecrosis.

Forty-one percent of patients with osteonecrosis compared with only 20% of controls were homozygous for the 4G/4G mutation in the plasminogen activator inhibitor-1 gene. This mutation causes increased hypofibrinolytic plasminogen activator inhibitor activity, resulting in decreased stimulated plasminogen activator activity. This observation lends support to the theory that procoagulants may play a significant role in the pathogenesis of osteonecrosis. A polymorphism in the plasminogen activator...
inhibitor-1 (PAI-1) gene has also been reported to be predictive of osteonecrosis in children with acute lymphoblastic leukemia.\textsuperscript{121}

Genetic variations in type and levels of lipoprotein (a) have been linked to osteonecrosis. Apo(a) is involved in lipid metabolism and the coagulation systems, and the Apo(a) low-molecular-weight phenotype is associated with an increased risk of osteonecrosis.\textsuperscript{122-124} Polymorphisms in the promoter for vascular endothelial growth factor (VEGF) and in the receptor for IL-23 were associated with osteonecrosis in the Korean population,\textsuperscript{125,126} reflecting the significance of the association of osteonecrosis with vascular disorders and autoimmune diseases, respectively.

**DIAGNOSIS**

**History and Physical Examination**

The diagnosis of osteonecrosis is generally made by history because many patients may not present until they develop hip pain. By the time the patient is clinically symptomatic, the disease may be quite advanced. Therefore a high index of suspicion is necessary for all patients on oral or parenteral steroids. Information that should be elicited from a good history should include any history of trauma; underlying disease; alcohol use; tobacco use; current medications; past medications; history of joint anomalies; presence of pain or limitation of motion; involvement in sports, especially high-impact sports; occupational history; gestational history; and the presence of liver disease or lipid abnormalities.

A good physical examination includes palpating the hip for tenderness, identification of limp, masses, leg-length discrepancy, the presence of masses, abnormal gait, muscle strength, and range of motion.

The Harris hip score is frequently used for evaluation of hip function and is also useful in monitoring the effectiveness of treatment (Figure 103-2).\textsuperscript{127-129} The Harris hip score is a multidimensional observational assessment based on eight items that address pain, walking function, daily activity, and range of motion. Scores range from 0 (maximum disability) to 100 (no disability).

**Radiologic Imaging**

When the diagnosis is suspected clinically, it can be confirmed by radiologic imaging studies. Earlier employed imaging techniques such as conventional radiography were inadequate in establishing the diagnosis because in the early stages of osteonecrosis radiographs may be completely normal. The earliest radiographic sign of osteonecrosis is the presence of a radiolucent crescent-shaped rim along the contour of the femoral head (crescent sign) (Figure 103-3). This appearance on radiographs is the result of structural collapse of a necrotic segment of subchondral trabecular bone. At this stage, the disease is already irreversible. Later, radiographs will begin to show sclerotic changes.

---

**Hip joint evaluation system**

<table>
<thead>
<tr>
<th>Date of assessment:</th>
<th>Name:</th>
<th>Medical record #:</th>
<th>DOB:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain</strong></td>
<td><strong>Distance walked</strong></td>
<td><strong>Activities — shoes, socks</strong></td>
<td><strong>Public transportation</strong></td>
</tr>
<tr>
<td>□ Totally disabled, crippled, pain in bed, bedridden</td>
<td></td>
<td>□ Bed and chair only</td>
<td>□ Unable to use</td>
</tr>
<tr>
<td>□ Marked pain, serious limitation of activities</td>
<td></td>
<td>□ Two or three blocks</td>
<td>□ Unable to fit or tie</td>
</tr>
<tr>
<td>□ Moderate pain, tolerable but makes concessions to pain, Some limitation of ordinary activity or work, May require occasional pain medication stronger than aspirin</td>
<td></td>
<td>□ Six blocks</td>
<td>□ With difficulty</td>
</tr>
<tr>
<td>□ Mild pain, no effect on average activities, rarely moderate pain with unusual activity, may take aspirin</td>
<td></td>
<td>□ Unlimited</td>
<td>□ With ease</td>
</tr>
<tr>
<td>□ Slight pain, occasional, no compromise in activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ None, or ignores it</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Support</strong></td>
<td><strong>Stairs</strong></td>
<td><strong>Sitting</strong></td>
<td><strong>Limb-length discrepancy</strong></td>
</tr>
<tr>
<td>□ Two crutches or not able to walk</td>
<td>□ Unable to do stairs in any manner</td>
<td>□ Unable to sit comfortably on any chair</td>
<td>□ cm</td>
</tr>
<tr>
<td>□ Two canes</td>
<td>□ Normally using a railing</td>
<td>□ On a high chair for 30 minutes</td>
<td></td>
</tr>
<tr>
<td>□ One crutch</td>
<td>□ Normally without using a railing</td>
<td>□ Comfortably, ordinary chair for 1 hour</td>
<td></td>
</tr>
<tr>
<td>□ Cane most of the time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Cane for long walks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Limb-length discrepancy</td>
<td></td>
</tr>
<tr>
<td><strong>Physician name:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Evaluator name:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Motions**

<table>
<thead>
<tr>
<th><strong>Hip flexion:</strong></th>
<th><strong>Abduction:</strong></th>
<th><strong>Internal rotation:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hip extension:</strong></td>
<td><strong>Adduction:</strong></td>
<td><strong>External rotation:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 103-2 Harris hip score.
early diagnosis of this condition depends on the fact that osteoblastic activity and blood flow are increased in the early stages of osteonecrosis. In an advanced stage of disease, the appearance may be one of increased activity in a subchondral distribution owing to osteoblastic activity at the reactive interface around the necrotic segment; however, the center of the osteonecrotic lesion may show much less radionuclide uptake (Figure 103-6) or even a complete lack of activity, reflecting decreased metabolism in the necrotic focus as a result of interruption of blood supply.

In addition to bone scintigraphy, single-photon emission computed tomography (SPECT) maximizes sensitivity. A study comparing conventional radiography, MRI, computed tomography (CT), and Tc-99m MDP three-phase bone scan in diagnosing bisphosphonate-associated osteonecrosis of the jaw showed that CT and MRI were the best at defining the extent of the disease, but that bone scan was the best at identifying disease at an early stage. Bone scan could be an excellent screening tool for the diagnosis of osteonecrosis before further characterization of the lesions using CT or MRI.

CT allows more detailed examination of the femoral head. A star-shaped structure, formed by weight-bearing bone trabeculae, gives the appearance of an asterisk on CT scan (the asterisk sign). This asterisk undergoes a characteristic change in ischemic bone necrosis of the femoral head, and this change was considered important for early detection of osteonecrosis. At a later stage, the collapse of necrotic bone can be well shown (Figure 103-7).

Currently, MRI is the “gold standard” for imaging of osteonecrosis. Most of the staging systems for osteonecrosis are now based on MRI appearance (Table 103-6). MRI of osteonecrosis can show changes earlier than conventional radiography or CT. It can also detect bone marrow edema,
a feature sometimes seen in the early phases of osteonecrosis that is not visible on conventional radiography or CT.

The typical MRI findings in osteonecrosis are intermediate or low signal intensity on T1-weighted images and high signal intensity on T2-weighted images (Figure 103-8). As the disease progresses, the subchondral necrotic lesion is surrounded by a low signal line on T1-weighted images. A high signal line is seen on T2-weighted images, central to the low signal line. This produces the “double-line” sign (Figure 103-9). In advanced osteonecrosis, the necrotic segment exhibits low signal intensity on both T1-weighted and T2-weighted images (Figure 103-10). MRI is done in the sagittal, coronal, and axial planes and includes T1-weighted and T2-weighted sequences. There is excellent correlation between histologic findings and MRI appearance (see Table 103-6).

MRI is an important tool in determining the extent of femoral head involvement in osteonecrosis. Three techniques are used to evaluate this. The first is estimating head involvement. This method was first proposed by Steinberg and colleagues in 1984, and it is defined by the appearance of abnormal signals on T1-weighted images. The degree of head involvement was classified into three categories: less than 15%, 15% to 30%, and greater than 30%. The second method used to evaluate extent is the index of

Figure 103-5  Increased density of the femoral head, loss of the normal spherical shape, and flattening of the superior aspect are characteristic radiographic features of osteonecrosis.

Figure 103-6  Bone scintigraphy of osteonecrosis of both femoral heads using Tc99m methylene diphosphonate showing moderate uptake of radiopharmaceutical at the site of the osteonecrotic segment in the right femoral head and markedly increased uptake at the site of bone repair (straight arrow). The left femoral head (curved arrow) exhibits early-stage disease.

Figure 103-7  A, Computed tomography scan shows osteonecrosis of the femoral head. Although there are several sclerotic foci within the trabecular bone, the integrity of the osseous structures is preserved and the femoral head exhibits normal spherical shape. B, In more advanced stage of osteonecrosis of the femoral head, note increased sclerosis in the posterior aspect (solid arrow) and subchondral collapse of necrotic bone anterolaterally (open arrow).

Figure 103-8  MRI is done in the sagittal, coronal, and axial planes and includes T1-weighted and T2-weighted sequences. There is excellent correlation between histologic findings and MRI appearance (see Table 103-6).
Markers of Disease

The ability to find consistent and reliable markers of disease is always a welcome tool, for diagnosis, determination of extent of the disease, or even determination of risk of acquiring the disease. The measurement of serum and urine carboxy-terminal cross-linking telopeptide of type I collagen (CTX-1), a marker of bone resorption, has been proposed as a method of evaluating the risk of osteonecrosis of the jaw secondary to bisphosphonate usage. Serum osteocalcin is another marker for bisphosphonate-related osteonecrosis of the jaw that has been suggested as a risk predictor because levels were significantly lower in the osteonecrosis group compared with a control group.135

Table 103-7 shows a comparison of various imaging techniques used in the diagnosis and staging of osteonecrosis. Hip arthroscopy is also used in the staging of osteonecrosis. In a study comparing radiography, MRI, and arthroscopy, there was only moderate correlation among the three methods. Arthroscopy was able to detect osteochondral degeneration, not detected by radiography or MRI in 36% of collapsed heads. Figure 103-11 is an algorithm for the diagnosis of osteonecrosis.

<table>
<thead>
<tr>
<th>Type of Appearance</th>
<th>Category of Observations</th>
<th>Histology</th>
<th>MRI Appearance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Fatlike</td>
<td>Premature fatty marrow development in the femoral neck or intertrochanteric region</td>
<td>Normal fat signal; Sclerotic margin may be seen circumscribing lesion</td>
</tr>
<tr>
<td>B</td>
<td>Bloodlike</td>
<td>Bone resorption; replacement by vascular granulation tissue</td>
<td>High signal intensity of inner border; low signal intensity of surrounding rim</td>
</tr>
<tr>
<td>C</td>
<td>Fluid-like</td>
<td>Bone marrow edema</td>
<td>Diffusely decreased signal on T1-weighted images; high signal on T2-weighted images</td>
</tr>
<tr>
<td>D</td>
<td>Fibrotic</td>
<td>Sclerosis owing to reinforcement of existing trabeculae at margin of live bone (repair tissue interface)</td>
<td>Decreased signal on T1-weighted and T2-weighted images</td>
</tr>
</tbody>
</table>

**TREATMENT**

**Surgical Treatment**

Most cases of osteonecrosis ultimately require surgical intervention. There are various surgical techniques ranging from core decompression to total hip replacement. Sometimes surgical procedures can be used in conjunction with

**Figure 103-8** A, On T1-weighted coronal magnetic resonance image of the left hip, the osteonecrotic segment in the subchondral portion of the femoral head shows low signal intensity. B, On T2-weighted coronal image, the necrotic bone exhibits high signal intensity, surrounded by a sclerotic low-signal rim.
nonsurgical approaches, as discussed later. The more advanced the disease, the more extensive the surgery.

The various surgical procedures used in the treatment of osteonecrosis include core decompression, structural bone grafting, vascularized fibula grafting, osteotomy, resurfacing arthroplasty, hemiarthroplasty, and total hip replacement. Table 103-8 shows the typical success rates for each of these procedures.

Arthroscopy is a valuable tool used in the treatment of osteonecrosis. It has been used to determine the position of the core decompression tract to the necrotic part of the femoral head, and arthroscopic débridement has been used in the treatment of osteonecrosis of the capitellum of the humerus in adolescents, Kienböck’s disease, and osteonecrosis of the scaphoid.

Core decompression, which involves the removal of a core of bone from the femoral neck and head, is indicated in less advanced stages of osteonecrosis. The core acts as a vent to reduce intraosseous pressure and intramedullary pressure, reversing ischemia and improving symptoms. Other benefits of core decompression include stimulation of angiogenesis, which leads to improved vascularity during the repair process. The effectiveness of core decompression in the treatment of nontraumatic osteonecrosis was illustrated in 34 patients with 54 affected hips. Mean age at presentation was 38 years. The patients were monitored for a mean duration of 120 months postsurgery. Success was defined as absence of symptoms, no further progression of disease, and no further surgery. Clinical success was established in 26 hips (48%), and radiographic success was established in 20 hips (37%).

Computer-assisted core decompression has been used to provide greater precision in directing the core into the ischemic area and to minimize the duration of radiation exposure to patients. Because early diagnosis improves outcome and there is a high incidence of developing osteonecrosis in a contralateral hip, core decompression is frequently done on both hips simultaneously. This approach adds little risk over unilateral core decompression with the benefit of better outcomes secondary to early surgical treatment of the contralateral hip.

In structural bone grafting, or bone impaction grafting, the bone graft is inserted into the necrotic segment through the core tract. The bone graft acts in similar fashion to a stent, providing support to overlying subchondral bone. The goal is to prevent collapse. This combination of procedures is frequently used in treating stage 1 or 2 osteonecrotic femoral heads. Allogeneic and autologous bone grafts, mostly harvested from the tibia or fibula, are used. When this technique was attempted in patients with stages 3 and 4 lesions, the outcome was generally poor (100% failure
Table 103-7  Comparative Sensitivity and Specificity of Diagnostic Radiologic Imaging Modalities in Osteonecrosis

<table>
<thead>
<tr>
<th>Radiologic Imaging</th>
<th>Earliest Sign Seen</th>
<th>Histologic Correlation</th>
<th>Stage</th>
<th>Degree of Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional radiograph</td>
<td>Crescent sign</td>
<td>Sclerotic rim of reactive bone</td>
<td>2</td>
<td>High</td>
</tr>
<tr>
<td>Computed tomography scan</td>
<td>Asterisk sign</td>
<td>Sclerotic rim surrounding a mottled area of osteolysis and sclerosis</td>
<td>2</td>
<td>High</td>
</tr>
<tr>
<td>Magnetic resonance image</td>
<td>Low signal intensity on T1-weighted images; high signal intensity on T2-weighted images</td>
<td>Bone marrow edema</td>
<td>1</td>
<td>High</td>
</tr>
<tr>
<td>Skeletal scintigraphy</td>
<td>Decreased uptake in subchondral distribution, “cold” spot</td>
<td>Osteonecrosis</td>
<td>1</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Increased uptake in subchondral distribution, “hot spot”</td>
<td>“Creeping substitution”</td>
<td>2</td>
<td>Low</td>
</tr>
</tbody>
</table>

after 2 to 4 years), with progression to collapse and further surgical procedures.\(^{138}\)

Vascularized structural bone grafting also uses the core tract to insert a corticocancellous bone graft into the femoral neck and head along with its vascular pedicle. The vascular pedicle is anastomosed to a nearby vessel, adding a source of blood to the graft. The results of vascularized fibular grafting in the treatment of hips with osteonecrosis showed a survival of 61% of hips at 5-year follow-up and 42% at a median time of 8 years.\(^{139}\) In another study, 197 patients with 226 osteonecrotic hips were treated with a combination of autologous cancellous bone impaction and pedicled iliac bone block transfer. The anastomosis was to the ascending branch of the lateral femoral circumflex artery. Fourteen hips required conversion to total hip arthroplasty because of collapse, severe pain, or both. Of the remaining

![Diagnostic algorithm for osteonecrosis. MRI, magnetic resonance image.](image-url)
1706 PART 15 | CARTILAGE, BONE, AND HERITABLE CONNECTIVE TISSUE DISORDERS

### Table 103-8  Surgical Treatment of Osteonecrosis

<table>
<thead>
<tr>
<th>Surgical Procedure</th>
<th>Rationale</th>
<th>Stages of Osteonecrosis</th>
<th>Outcome</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core decompression</td>
<td>Reduction of intraosseous and intramedullary pressure</td>
<td>Early stages</td>
<td>37% radiographic success, 48% clinical success</td>
<td>Success rate depends on disease stage</td>
</tr>
<tr>
<td>Structural bone grafting</td>
<td>Provide support to overlying subchondral bone</td>
<td>1 or 2</td>
<td>Poor in advanced disease</td>
<td>100% failure rate in stages 3 and 4</td>
</tr>
<tr>
<td>Vascularized fibula grafting</td>
<td>Increase blood flow to graft</td>
<td>2 to 4</td>
<td>96% success in stage 2, 90% in stage 3, and 57% in stage 4</td>
<td></td>
</tr>
<tr>
<td>Osteotomy</td>
<td>Shifting position of osteonecrotic segment out of weight-bearing region</td>
<td>2 and 3</td>
<td>Not available</td>
<td>An alternative to total hip arthroscopy in later stages of disease</td>
</tr>
<tr>
<td>Resurfacing arthroplasty</td>
<td>Preservation of bone and joint mechanics with metallic or ceramic shell over femoral head</td>
<td>Later stages</td>
<td>Mean 7-year success rate is 90%</td>
<td></td>
</tr>
<tr>
<td>Hemiarthroplasty</td>
<td>Replacement of femoral head, preservation of anatomic acetabulum</td>
<td>Later stages</td>
<td>Failure rate for unilateral hemiarthroplasties is 50%-60% at 3 years, for bilateral hemiarthroplasties is 44%</td>
<td>Various techniques available, some with better outcome</td>
</tr>
<tr>
<td>Total hip replacement</td>
<td>Complete replacement of the hip joint</td>
<td>Late stages</td>
<td>17.4% required revision after 10 years</td>
<td>Eventually most patients will require multiple hip replacements</td>
</tr>
</tbody>
</table>

212 hips, 92% were considered a clinical success and 76% were considered radiographically successful. The success rate declined from stage 2 to stage 4 hips (96% for stage 2 hips, 90% for stage 3 hips, and 57% for stage 4 hips). Free vascularized fibula grafting has been compared favorably with other modes of surgical treatment.\(^\text{41}\)

Osteotomy of the femur involves shifting the position of the osteonecrotic segment by making a cut in the proximal femur so that the osteonecrotic segment is rotated or flexed out of the weight-bearing region of the acetabulum and replacing the weight-bearing region with viable bone. Healing of the necrotic region can proceed without the stress of weight bearing. Several different osteotomy techniques have been attempted to salvage hips in stage 2 or 3 osteonecrosis.

Resurfacing arthroplasty uses a metallic or ceramic shell placed over a femoral head that has been débrided of the necrotic area. The potential advantages of resurfacing arthroplasty include preservation of joint mechanics, bone conservation,\(^\text{42}\) more physiologic loading of the bone, a lower incidence of perioperative complications, and easier conversion to total hip arthroplasty in case of failure.\(^\text{143}\)

Complications of this procedure include femoral neck fractures, a secondary osteonecrosis when the procedure is done for other reasons,\(^\text{44}\) and increased metal ion levels.\(^\text{145}\) Resurfacing arthroplasty has been recommended for patients with later-stage osteonecrosis including those with femoral head collapse.\(^\text{46}\) A retrospective study compared the results of limited femoral head resurfacing and total hip arthroplasty in 30 consecutive patients with Steinberg stage 3 or 4 disease. The survival rate at a 7-year mean follow-up period for the resurfacing group was 90%, whereas the survival rate at a mean 8-year follow-up for the total hip arthroplasty group was 93%.\(^\text{47}\) A recent level 3 therapeutic study showed that hip resurfacing success rates at a 5-year follow-up were comparable with those of total hip arthroplasty in osteonecrosis patients younger than 25 years of age.\(^\text{148}\)

In hemiarthroplasty, only part of the hip joint is replaced. The original acetabulum is preserved, but the femoral head is replaced with a prosthesis. Two kinds of prostheses are used—a unipolar prosthesis and a bipolar prosthesis. In a unipolar prosthesis, the articulation is between the artificial femoral head and the acetabulum. In the bipolar prosthesis, presently the most frequently used, the articulation is within the prosthesis itself. Failure rates for hemiarthroplasties in osteonecrosis are 50% to 60% at 3 years for unipolar prostheses and 44% for bipolar prostheses. Another study evaluated the success rate of Charnley/Bicentric hemiarthroplasty in the treatment of Ficat and Arlet stage 3 osteonecrosis of the femoral head. Failures include three hips that needed to be revised to cementless total hip replacement, two hips with radiographic changes of loosening and imminent failure, and one hip with progressive loss of joint space and secondary degenerative changes. The success rate was 84.2% after a mean of 56 months.

Total hip arthroplasty is complete replacement of the hip joint with a prosthesis including the femoral head and the acetabulum. In a study of 55 consecutive hip arthroplasty procedures, cementless total hip arthroplasty was shown to provide favorable results in advanced-stage osteonecrosis of the femoral head. Although 10 of the 48 hips available for follow-up after a minimum of 5 years required revision, all of these patients had Ficat and Arlet stage 3 or 4 disease. A study of 53 hips in 41 patients treated with cemented total hip replacement showed that at a minimum of 10 years of follow-up, 17.4% required revision. Compared with cemented total hip replacements done for other conditions, osteonecrosis had a greater risk for loosening of acetabular and femoral components. A survivorship analysis of cemented total hip replacements in renal transplant patients with osteonecrosis of the femoral head showed that there was excellent survival after 10 years (98.8%). After 20 years, the survival rate decreased to 63.8%.

In osteonecrosis of the jaw, the most common surgical procedure is resection of the affected bone.\(^\text{149}\) Conservative
management does not seem to alter the natural course of the disease. Electrical stimulation has been used in the treatment of osteonecrosis, in conjunction with core decompression. Electrical stimulation enhances osteogenesis and neovascularization. It also alters the balance between osteoblast and osteoclast activity, resulting in increased bone deposition and decreased bone resorption. Delivery of electrical stimulation can be done by direct current (DC), pulsed electromagnetic field, and capacitance coupling. The success of electrical stimulation in the treatment of osteonecrosis has been rather mediocre. Eleven hips in eight patients with Ficat stage 2 osteonecrosis who underwent core decompression and placement of an electric stimulating coil within the core in the anterosuperior segment of the femoral head were studied. Of these, five hips required reoperation and six hips had progressive deterioration 13 months after initial placement of the coil. In addition, there was little histologic evidence that the coil did indeed generate new bone deposition around itself.

On the other hand, a study compared the effectiveness of conservative nonsurgical treatment with core decompression with or without direct current electrical stimulation. The clinical symptom scores and the rate of progression to arthroplasty were best in the group with core decompression treatment has also been used but carries a higher recurrence rate. A larger extent of surgical excision and a higher number of surgical débridements were associated with a lower recurrence rate. Other modes of surgical therapy for osteonecrosis of the jaw include bone-contouring procedures; fluorescence-guided bone-contouring procedures; and segmental osteotomies, but these are generally reserved for more severe cases. Nonsurgical treatment including hyperbaric oxygen therapy and low-intensity laser therapy are controversial but have been used to treat osteonecrosis of the jaw.

**Nonsurgical Approaches**

The key to the successful treatment of osteonecrosis is early detection. The choice of conservative nonsurgical versus more aggressive surgical options depends on the clinical and pathologic staging of the disease. Figure 103-12 is an algorithm for the treatment of osteonecrosis.

Nonsurgical treatment of osteonecrosis of the femoral head includes refraining from weight bearing on the affected joint, analgesic and anti-inflammatory medications, and physiotherapy. Conservative medical treatment is effective only in the early stages for symptomatic relief. Nonsurgical
and DC electrical stimulation and worst in the nonoperative group. Capacitive coupling can be done with or without core decompression and grafting. Core decompression and grafting were done on 40 patients with stage 1 to 3 osteonecrosis; half of the patients were active capacitive coupling units with electrodes over the femoral head for 6 months. The control group was 25 patients with osteonecrosis who were treated conservatively. Two- and 4-year follow-up showed that core decompression with or without capacitive coupling provided better clinical and radiologic outcome than conservative treatment. Capacitive coupling did not improve the results further when used with core decompression and grafting.

Extracorporeal shock wave therapy has been used in the treatment of osteonecrosis of the femoral head. A study of 48 patients and 57 hips compared extracorporeal shock wave therapy with core decompression and bone grafting. Twenty-three patients with 29 affected hips were assigned to the shock wave group, and the remaining patients and hips received surgical treatment. The patients in the shock wave group were given treatment of 6000 pulses of shock waves at 28 kV to the affected hip. The patients were evaluated radiographically and by their reports of symptoms (pain), Harris hip scores, and quality of life (daily work activity assessment). Shock wave therapy produced better results than the nonvascularized bone grafting procedure, with comparatively less progression of disease. In 35 patients with 47 osteonecrotic hips, the use of shock wave therapy led to improvements in serum nitric oxide levels, angiogenic factors such as VEGF, and osteogenic factors such as bone morphogenetic protein-2 (BMP-2) and osteocalcin. Levels of inflammatory markers were reduced. It is interesting to note that although these changes did not persist beyond several months, the clinical and radiographic improvement, present in 83% of hips, was present after 12 months.122

Conservative treatment of osteonecrosis of the talus is not promising, and the affected ankles generally continue to progress, requiring either core decompression or arthrodesis. Conservative treatment of bisphosphonate-induced osteonecrosis of the jaw includes cessation of bisphosphonate usage or surgical debridement. Good oral hygiene, regular dental assessment, and avoidance of dental procedures during bisphosphonate usage can prevent onset of osteonecrosis.

## RECENT DEVELOPMENTS

### Prevention versus Treatment

A recent study evaluated the role of antioxidants in the treatment of osteonecrosis. Japanese white rabbits were divided into two groups and fed either a normal diet or a normal diet supplemented with α-tocopherol. Osteonecrosis developed in 14 of 20 rabbits in the control group but only in 5 of 21 rabbits in the experimental group. This suggests that oxidative stress may play a role in the pathogenesis of osteonecrosis and that there may potentially be a role for antioxidants such as vitamin E.133

A group of researchers studied the use of adrenocorticotropic hormone (ACTH) in rabbits to prevent corticosteroid-induced osteonecrosis and found that if ACTH is administered along with depot methylprednisolone acetate (DepoMedrol), osteonecrosis is reduced. The authors of this study believe that ACTH enhances osteoblast support and stimulates the production of vascular endothelial growth factor (VEGF), which stimulates the generation of new blood vessels. The result is an increase in blood flow to the vulnerable areas of bone, preventing cell death and reducing the likelihood of osteonecrosis.154

### Mesenchymal Stem Cells

Corticosteroids interfere with the balance of adipogenesis and osteogenesis in the differentiation of mesenchymal stem cells. Corticosteroids shunt uncommitted osteoprogenitor cells in the bone marrow into the adipocytic pathway, leading to reduced osteoblast formation. Corticosteroids have also been shown to reduce vascular endothelial growth factor, which leads to a reduction in new blood vessel formation and potentially can lead to bone death. Alcohol has a similar effect on the differentiation of progenitor cells.

The balance between adipogenesis and osteogenesis has been targeted as a potential site for the treatment of osteonecrosis. Multipotential mesenchymal stem cells from femoral bone marrow near osteonecrosis sites are able to express messenger RNA aggrecan and type II collagen. Both are deposited into the bone matrix. These features are characteristic of chondrogenic differentiation. The mesenchymal stem cells can be differentiated into osteocytic lineage in vitro.

A pilot study evaluating the effectiveness of implantation of autologous bone marrow cells in the treatment of osteonecrosis used core decompression to implant stem cells into the necrotic lesions of the femoral head. The patients were divided into two groups—one that received core decompression alone as treatment for osteonecrosis (the control group) and one that received autologous bone marrow cell implantation along with core decompression (the treatment group). The patients were followed for 24 months, and at that time, 5 of 8 hips in the control group, but only 1 of 10 in the treatment group, advanced to stage 3 osteonecrosis. In addition, there was greater improvement in pain and joint symptoms in the treatment group and the treatment seemed to be safe. Because of the small number of patients involved, further studies are necessary to confirm these results.

Twenty-eight patients with 44 necrotic hips were treated with percutaneous decompression and autologous bone marrow mononuclear cell infusion. Patients were followed for a minimum of 2 years and evaluated for clinical and radiographic progression of the disease. There seemed to be overall slowing in the progression of the disease stage. The mean Harris hip score improved from 58 to 86.

### OUTCOME

The natural history of osteonecrosis depends on the size of the infarcted segment, the site of occurrence, and the clinical and radiologic staging of the disease. At the onset of the disease, range of motion may be well preserved but gradually deteriorates over time. In the early stages of the disease, when it is still reversible, patients may be asymptomatic.
Many patients therefore present with advanced disease. Although spontaneous resolution of femoral head osteonecrosis can occur, it is rare and occurs only when lesion size is small. A study of the prognosis of osteonecrosis of the femoral head as a function of symptoms (pain) and radiographic findings showed that in patients who were asymptomatic and had normal radiographs, progression of the disease was slow, with only 1 of 23 hips progressing to pain and radiographic changes after 5 years. If radiographic changes are already present, disease progresses to pain in 14 of 19 patients after 5 years. In a study of stage 1 osteonecrotic lesions of the hip diagnosed with MRI, 40 patients were followed for an average of 11 years. All patients had stage 1 lesions on the contralateral hip. Overall, 35 of the 40 stage 1 hips became symptomatic and 29 hips showed collapse. The mean interval between diagnosis and collapse was 92 months, whereas the mean interval between symptoms and diagnosis was 80 months. Most stage 1 hips eventually progress to a more advanced stage, requiring surgery, so these hips should be monitored closely.

**SUMMARY**

Osteonecrosis is a potentially debilitating condition with significant morbidity despite medical interventions or surgery. Corticosteroids are the most common cause of osteonecrosis, and corticosteroid-induced osteonecrosis can be reproduced in animal models. The pathogenesis of osteonecrosis is multifaceted and still not completely understood. Why is it that corticosteroid-induced osteonecrosis is more common in patients with certain underlying diseases and not in others? Is there a genetic basis for osteonecrosis? Common pathogenic mechanisms known to be involved in osteonecrosis include osteoblast/osteoclast survival and apoptosis, lipid metabolism, and coagulation abnormalities. However, it is still unclear how these mechanisms interrelate with each other. In order to better appreciate the risk factors involved in osteonecrosis, a more complete understanding of the pathogenesis is necessary. Until then, the physician should always maintain a high index of suspicion for osteonecrosis whenever known risk factors are present, especially use of corticosteroids and alcohol.

**Selected References**


Full references for this chapter can be found on www.expertconsult.com.
References


bone early after steroid administration is involved in the pathogenesis of
111. Kerachian MA, Seguin C, Harvey EJ: Glucocorticoids in osteonecro-
sis of the femoral head: a new understanding of the mechanisms of
113. Feng Y, Yang SH, Xiao BJ, et al: Decreased in the number and func-
tion of circulation endothelial progenitor cells in patients with avas-
114. Jones JP Jr, Ramirez S, Doty SB: The pathophysiologic role of fat in
115. Tokuhara Y, Wakiyama S, Oda Y, et al: Low levels of steroid-
metabolizing hepatic enzyme (cytochrome P450 3A) activity may
enable responsiveness to steroids and may increase risk of steroid-
induced osteonecrosis even with low glucocorticoid dose, J Orthop Sci
116. Saito S, Ohzono K, Ono K: Early arteriopathy and postulated patho-
genesis of osteonecrosis of the femoral head. The intracapital arteri-
toxicity of idiopathic osteonecrosis, Orthop Clin North Am 16:669–
associated with ischemic necrosis of bone in systemic lupus erythe-
polymorphisms in patients with nontraumatic femoral head osteone-
inhibitor-1 gene, hypobilirubinemia, and osteonecrosis, Clin Orthop Relat
polymorphism predicts osteonecrosis in children with acute lympho-
blastic leukemia: a report from the Children’s Oncology Group, Blood
122. Hirata T, Fujioka M, Takahashi KA, et al: ApoB C762T polymor-
phism predicts risk for steroid-induced osteonecrosis of the
phenotype of Apo(a) is a risk factor of corticosteroid-induced osteo-
nerosis of the femoral head after renal transplant, J Rheumatol
A study of 101 hips treated with vascularized fibular grafting, J Bone
128. Kohornilas AV, Beris AE, Lykissas MG, et al: Femoral head osteo-
 necrosis: why choose free vascularized fibula grafting, Microsurgery
129. Venditelli PA, Lavigne M, Girard J, Roy AG: A randomised study
comparing resection of acetaluber bone at resurfacing and total hip
130. Greula MJ: Resurfacing arthroplasty in osteonecrosis of the hip,
femoral heads after failed resurfacing arthroplasty of the hip, J Bone
132. Shimmin AJ, Bare J, Back DL: Complications associated with hip
hemiarthroplasty for the treatment of osteonecrosis of the femoral head,
134. Mont MA, Rajadhyaksha AD, Hungerford DS: Outcomes of limited
resurfacing arthroplasty compared with total hip arthroplasty
who have osteonecrosis and are 25 years or under, Clin Orthop Relat
136. Carlson ER, Basile JD: The role of surgical resection in the manage-
ment of bisphosphonate-related osteonecrosis of the jaws, J Oral
137. Poutte C, Bauer F, Tischer T, et al: Fluorescence-guided bone resec-
tion in bisphosphonate-associated osteonecrosis of the jaws, J Oral
treatment and bisphosphonate-induced osteonecrosis of the jaw: a
139. Wang CJ, Yang YJ, Huang CC: The effects of shockwave on systemic
concentrations of nitric oxide level, angiogenesis and osteogenesis
140. Kuribayashi M, Fujioka M, Takahashi KA, et al: Vitamin E prevents
glucocorticoid-induced osteonecrosis of bone, Proc Natl Acad Sci U
142. Crues RL: Steroid-induced avascular necrosis of the head of the
humerus, Natural history and management, J Bone Joint Surg Br
162. Abylaev Zh A, Bukhanal AI: [Possible mechanisms of the development of osteoporosis in patients with chronic lead poisoning], Gig Tr Prof Zabol 2:3–31, 1990.


Relapsing polychondritis (RPC) is a rare autoimmune disease of unknown etiology, characterized by episodic inflammation of cartilaginous structures throughout the body (e.g., the ears, nose, upper respiratory tract, chest wall, joints). Associated diseases such as vasculitis or myelodysplastic syndromes occur in up to one-third of cases. The severity of the inflammatory process frequently requires the use of immunosuppressive agents. Prognosis is variable with a 5-year survival rate of 45% to 95% depending on organ involvement and disease or treatment complications.

**EPIDEMIOLOGY**

The worldwide incidence of RPC is unknown, but figures from Rochester, Minnesota, suggest an annual incidence of 3.5 per million in that community. RPC occurs in all racial groups and with similar frequency in men and women. Peak age of onset is 40 to 50 years, but it has been described in children and in people older than 80 years of age. No familial pattern of inheritance has been shown.

**PATHOLOGY**

Cartilage has a cellular component consisting of chondrocytes and an extracellular matrix made from interlinking fibrils of type II collagen, other collagens, hydrophilic proteoglycan aggregates, and a variety of matrix proteins. One such protein, matrilin-1, is exclusive to the respiratory tract, ears, and xiphisternum in adults and is only found in articular cartilage before skeletal maturity. Cartilage is an avascular structure that derives essential nutrients from adjacent tissue. The composition of cartilage confers resilience to external compressive forces. In normal adults, turnover of cartilage components is slow, with incomplete repair processes. Degradation of collagen fibrils occurs with age resulting from imbalances between naturally occurring proteolytic enzymes and their inhibitors.

In RPC, typical features on hematoxylin-eosin staining include a loss of normal cartilage basophilia and a perichondrial inflammatory infiltrate composed of lymphocytes, neutrophils, eosinophils, and plasma cells. Necrosis and a reduction in cartilage components may be observed. The cartilage is subsequently replaced by fibrotic tissue. Immunofluorescence may demonstrate immunoglobulin and complement components in the perichondrial tissue and vessels.

**PATHOGENESIS**

Although the pathogenesis of RPC is unknown, evidence supports the concept of an autoimmune process. Researchers have observed autoantibodies to types II, IX, and XI collagen in patients with this disease, while autoantibodies to matrilin-1 have been detected in those with respiratory tract involvement. Specific cytokines such as monocyte chemoattractant protein-1, macrophage inflammatory protein 1b, and interleukin-8 are significantly elevated in active RPC compared with controls. Furthermore, there is a human leukocyte antigen (HLA) class II association with RPC. A twofold increase in HLA-DR4 has been found in RPC, although a specific subtype has not been identified.

Several animal models have helped elucidate some of the pathogenic mechanisms underlying RPC. Chondritis can be induced in certain rat or mice species following injection of type II collagen. In addition, mice expressing HLA-DQ6ab8ab transgenes develop a spontaneous form of polychondritis with auricular, nasal, and joint involvement. Experiments using the matrilin-1-induced rodent model of RPC have demonstrated the importance of T cells, B cells, and complement components in pathogenesis of this disease. Deletion of interleukin (IL)-10 in this model resulted in increased disease severity, suggesting a role for this endogenous cytokine in suppressing episodic inflammation.

The processes involved in the initiation of disease, escalation of the inflammatory response, and subsequent cartilage destruction are poorly understood. In a healthy individual, cartilage is an immunologically privileged site. However, in RPC, cartilage components are exposed and vulnerable to immunologic attack, resulting in a perpetuating process of systemic inflammation and local tissue damage. In genetically susceptible individuals, cartilage...
Table 104-1  Diagnostic Criteria

<table>
<thead>
<tr>
<th>Major Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proven inflammatory episodes of ear cartilage</td>
</tr>
<tr>
<td>Proven inflammatory episodes of nose cartilage</td>
</tr>
<tr>
<td>Proven inflammatory episodes of laryngotracheal cartilage</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Minor Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye inflammation</td>
</tr>
<tr>
<td>Hearing loss</td>
</tr>
<tr>
<td>Vestibular dysfunction</td>
</tr>
<tr>
<td>Sero-negative inflammatory arthritis</td>
</tr>
</tbody>
</table>

Diagnosis is made by two major criteria or one major plus two minor criteria. Histologic examination of affected cartilage is not required.


Microtrauma or molecular mimicry between an infectious/inciting agent and cartilage structures might instigate a cycle of inflammation similar to other autoimmune disorders.

**CLINICAL FEATURES**

The diagnostic criteria for RPC are outlined in Table 104-1. Table 104-2 documents the frequency of clinical presentations.

**Otorhinologic Disease**

The most characteristic manifestation of RPC is auricular chondritis, in which the patient develops acute pain, redness, and swelling of the cartilaginous upper two-thirds of the outer ear, sparing the lobe (Supplemental Figure 104-1 on www.expertconsult.com). It may be unilateral or bilateral and resolves spontaneously over a period of several weeks. Repeated episodes result in visible damage to the ear, with a deformed, flaccid appearance (Figures 104-1A and 104-2). Although classic in RPC, it is the presenting feature in only 40%, whereas up to 85% of patients eventually develop this feature over the course of their disease. Such swelling of the outer ear causes temporary conductive deafness. However, sensory-neural deafness may result from an associated vasculitis of the internal auditory artery or its branches, leading to additional vertigo in some patients. A similar pattern of chondritis in the nose may cause collapse of the nasal bridge and an alteration of the facial appearance—the so-called saddle-nose deformity (Figure 104-1B).

**Respiratory Disease**

Involvement of the upper respiratory tract may be life threatening and should be investigated at an early stage in the disease course. Approximately 50% of patients with RPC develop respiratory problems, which are associated with a worse prognosis. Hoarseness, dysphonia, a persistent dry cough, or anterior neck tenderness may indicate laryngeal or tracheal disease. Inflammation of the tracheobronchial tree may lead to tracheomalacia, dynamic obstruction, and acute respiratory failure, whereas repeated inflammatory episodes cause subglottic stenosis, chronic dyspnea, and increased susceptibility to infection. Obstruction may be precipitated by attempts at intubation or bronchoscopy. Computed tomography and bronchoscopy appearances of one patient with RPC are demonstrated in Figure 104-1C and D. The development of costochondritis may result in chest pain, leading to additional respiratory symptomatology.

**Cardiovascular Disease**

Large and small vessel disease can occur during the course of RPC in up to 10% of cases. Inflammation of the aorta, most likely to occur at the level of the root or arch, leads to aneurysm formation and aortic incompetence, which may develop acutely with minimal prior symptoms. Aortic valve disease may be due to progressive dilatation of the aortic ring rather than to active local inflammation. Mitral regurgitation may result from valvitis or papillary muscle involvement, whereas myocarditis may lead to cardiac failure and conduction system abnormalities. Small vessel vasculitis has also been reported and may cause end-organ damage to the skin, kidneys, testes, sclera, cochleovestibular system, or other areas. Arterial and venous thromboses have been described in association with RPC.

**Eye Disease**

Approximately 50% of patients with RPC will develop ocular disease, most commonly episcleritis and scleritis. Uveitis, retinal vasculitis, and optic neuritis are rare complications. Keratoconjunctivitis sicca may occur with associated Sjogren’s syndrome. Extraocular muscle palsy, periorbital edema, and proptosis may also occur as part of the RPC symptom complex.

**Renal Disease**

Proteinuria and/or hematuria may be observed on urinalysis in up to 26% of patients with RPC. Renal pathology has been reported to show mesangial proliferation and segmental necrotizing crescentic glomerulonephritis.

Table 104-2  Clinical Features of Relapsing Polychondritis

<table>
<thead>
<tr>
<th>Clinical Feature</th>
<th>Frequency At Presentation</th>
<th>Frequency In Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auricular chondritis</td>
<td>39</td>
<td>85</td>
</tr>
<tr>
<td>Nasal chondritis</td>
<td>24</td>
<td>54</td>
</tr>
<tr>
<td>Arthritis</td>
<td>36</td>
<td>52</td>
</tr>
<tr>
<td>Eye disease</td>
<td>19</td>
<td>51</td>
</tr>
<tr>
<td>Laryngotracheal bronchial disease</td>
<td>26</td>
<td>48</td>
</tr>
<tr>
<td>Hearing loss</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>Rash</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>Systemic vasculitis</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Valvular dysfunction</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Costochondritis</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Supplemental Figure 104-1  A-F, This series of photographs depicts the ear of a 56-year-old man with relapsing polychondritis. His disease was uncontrolled for 1 year during which the damage shown developed in the cartilaginous structures of the ear.
Peripheral limb ulcers, livedo reticularis, panniculitis, superficial thrombophlebitis, urticaria, and angioedema. Leuko-cytoclastic vasculitis, dermal vessel thrombosis, or septal panniculitis may manifest on biopsy.

Joint Disease

Musculoskeletal manifestations are common and develop in 75% over the course of the disease. Involvement of the joints may be oligoarticular or polyarticular, most commonly affecting the ankles, wrists, hands, and feet. However, other joints, including the sacroiliac joints, may also show evidence of disease, although it is possible that the presence of sacroiliitis relates to the coexistence of a spondyloarthropathy. The arthritis associated with RPC is typically episodic and nonerosive, and it does not correlate with activity of RPC.

ASSOCIATED DISORDERS

RPC has been described in association with a variety of other disorders including hematologic, rheumatic, and vasculitic syndromes. Its coexistence with Behçet's disease has led to the acronym MAGIC syndrome: mouth and genital...
from RPC. *Pseudomonas aeruginosa* or *Staphylococcus aureus* are potential causes of unilateral otitis externa, particularly in immunocompromised patients. Other infections associated with chondritis include leprosy and syphilis. Damage to the nose cartilage may be due to trauma, local infections, or granulomatous lesions such as antineutrophil cytoplasmic antibody (ANCA)-associated granulomatous vasculitis or lethal midline granuloma. Intranasal cocaine use should also be considered. A wide variety of inflammatory, rheumatic, and infectious conditions may mimic some features of RPC, particularly when multiple organs are involved. Cogan’s syndrome is an inflammatory autoimmune condition that results in keratitis and vestibulo-auditory symptoms and may be associated with vascular inflammation, although it is not linked with chondritis. Causes of systemic vasculitis (e.g., ANCA-associated vasculitis, polyarteritis nodosa, Behçet’s syndrome, rheumatoid arthritis, or those associated with connective tissue disorders) should be taken into consideration, especially when the presentation is atypical. Valvular heart disease, particularly aortic incompetence and aortic root dilatation, may occur in conditions such as Marfan’s syndrome, syphilis, and idiopathic medial cystic necrosis and can also develop in relation to diseases that may coincide with RPC (e.g., ankylosing spondylitis).

**INVESTIGATIONS**

**Routine Laboratory Tests**

Routine laboratory tests should be performed and may demonstrate nonspecific findings such as anemia or mild thrombocytosis. The acute phase response is typically elevated during periods of inflammation. Urinalysis should always be performed to determine the presence of subclinical renal disease or infection. Autoimmune tests are nonspecific, except where associated diseases are present concurrently.

**Tissue Sampling/Histopathology**

In typical presentations of RPC, a biopsy is not indicated for diagnosis. However, where the diagnosis is unclear, a biopsy of inflamed cartilage may show evidence of perichondritis, as described earlier. Skin rashes may demonstrate leukocytoclastic vasculitis; other frequent findings include panniculitis, neutrophilic dermatoses, or dermal vessel occlusion.

**Pulmonary Investigations**

Clinicians should conduct pulmonary function tests in all suspected cases of RPC and should be repeated in the event of new respiratory symptoms. If abnormal, computed tomography of the chest is recommended to detect the presence of tracheal or bronchial stenoses or dynamic airway collapse, which may be evident only during the expiratory phase of the respiratory cycle. Extrapulmonary disease from costochondritis may cause a restrictive pattern on pulmonary function testing.

Bronchoscopy is warranted only if specifically indicated in order to avoid inadvertent damage to the upper airways, which poses the risk of precipitating respiratory failure. However, indirect laryngoscopy is helpful to monitor airway

---

Table 104-3  Diseases Associated with Relapsing Polychondritis

<table>
<thead>
<tr>
<th>Rheumatic Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Seronegative spondyloarthropathies</td>
</tr>
<tr>
<td>Connective tissue disorders</td>
</tr>
<tr>
<td>Vasculitides</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hematologic Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myelodysplasia</td>
</tr>
<tr>
<td>Pernicious anemia</td>
</tr>
<tr>
<td>Acute lymphoblastic leukemia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Endocrine Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I diabetes mellitus</td>
</tr>
<tr>
<td>Thyroid disorders: Hashimoto’s/Graves’ disease/hypothyroidism</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>Primary biliary cirrhosis</td>
</tr>
<tr>
<td>Retroperitoneal fibrosis</td>
</tr>
<tr>
<td>Myasthenia gravis</td>
</tr>
</tbody>
</table>
involved. Bronchoscopic ultrasound can also be used to monitor disease.\textsuperscript{18}

**Cardiac Investigations**

Echocardiography is necessary to evaluate the aortic root and heart valves.

Many systemic inflammatory diseases are associated with accelerated atherosclerosis, and this may also be true for patients with RPC, although due to small numbers of cases, this has not been widely described. Nevertheless, other cardiac disease risk factors should be ascertained.

**Ocular Investigations**

Routine retinoscopy should be performed at diagnosis. Referral to a specialist unit is recommended if abnormalities are detected and/or if symptoms develop.

**Musculoskeletal Tests**

Plain radiographs of involved joints may show joint space narrowing or periarticular osteopenia. Erosions are not observed, unless there is an associated rheumatic disorder. Dual x-ray absorptiometry may demonstrate osteopenia as a consequence of the underlying inflammation or general debility.

**Additional Investigations**

At present genetic testing offers little added benefit. Markers of cartilage autoimmunity such as antibodies to types II, IX, or XI collagen or matrilin-1 are not routinely available, and their role in disease activity monitoring is unclear. A recent report highlighted the relevance of antiglutamate receptor GluR\textsubscript{2} autoantibodies in the cerebrospinal fluid and sera of a patient with limbic encephalitis.\textsuperscript{29} The potential use of other biomarkers such as urinary type II collagen breakdown products has been demonstrated.\textsuperscript{19} Additionally, a recent study showed that serum levels of cartilage oligomeric matrix protein (COMP) are elevated during disease flares in RPC.\textsuperscript{40}

**TREATMENT**

Nonsteroidal anti-inflammatory drugs and low-dose corticosteroids may successfully control minor inflammation of the nose, ear, or chest wall. Dapsone may also be used for non-critical cartilaginous inflammation,\textsuperscript{1,23} and colchicine is of reported use in auricular chondritis.\textsuperscript{41} However, the presence of potentially life-threatening disease or severe chondritis requires high-dose prednisone, typically at a dose of 1 mg/kg, reducing slowly as the inflammation recedes.\textsuperscript{1,23} Although the RPC is usually steroid responsive, it can be difficult to withdraw these drugs completely without risking a flare in disease activity and many patients require maintenance doses of prednisone and consideration of alternative immunosuppression. Disease-modifying treatment is indicated in patients who require a steroid-sparing agent and in whom serious manifestations such as vascular or respiratory disease have developed.

Due to the rarity of this disease, there are no controlled trials of therapeutic interventions in RPC. However, methotrexate, cyclophosphamide, mycophenolate mofetil, azathioprine, chlorambucil, and cyclosporine have had successful therapeutic outcomes in some patients.\textsuperscript{54-61} Varied outcomes have been reported with leflunomide.\textsuperscript{52,51} Plasmapheresis has been employed in recalcitrant cases.\textsuperscript{4,23} There are several case reports and small case series detailing the use of biologic agents in this disease. However, there have been both positive and negative reports in relation to the anti–tumor necrosis factor agents and anti–B cell drugs.\textsuperscript{24,52-57} Interleukin-1 receptor antagonist has demonstrated benefit in a small number of reported cases.\textsuperscript{58,59} Tocilizumab, an anti-IL-6 agent, has had beneficial effects in this disease.\textsuperscript{60} A successful outcome for autologous stem cell transplantation in treatment-resistant RPC has also been reported.\textsuperscript{61} It is not clear, however, if any of these treatments overtly influence the natural history of this disease.

A broad approach to the management of RPC is essential. Awareness of disease complications and recognition of early pathology may be lifesaving. Special consideration should be given to the management of upper airway involvement, and the anesthesia team should be made aware of disease in this region before surgical intervention. Symptomatic airway obstruction might require tracheostomy, tracheal stenting, or nighttime positive pressure ventilation to prevent airway collapse while the patient is asleep. There is limited experience with laryngotracheal reconstruction in patients with this disease, and such interventions should only take place during disease quiescence.\textsuperscript{62} Heart surgery to treat aortic root dilatation, aneurysm formation, or valvular disease has a good outcome in most cases, although postoperative dehiscence is a complication in 12%.\textsuperscript{53} Most cases require postoperative immunosuppression. For patients with sensorineural deafness, cochlear implantation is usually beneficial with significant restoration of hearing.\textsuperscript{54} Neither hormonal activity nor pregnancy is thought to influence disease activity in RPC.\textsuperscript{63} However, several case reports have described the treatment of disease flares during pregnancy, typically with corticosteroids, with successful maternal and fetal outcomes.\textsuperscript{63-65}

**OUTCOME**

RPC is an episodic, progressive disease that results in tissue destruction of target organs. Some patients have a mild course, complicated only by recurrent chondritis, whereas others develop potentially life-threatening problems. The most common cause of death is pulmonary infection, arising as a result of the disease or its immunosuppressive treatment.\textsuperscript{6} In young patients, adverse prognostic factors at presentation include anemia, hematuria, upper airway disease, arthritis, and saddle nose deformity, whereas in older patients (more than 51 years of age), only anemia at presentation predicts increased mortality.\textsuperscript{20} The presence of vasculitis worsens prognosis, with a 5-year survival rate of 45%.\textsuperscript{21} In 1986 Michet and and colleagues\textsuperscript{20} reported the 10-year survival rate to be 55%, whereas in 1998, perhaps due to improvements in disease monitoring and treatment, Trentham and colleagues\textsuperscript{8} described a 94% survival rate.
Future Directions

An international database of RPC, detailing modes of presentation, clinical characteristics, and therapeutic responses, would be a valuable resource in learning more about the epidemiology and treatment of RPC. It would also help clarify some of the discrepancies in the current literature relating to presentation and prognosis of this rare disease. New therapies, with specific cellular or molecular targets, offer promise to patients with chronic inflammatory diseases but should ideally be accompanied by definitive knowledge of their mechanism of action in particular disease subtypes. Increased awareness of the immunogenic precipitants involved in RPC may help to guide future treatments and reduce morbidity and mortality in such patients.

References

2. Deleted in proofs.
23. Deleted in proofs.

The references for this chapter can also be found on www.expertconsult.com.
Heritable disorders of connective tissues are a diverse group of disorders and can be associated with extreme variation in height ranging from very short (dwarfs) to tall stature.

The osteochondrodysplasias or skeletal dysplasias are a heterogeneous group of more than 450 disorders frequently associated with profound short stature and orthopedic complications.

The molecular mechanisms have been elucidated in many of these disorders providing for improved clinical diagnosis and reproductive choices for affected individuals and their families.

Mechanism-based treatment options that might improve the quality of life and life span in individuals affected with osteogenesis imperfecta and Marfan syndrome are being investigated. Newer advances in understanding the underpinnings of altered pathways in these disorders are providing potential new targets for treatment.

Heritable disorders of connective tissues are a heterogeneous group of disorders characterized by abnormalities in skeletal tissues including cartilage, bone, tendon, ligament, muscle, and skin. These disorders, originally defined by McKusick, have been classified on the basis of clinical findings and molecular criteria. They are subclassified into disorders that primarily affect cartilage and bone (the skeletal dysplasias) and disorders that have a more profound effect on connective tissue including Ehlers-Danlos syndrome (EDS), Marfan syndrome, and other disorders manifested by abnormal extracellular matrix molecules. The skeletal dysplasias are associated with abnormalities in the size and shape of the appendicular and axial skeleton and frequently result in disproportionate short stature. Until the early 1960s, most individuals with short stature were considered to have pituitary dwarfism, achondroplasia (short-limb dwarfism), or Morquio disease (short-trunked dwarfism). Presently, there are more than 450 well-characterized disorders that are classified primarily on the basis of clinical, radiographic, and molecular criteria. Disorders of connective tissue are genetic defects that result from mutations in genes that encode extracellular matrix proteins, transcription factors, tumor suppressors, signal transducers, enzymes, chaperones, intracellular binding proteins, ribonucleic acid (RNA) processing molecules, and genes of unknown function.

SKELETAL DYSPLASIAS

The skeletal dysplasias, or osteochondrodysplasias, are defined as disorders that are associated with a generalized abnormality in the skeleton. Although each skeletal dysplasia is relatively rare, collectively, the birth incidence of these disorders is almost 1 in 5000. These disorders range in severity from “precocious” arthropathy to perinatal lethality owing to pulmonary insufficiency. Individuals with these disorders can have significant orthopedic, neurologic, and psychological complications. Many of these individuals seek medical attention for orthopedic complaints owing to ongoing pain, arthritic complaints in large joints, and back pain primarily caused by ongoing abnormalities in bone and cartilage frequently leading to spinal stenosis.

Embryology

The human skeleton (from the Greek, skeletos, “dried up”) is a complex organ consisting of 206 bones (126 appendicular bones, 74 axial bones, and 6 ossicles). The skeleton including tendons, ligaments, and muscles in addition to cartilage and bone has multiple embryonic origins and serves many key functions throughout life such as linear growth, mechanical support for movement, a blood and mineral reservoir, and protection of vital organs.

The patterning and architecture of the skeleton occurs during fetal development (see Chapter 4). During that period, the number, size, and shape of the future skeletal elements are determined, a process that is under complex genetic control. Uncondensed mesenchyme undergoes cellular condensations (cartilage anlagen) at the sites of future bones, and this occurs via two mechanisms. In the process of endochondral ossification, mesenchyme first differentiates into a cartilage model (anlagen), and then the center of the anlagen degrades, mineralizes, and is removed by osteoclast-like cells. This process spreads up and down the bones and allows for vascular invasion and influx of osteoprogenitor cells. The periosteum in the midshaft region of the bone produces osteoblasts, which synthesize the cortex; this is known as the primary ossification center.

At the ends of the cartilage anlagen, a similar process leading to the removal of cartilage occurs (secondary center of ossification), leaving a portion of cartilage model “trapped” between the expanding primary and secondary ossification centers. This area is referred to as a cartilage growth plate or epiphysis. Four chondrocyte cell types exist in the growth plate: reserve, resting, proliferative, and hypertrophic. These growth plate chondrocytes undergo a tightly regulated program of proliferation, hypertrophy, degradation, and replacement by bone (primary spongiosa). This is the major mechanism of skeletogenesis and is the
mechanism by which bones increase in length, and the articular surfaces increase in diameter. In contrast, the flat bones of the cranial vault and part of the clavicles and pubis form by intramembranous ossification, whereby fibrous tissue, derived from mesenchymal cells, differentiates directly into osteoblasts, which directly lay down bone. These processes are under specific and direct genetic control, and abnormalities in the genes that encode these pathways frequently lead to skeletal dysplasias.

**Cartilage Structure**

Collagen accounts for two-thirds of the adult weight of adult articular cartilage and provides significant strength and structure to the tissue (see Chapter 3). Collagens are a family of proteins that consist of single molecules (monomers) that combine into three polypeptide chains to form a triple helix structure. In the triple helix, every third amino acid is a glycine residue and the general chain structure is denoted as Gly-X-Y, where X and Y are commonly proline and hydroxyproline. The collagen helix can be composed of identical chains (homotrimeric), as in type II collagen, or can consist of different collagen chains (heterotrimeric), as seen in collagen type XI.

Collagens are widely distributed throughout the body, and 33 collagen gene products are expressed in a tissue-specific manner, leading to 19 triple helical collagens. Collagens are classified further by the structures they form in the extracellular matrix. The most abundant collagens are the fibril-forming types (I, II, III, V, and XI), and their extensive cross-linking provides mechanical strength that is necessary for high stress tissue such as cartilage, bone, and skin. Another collagen species is the fibril-associated collagens with interrupted triple helices, which include collagen types IX, XII, XIV, and XVI. These collagens interact with fibril collagen and other extracellular molecules including aggrecan, cartilage oligomeric matrix protein, and other sulfated proteoglycans. Collagen types VIII and X are non-fibrillar, short-chain collagens, and type X collagen is the most abundant extracellular matrix molecule expressed by hypertrophic chondrocytes during endochondral ossification. The major collagens of articular cartilage are fibrillar collagen types II, IX, X, and XI. In developing cartilage, the core fibrillar network is a cross-linked copolymer of collagens II, IX, and XI. Mutations in genes that encode these collagens and proteins involved in their processing result in various skeletal dysplasias and highlight the importance of these molecules in skeletal development.

**Classification and Nomenclature**

As mentioned earlier, in the 1970s, there was recognition of the genetic and clinical heterogeneity of heritable disorders of connective tissue and a new awareness of the complexity of these disorders. As a result, there have been multiple attempts to classify these disorders in a manner that clinicians and scientists could use effectively to diagnose and determine their pathogenicity. The initial categories were pure descriptive and clinically based. With the more recent explosion in determining the genetic basis of these diseases, the classification has evolved into one that combines the older clinical one (including the eponyms and Greek terms) and blends these disorders into families that share a molecular basis or pathway. The most recent updated classification can be found at www.isds.ch. Some of the chondrodysplasia families are listed in Table 105-1.

The most widely used method for differentiating the skeletal disorders has been through the detection of skeletal radiographic abnormalities. Radiographic classifications are based on the different parts of the long bones that are abnormal (epiphyses, metaphyses, and diaphyses) (Figure 105-1). These epiphyseal, metaphyseal, and diaphyseal disorders can be differentiated further depending on whether or not the spine is involved (spondyloepiphyseal, spondyloepimetaphyseal, and spondyloepimetaphyseal dysplasias). The classes of these disorders can be differentiated further into distinct disorders on the basis of other clinical and radiographic findings.

**Clinical Evaluation and Features**

The skeletal dysplasias are generalized disorders of the skeleton and usually result in disproportionate short stature. Affected individuals usually present because they are disproportionately short. This finding needs to be documented on the appropriate growth curves for gender and ethnicity if possible. As a generalization, most individuals with disproportionate short stature have skeletal dysplasias, and individuals with proportionate short stature have endocrine, nutritional, or prenatal-onset growth deficiency or other disorders. Exceptions to the rule include congenital hypothyroidism, which is associated with disproportionate short stature, and disorders such as osteogenesis imperfecta (OI) and hypophosphatasia can be associated with normal body proportions.

A disproportionate body habitus may not be immediately visible on physical examination. Anthropometric dimensions such as upper-to-lower segment (U/L) ratio, sitting height, and arm span must be measured when considering the possibility of a skeletal dysplasia and should be measured in centimeters. Sitting height is an accurate measurement...
Table 105-1 Classification of the Chondrodysplasias

<table>
<thead>
<tr>
<th>Dysplasia</th>
<th>Inheritance</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Achondroplasia Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achondroplasia</td>
<td>AD</td>
<td>FGFR3</td>
</tr>
<tr>
<td>Thanatophoric dysplasia, type I</td>
<td>AD</td>
<td>FGFR3</td>
</tr>
<tr>
<td>Thanatophoric dysplasia, type II</td>
<td>AD</td>
<td>FGFR3</td>
</tr>
<tr>
<td>Achondroplasia</td>
<td>AD</td>
<td>FGFR3</td>
</tr>
<tr>
<td>Hypochondroplasia</td>
<td>AD</td>
<td>FGFR3</td>
</tr>
<tr>
<td>SADDAN</td>
<td>AD</td>
<td>FGFR3</td>
</tr>
<tr>
<td>CATSHL</td>
<td>AD</td>
<td>FGFR3</td>
</tr>
<tr>
<td>Osteoeplogephon dysplasia</td>
<td>AD</td>
<td>FGFR1</td>
</tr>
<tr>
<td><strong>Severe Spondyloeplastic Dysplasias</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achondrogenesis IA</td>
<td>AR</td>
<td>GMAP210</td>
</tr>
<tr>
<td>Oispsmodyplasia</td>
<td>AR</td>
<td>Unknown</td>
</tr>
<tr>
<td>Schneckenbenck dysplasia</td>
<td>AR</td>
<td>SLC3SD1</td>
</tr>
<tr>
<td>Spondyloepiheysal dysplasia, type Sedaghatian</td>
<td>AD</td>
<td>TRPV4</td>
</tr>
<tr>
<td><strong>TRPV4/Metatropic Dysplasias Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metatropic dysplasia</td>
<td>AD</td>
<td>TRPV4</td>
</tr>
<tr>
<td>Spondyloepiphsyal dysplasia (Kozlowski type)</td>
<td>AD</td>
<td>TRPV4</td>
</tr>
<tr>
<td>Parastrammetic dysplasia</td>
<td>AD</td>
<td>TRPV4</td>
</tr>
<tr>
<td>Brachyolmia (AD)</td>
<td>AD</td>
<td>TRPV4</td>
</tr>
<tr>
<td><strong>Short Rib Dysplasia (Polydactyly) Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-rib polydactly type I/II</td>
<td>AR</td>
<td>DYNC2CH1</td>
</tr>
<tr>
<td>IFT80</td>
<td></td>
<td>WDR35</td>
</tr>
<tr>
<td>NDK1</td>
<td></td>
<td>DYNC2CH1</td>
</tr>
<tr>
<td>IFT80</td>
<td></td>
<td>EVC1, EVC2</td>
</tr>
<tr>
<td>Thoracolaryngopelvic dysplasia</td>
<td>AD</td>
<td>EVC1, EVC2</td>
</tr>
<tr>
<td><strong>Filamin-Related Disorders</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atelosteogenesis I</td>
<td>AD</td>
<td>FLNB</td>
</tr>
<tr>
<td>Atelosteogenesis III</td>
<td>AD</td>
<td>FLNB</td>
</tr>
<tr>
<td>Larsen syndrome</td>
<td>XLR</td>
<td>FLNA</td>
</tr>
<tr>
<td>Otopalato-digital syndrome type II</td>
<td>XLR</td>
<td>FLNA</td>
</tr>
<tr>
<td>Osteoeplyplasia, Melnick-Needles</td>
<td>XLD</td>
<td>FLNA</td>
</tr>
<tr>
<td><strong>Diastrophic Dysplasia Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achondrogenesis IB</td>
<td>AR</td>
<td>DTDST</td>
</tr>
<tr>
<td>Achondrogenesis II</td>
<td>AR</td>
<td>DTDST</td>
</tr>
<tr>
<td>Diastrophic dysplasia</td>
<td>AR</td>
<td>DTDST</td>
</tr>
<tr>
<td>Recessive multiple epiphyseal dysplasia</td>
<td>AR</td>
<td>DTDST</td>
</tr>
<tr>
<td><strong>Dyssegmental Dysplasia Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyssegmental dysplasia</td>
<td>AR</td>
<td>HSPG2</td>
</tr>
<tr>
<td>Silverman-Handmaker type</td>
<td>AR</td>
<td>HSPG2</td>
</tr>
<tr>
<td>Dyssegmental dysplasia</td>
<td>AR</td>
<td>HSPG2</td>
</tr>
<tr>
<td>Rolland-Derbuquis</td>
<td>AR</td>
<td>HSPG2</td>
</tr>
<tr>
<td>Schwartz-Jampel syndrome</td>
<td>AR</td>
<td>HSPG2</td>
</tr>
<tr>
<td><strong>Type II Collagenopathies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achondrogenesis II</td>
<td>AD</td>
<td>COL2A1</td>
</tr>
<tr>
<td>Kniest dysplasia</td>
<td>AD</td>
<td>COL2A1</td>
</tr>
<tr>
<td>Spondyloepiphyseal dysplasia conginita</td>
<td>AD</td>
<td>COL2A1</td>
</tr>
<tr>
<td>Spondyloepiphyseal dysplasia</td>
<td>AD</td>
<td>COL2A1</td>
</tr>
<tr>
<td>Strudwick type</td>
<td>AD</td>
<td>COL2A1</td>
</tr>
<tr>
<td>Spondeleophipheral dysplasia</td>
<td>AD</td>
<td>COL2A1</td>
</tr>
<tr>
<td>Spondeleophipheral dysplasia</td>
<td>AD</td>
<td>COL2A1</td>
</tr>
<tr>
<td>Arthro-opthamalopathy (Stickler syndrome)</td>
<td>AD</td>
<td>COL2A1</td>
</tr>
<tr>
<td><strong>Type XI Collagenopathies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stickler dysplasia</td>
<td>AD</td>
<td>COL11A1</td>
</tr>
<tr>
<td>OSMED</td>
<td>AR</td>
<td>COL11A2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dysplasia</th>
<th>Inheritance</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weissenbacher-Zweymuller syndrome</td>
<td>AD</td>
<td>COL11A2</td>
</tr>
<tr>
<td>Fibrochondrogeneis</td>
<td>AR</td>
<td>COL11A1</td>
</tr>
<tr>
<td>Other Spondyloepi-(meta)-physeal [SE(M)D] Dysplasia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spondyloepimetaphyseal dysplasia, Pakistani type</td>
<td>AR</td>
<td>ATPSK2</td>
</tr>
<tr>
<td>Spondyloepiphysyal dysplasia tara</td>
<td>XLR</td>
<td>SBDS</td>
</tr>
<tr>
<td>Progressive pseudorheumatoid dysplasia</td>
<td>AR</td>
<td>WISP3</td>
</tr>
<tr>
<td>Dygyge-Melchior-Clausen dysplasia</td>
<td>AR</td>
<td>FLJ90130</td>
</tr>
<tr>
<td>Wolctol-Ralison dysplasia</td>
<td>AR</td>
<td>EIF2AK3</td>
</tr>
<tr>
<td>Acrocapitofemoral dysplasia</td>
<td>AR</td>
<td>IHH</td>
</tr>
<tr>
<td>Schimke immuno-osseous dysplasia</td>
<td>AR</td>
<td>SMARCAL1</td>
</tr>
<tr>
<td>Spondytrarime</td>
<td>AR</td>
<td>Unknown</td>
</tr>
<tr>
<td>Spondyloepiheysyal dysplasia, type corner fracture</td>
<td>AD</td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>Multiple Epiphysyal Dysplasia and Pseudoachondroplasia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple epiphysyal dysplasia</td>
<td>AD</td>
<td>COL9A1</td>
</tr>
<tr>
<td>COL9A2</td>
<td></td>
<td>COL9A3</td>
</tr>
<tr>
<td>MATN</td>
<td></td>
<td>COMP</td>
</tr>
<tr>
<td>Pseudoachondroplasia</td>
<td>AD</td>
<td>COMP</td>
</tr>
<tr>
<td><strong>Chondroplasia Punctata</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chondroplasia punctata, rhizomeric type</td>
<td>XLD</td>
<td>PEX7</td>
</tr>
<tr>
<td>DHAPAT</td>
<td></td>
<td>AGPS</td>
</tr>
<tr>
<td>EBP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chondroplasia punctata, Conradi-Hunermann type</td>
<td>XLR</td>
<td>LBR</td>
</tr>
<tr>
<td>Hydrops-ectopic calcifications-moth-eaten bones</td>
<td>XLR</td>
<td>ARSE</td>
</tr>
<tr>
<td>Chondroplasia punctata, brachytelepahalic type</td>
<td>XLR</td>
<td>ARSE</td>
</tr>
<tr>
<td>Chondroplasia punctata, tibial-metacarpal type</td>
<td>XLD</td>
<td>ARSE</td>
</tr>
<tr>
<td><strong>Metaphysyal Dysplasias</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metaphysyal chondroplasia, type Jansen</td>
<td>AD</td>
<td>PTHrP</td>
</tr>
<tr>
<td>Eiken dysplasia</td>
<td>XLD</td>
<td>PTHrP</td>
</tr>
<tr>
<td>Bloomstrand dysplasia</td>
<td>AD</td>
<td>COL10A1</td>
</tr>
<tr>
<td>Metaphysyal chondroplasia, type Schmidt</td>
<td>AR</td>
<td>RMRP</td>
</tr>
<tr>
<td>Metaphysyal chondroplasia, McKusick type</td>
<td>AR</td>
<td>SBDS</td>
</tr>
<tr>
<td>Metaphysyal chondroplasia, with pancreatic insufficiency, and cyclin neutropenia</td>
<td>AR</td>
<td>ADA</td>
</tr>
<tr>
<td><strong>Brachyolmia Spondylodysplasias</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brachyolmia (Hobek type)</td>
<td>AR</td>
<td>Unknown</td>
</tr>
<tr>
<td>(Toledo type)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brachyolmia (Maroteaux type)</td>
<td>AR</td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>Rhizo- and Mesomelic Dysplasias</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omodyplasia</td>
<td>AR</td>
<td>Glypican 6</td>
</tr>
<tr>
<td>Dyschondrosteosis</td>
<td>XLD</td>
<td>SHOX</td>
</tr>
<tr>
<td>Mesomelic dysplasia, type Lange</td>
<td>XLR</td>
<td>SHOX</td>
</tr>
<tr>
<td>Mesomelic dysplasia, type Robinow</td>
<td>AD</td>
<td>ROR2</td>
</tr>
<tr>
<td>Mesomelic dysplasia, Kantapura type</td>
<td>AD</td>
<td>Duplicated in the Hox cluster</td>
</tr>
</tbody>
</table>

Continued
of head and trunk length, but it requires special equipment for precise measurements. U/L ratios are easy to obtain and provide an accurate measurement of proportion. The lower segment is measured from the symphysis pubis to the floor at the inside of the heel. The upper segment is measured by subtracting the lower segment measurement from the total height. McKusick18−19 has published standard U/L segment ratios for whites and African-Americans across ages. An average-height white child 8 to 10 years old has a U/L segment ratio of approximately 1 and as an adult has a U/L segment ratio of 0.95. Individuals presenting with disproportionate short stature have altered U/L segment ratios depending on whether they have short limbs, short trunk, or both. An individual with short limbs and normal trunk has an increased U/L segment ratio, and an individual with normal limbs but short trunk has a diminished U/L segment ratio (Figure 105-2). Another means of determining if there is disproportion is based on arm span measurements, which are close to total height in an average-proportioned individual. A short-limbed individual has an arm span considerably shorter than the height.

As in any disorder that has a genetic basis, it is crucial to obtain an accurate family history. This should include any history of previously affected children or parental consanguinity. The skeletal dysplasias are genetically heterogeneous and can be inherited as autosomal dominant, autosomal recessive, X-linked recessive, and X-linked dominant disorders, and rarer genetic mechanisms of disease including germline mosaicism, uniparental disomy, and chromosomal rearrangements have been seen.15−18 For many patients and families, accurate diagnosis and recurrence risk can have a significant impact on their reproductive decisions. Another consideration for patients with short stature is that there is increased nonrandom mating, which leads to reproductive outcomes that have been previously unknown.19,20 Homozygous achondroplasia is lethal, and many newborns who inherit two dominant mutations (compound heterozygotes) die early with severe abnormalities of the skeleton.21 It is also important to obtain an accurate history relative to the onset of short stature and whether it developed immediately in the postnatal period or was noticed at age 2 or 3. Of the 450 skeletal dysplasias, approximately 100 of them have onset in the prenatal period, but many affected individuals do not develop disproportionate short stature and joint discomfort until childhood.22,23

A detailed physical examination may reveal a diagnosis or help differentiate the most likely group of possible diagnoses. It is crucial when disproportion and short stature have been established and the limbs are involved to determine which segment is involved: upper segment (rhizomelic—humerus and femur); middle segment (mesomelic—radius, ulna, tibia, and fibula); and distal segment (acromelic—hands and feet). Numerous head and facial dysmorphisms are seen in the skeletal disorders. Affected individuals frequently have disproportionately large heads. Frontal bossing and flattened nasal bridge are characteristic of achondroplasia, one of the most common skeletal dysplasias.24 Cleft palate and micrognathia are commonly found in the type II collagen abnormalities, abnormally flattened midface with

### Table 105-1 Classification of the Chondrodysplasias—cont’d

<table>
<thead>
<tr>
<th>Dysplasia</th>
<th>Inheritance</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acromelic and Acromesomelic Dysplasias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acromelic dysplasia</td>
<td>AD</td>
<td>Fibrillin 1</td>
</tr>
<tr>
<td>Geleophysic dysplasia</td>
<td>AD</td>
<td>Fibrillin 1</td>
</tr>
<tr>
<td>Trichorhinophalangeal dysplasia, type I</td>
<td>AD</td>
<td>TRPS1</td>
</tr>
<tr>
<td>Trichorhinophalangeal dysplasia, type II</td>
<td>AD</td>
<td>TRPS2</td>
</tr>
<tr>
<td>Acrodysostosis</td>
<td>AD</td>
<td>PRKAR1A</td>
</tr>
<tr>
<td>Grebe dysplasia</td>
<td>AR</td>
<td>CDMP1</td>
</tr>
<tr>
<td>Acromesomelic dysplasia, Hunter-Thompson</td>
<td>AR</td>
<td>CDMP1</td>
</tr>
<tr>
<td>Acromesomelic dysplasia, type Maroteaux</td>
<td>AR</td>
<td>NPB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dysplasia with Prominent Membranous Bone Involvement</th>
<th>Inheritance</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campomelic dysplasia</td>
<td>AD</td>
<td>SOX9</td>
</tr>
<tr>
<td>Stüve-Wiedemann dysplasia</td>
<td>AR</td>
<td>LIFR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bent Bone Dysplasias</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Desbuquois syndrome</td>
<td>AR</td>
<td>CANT1</td>
</tr>
<tr>
<td>Pseudodiastrophic dysplasia</td>
<td>AR</td>
<td>Unknown</td>
</tr>
<tr>
<td>Spondyloepimetaphyseal dysplasia with joint laxity</td>
<td>AR</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multiple Dislocations with Dysplasias</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Maroteaux-Lamy syndrome</td>
<td>AR</td>
<td>CNT1</td>
</tr>
<tr>
<td>Pseudopseudohypoparathyroidism</td>
<td>AR</td>
<td>TRPV4</td>
</tr>
<tr>
<td>Hypochondroplasia</td>
<td>AR</td>
<td>TRPV4</td>
</tr>
<tr>
<td>Hunter-Thompson syndrome</td>
<td>AR</td>
<td>TRPV4</td>
</tr>
</tbody>
</table>

**Figure 105-2** Upper segment length/lower segment length (U/L) in 8- to 10-year-old individuals with short limb and short trunk dwarfism. The child on the left has short limbs and an increased U/L ratio; the child on the right has a short trunk and reduced U/L ratio.
a turned-up nose is frequently found in the chondrodysplasia punctata disorders, and abnormal swollen pinnae are seen in diastrophic dysplasia. Individuals with skeletal dysplasias should be screened for ophthalmologic and hearing abnormalities because many of these disorders are associated with eye abnormalities and hearing loss.

Further evaluation of the hands and feet can lead to further differentiation of these disorders. Postaxial polydactyly is characteristically found in chondroectodermal dysplasia and the short-rib polydactyly disorders (see Table 105-1). Short, hypermobile, radially displaced thumbs are seen in diastrophic dysplasia. Nails can be abnormally hypoplastic in chondroectodermal dysplasia and short and broad in cartilage hair hypoplasia. Clubfeet may be seen in many disorders including Kniest dysplasia, spondyloepiphyseal dysplasia congenita, Larsen syndrome, varying forms of osteogenesis imperfecta, and diastrophic dysplasia. Bone fractures occur most commonly in two types of disorders—those that result from undermineralized bone (OI, hypophosphatasia, achondrogenesis IA), or those that result from overmineralized bone (osteopetrosis syndromes and dysosteosclerosis).

Organ systems other than the skeleton can be involved, although rarely. Congenital cardiac defects are seen in chondroectodermal dysplasia (atrial septal defects), the short-rib polydactyly disorders (complex outlet defects including isolated ventricular septal defects), and Larsen syndrome (ventricular septal defects). Gastrointestinal anomalies are rare among the skeletal disorders, but congenital megacolon can be seen in cartilage hair hypoplasia, malabsorption syndrome in Schwachmann-Diamond syndrome, and omphaloceles in otopalatodigital syndrome and atelosteogenesis I.

**Diagnosis and Testing**

After obtaining a thorough family history and physical examination, the next step is to obtain a full set of skeletal radiographs. A full series of skeletal views includes anterior, lateral, and Towne views of the skull; anterior and lateral views of the entire spine; and anteroposterior views of the pelvis and extremities, with separate views of the hands and feet, especially after the newborn period. Most of the important clues to diagnosis are in skeletal radiographs that are obtained before puberty. When the epiphyses have fused to the metaphyses, determining the precise diagnosis can be extremely challenging. If an adult is evaluated, all attempts should be made to obtain any available childhood radiographs. Many subtle clues in these skeletal radiographs can lead to precise diagnosis. Demonstrating punctate calcifications in the areas of the epiphyses in the chondrodysplasia punctata disorders, multiple ossification centers of the calcaneus in more than 20 disorders, and the type of hand shortening can aid in differentiating many disorders.

After obtaining radiographs, close attention should be paid to the specific parts of the skeleton (spine, limbs, pelvis, skull) involved and to the location of the lesions (epiphyses, metaphyses, and vertebrae) (Figure 105-3). As mentioned earlier, these radiographic abnormalities can change with age, and if available, radiographs across a few years or decades aid in diagnosis. Fractures can be seen in OI (all types) (Figure 105-4; see Table 105-1) and severe hypophosphatasia. In older individuals, fractures may be seen in disorders associated with increased mineralization such as the osteopetrosis syndromes and dysosteosclerosis. When a thorough evaluation of the radiographs reveals abnormalities, but a diagnosis still cannot be made, resources are available. The International Skeletal Dysplasia Registry and European Skeletal Dysplasia Network are available to provide diagnosis for these rare disorders.

Morphologic studies of chondro-osseous tissue have revealed specific abnormalities in many of the skeletal dysplasias. In these disorders, histologic evaluation of chondro-osseous morphology can aid in making an accurate diagnosis, and absence of histopathologic alterations can rule out diagnoses. These studies need to be done on cartilage growth plate, and although commonly performed on perinatal lethal skeletal disorders at autopsy, obtaining...
growth plate histology on individuals with nonlethal disorders is difficult. If affected individuals (children) are undergoing surgery, an iliac crest biopsy specimen can be evaluated.

Histomorphology studies done on these disorders have led to important insights on the pathogenesis of these disorders. On morphologic grounds, the chondrodysplasias can be broadly classified into disorders (1) that have a qualitative abnormality in endochondral ossification, (2) that have abnormalities in cellular morphology, (3) that have abnormalities in matrix morphology, and (4) in which the abnormality is primarily localized to the area of chondro-osseous transformation. In thanatophoric dysplasia, there is a defect in endochondral ossification with a short, almost hypertrophic zone; shortened proliferative zone; and overgrowth of the periosteum. In pseudoachondroplasia, there is a distinct lamellar pattern (alternating electron-dense and electron-lucent lamellae) in the rough endoplasmic reticulum of chondrocytes (Figure 105-5) and a grossly abnormal matrix in diastrophic dysplasia, which leads to a characteristic ring around the chondrocytes. All of these findings are characteristic and diagnostic for these disorders and illustrate how morphometry studies can have an integral part in the investigation of these disorders.

There has been significant progress in gene identification in these disorders, which has impact for affected individuals. As illustrated in Table 105-1, for disorders in which the gene is identified, molecular diagnostic testing is potentially available. Molecular diagnosis can be used to confirm a clinical and radiographic diagnosis, predict carrier status in families at risk for a recessive disorder, and, for some individuals, allow for prenatal diagnosis of at-risk fetuses. Because these are rare disorders, commercial testing is not always readily available; however, GeneTests (www.genetests.org) is a publically funded medical genetics website developed for physicians that provides information on diseases and available genetic testing.

**Management and Treatment**

The optimal management of this diverse set of disorders requires an understanding of the medical, skeletal, and psychosocial consequences. This is often best accomplished by centers that have a multidisciplinary approach, which includes adult and pediatric physicians, orthopedists, rheumatologists, otolaryngologists, neurologists, neurosurgeons, and ophthalmologists who are committed to the care of these patients.

Most medical complications in these disorders result from orthopedic complications, and they vary depending on the specific disorder. In disorders associated with significant odontoid hypoplasia such as Morquio disease, type II collagenopathies, metatropic dysplasia, and Larsen syndrome, flexion-extension films should be monitored at regular intervals to assess for C1-C2 subluxation. Many experts in the field now believe that all individuals with skeletal dysplasias should have evaluation of their cervical spine, regardless of diagnosis. If there is evidence for subluxation, surgery for C1-C2 fixation is indicated. Genu varum—lateral curvature of the lower extremity—is common in many skeletal disorders caused by overgrowth of the fibula; this causes knee or ankle pain in many individuals, especially children, and correction by osteotomy should be
considered. Children and adults with skeletal dysplasias should have regular eye and hearing examinations because they are at increased risk for myopia, retinal degeneration, glaucoma, and hearing loss depending on the disorder.

Frequently, patients with these disorders have significant joint pain and in some cases joint limitations. Because most of these disorders result from mutations in genes crucial to cartilage function, the cartilage at the joint surfaces may not provide adequate support and cushioning function. Many of these patients seek attention for joint pain. Evaluation should include radiographs and magnetic resonance imaging (MRI), when appropriate, to determine the etiology of the pain. In some disorders such as the type II collagenopathies, pseudoachondroplasia, multiple epiphyseal dysplasia, and cartilage hair hypoplasia, by adulthood, so little cartilage remains at the knee or hips that joint replacement is indicated for pain relief. Lastly, overweight in adults with short stature is an ongoing issue and contributes to inactivity, loss of function, adult-onset diabetes, hypertension, and coronary disease.

**Achondroplasia**

Achondroplasia is the most common of the nonlethal skeletal dysplasias (approximately 1 in 20,000) and serves as an example on how to approach these disorders. Most affected individuals are of normal intelligence, have a normal life span, and lead independent and productive lives. The mean final height in achondroplasia is 130 cm for men and 125 cm for women; specific growth charts have been developed to document and track linear growth, head circumference, and weight in these individuals.

In early infancy, there is potentially serious compression of the cervicomedullary spinal cord secondary to a narrow foramen magnum, cervical canal, or both. Clinically, these infants have central apnea, sleep apnea, profound hypotonia, motor delay, emesis while forward positioned in car seats, or excessive sweating. MRI with flow studies is necessary to document the obstruction; if present, obstruction requires decompressive surgery. Other complications include nasal obstruction, venous distention, thoracolumbar kyphosis, and hydrocephalus in a few individuals.

From early childhood, and as children begin to walk around 22 to 24 months, they develop several orthopedic manifestations, which include progressive bowing of the legs owing to fibular overgrowth, lumbar lordosis, and hip flexion contractures. Recurrent ear infections can lead to chronic serous otitis media and deafness. Tympanic membrane tube placement is indicated in many of these patients. Craniofacial abnormalities lead to dental malocclusion, and appropriate treatment is necessary. In adults, the main potential medical complication is impingement of the spinal root canals. This complication can be manifested by lower limb paresthesias, claudication, clonus, or bladder or bowel dysfunction. It is crucial that these complaints are addressed because without appropriate decompression surgery, spinal cord paralysis may result.

Growth hormone has not been effective in increasing height in this disorder. Surgical limb lengthening has been employed successfully to increase limb length by 12 inches, but this technique needs to be done during the teen years and is performed over a 2-year period and is associated with complications. Recent advances in the molecular understanding underlying achondroplasia have identified molecular targets to potentially treat the disorder, thus improving height and orthopedic complications. Achondroplasia results from heterozygosity for mutations in the gene that encodes fibroblast growth factor receptor 3 (FGFR3). The mutation causes constitutive activation of the receptor leading to increased MAPK signaling with elevated levels of ERK1/2 phosphorylation. Molecules targeted to the tyrosine kinase domain of the receptor and those that diminish ERK signaling have shown efficacy in tissue and animal models. Throughout their lives, individuals with achondroplasia and other skeletal dysplasias and their families experience various psychosocial challenges. These challenges can be addressed by specialized medical and social support systems. Interactions with advocacy groups such as Little People of America (www.lpaonline.org) can provide emotional support and medical information.

**Biochemical and Molecular Abnormalities**

Similarities in clinical and radiographic findings and histomorphology have placed bone dysplasia into families. These families share common pathophysiological or pathway mechanisms. In recent years, there has been an explosion in understanding of the basic biology of these disorders. This explosion has resulted from the successful human genome project, which improved various methodologies including candidate gene approach, linkage analysis, positonal cloning, human/mouse synteny, array comparative genomic hybridization, and massive parallel sequencing (whole exome or whole genome analyses) allowing for identification of the disease genes (see Table 105-1). With gene discovery in the vast number of these osteochondrodysplasias, these genes can be placed into several categories designed to understand their pathogenesis: (1) defects in extracellular proteins; (2) defects in metabolic pathways (enzymes, ion channels, and transporters; (3) defects in folding and degradation of macromolecules; (4) defects in hormones and signal transduction; (5) defects in nuclear proteins; (6) defects in oncogenes and tumor-suppressor genes; (7) defects in RNA and deoxyribonucleic acid (DNA) processing molecules; (8) defects in intracellular structural and organelle proteins; (9) microRNAs; and (10) genes of unknown function. There are still skeletal dysplasias for which the gene and mechanism of disease are unknown. Following are descriptions of some of the molecular mechanisms involved in the skeletal dysplasias.

**DEFECTS IN EXTRACELLULAR STRUCTURAL PROTEINS**

**Type II Collagen and Type XI Collagen**

Because type II collagen was found primarily in cartilage, the nucleus pulposus, and the vitreous of the eye, it was hypothesized that skeletal disorders with significant spine and eye abnormalities would result from defects in type II collagen. Type II collagen defects have been identified in a spectrum of disorders ranging from lethal to mild arthropathy, which include achondrogenesis II, hypochondrogenesis, spondyloepiphyseal dysplasia congenita,
spondyloepimetaphyseal dysplasia, Strudwick type, Kniest dysplasia, Stickler syndrome, spondyloperipheral dysplasia, and "precocious" familial arthropathy. These disorders are referred to as type II collagenopathies, and they all result from heterozygosity for mutations in COL2A1. Biochemical analysis of cartilage derived from these individuals shows electrophoretically detectable abnormal type II collagen. Type I collagen is not normally present in cartilage, but in the presence of abnormal type II collagen there is increased type I collagen in the growth plate.

Mutations that result in a substitution for a triple helical glycine residue seem to be the most common type of mutation. There are some correlations between the location of the mutation and the disease phenotype. In spondyloepimetaphyseal dysplasia, the glycine substitutions are scattered throughout the molecule; however, in Kniest dysplasia, the mutations are in the more amino-terminal end of the molecule. Stickler syndrome, a disorder of mild short stature, arthropathy, and high-grade myopia (see Table 105-1), is genetically heterogeneous and results from mutations in COL2A1 and COL11A1, and nonocular forms result from mutations in COL11A2. In Stickler syndrome, the COL2A1 and COL11A1 mutations tend to be nonsense mutations resulting in premature translation stop codons; however, patients with COL11A1 mutations tend to have a more severe eye phenotype and hearing loss than patients with COL2A1 mutations.

Individuals heterozygous for various COL11A2 mutations have a nonocular form of Stickler syndrome, consistent with the absent expression of COL11A1 in the vitreous humor. Oto- spondylo-megaepiphyseal dysplasia is a rare autosomal recessive disorder caused by loss of function mutations in COL11A2. This disorder has radiographic similarities to Kniest dysplasia but is associated with profound sensorineural hearing loss and lack of ocular involvement. Recent discoveries have extended the spectrum of disease for type XI collagen. Autosomal recessive fibrochondrogenesis, a severe skeletal dysplasia, highly associated with lethality, results from mutations in the two genes that encode type XI, COL11A1 and COL11A2. Type II and XI collagens form a heterotypic fibril in the cartilage matrix and not surprisingly, there is significant clinical overlap in the disorders due to mutation in the genes that encode these collagens.

**Cartilage Oligomeric Matrix Protein**

Heterozygosity for mutations in cartilage oligomeric matrix protein leads to pseudoachondroplasia and multiple epiphyseal dysplasia. Cartilage oligomeric matrix protein is a member of the thrombospondin family of proteins and consists of an epidermal growth factor domain and calcium binding, calmodulin domain. In pseudoachondroplasia and multiple epiphyseal dysplasia, disease-producing mutations occur in the calmodulin domain, with a few in the globular carboxy-terminal domain (Figure 105-6). As opposed to pseudoachondroplasia, multiple epiphyseal dysplasia results from heterozygosity for mutations in numerous genes (COL9A1, COL9A2, COL9A3, and MATRILIN3), and there is a recessive form due to mutations in the DTDST gene. Both these disorders are associated with significant early destruction of cartilage with many affected individuals undergoing hip and knee replacements at an early age.

**Defects in Metabolic Pathways**

Defects in metabolic pathways comprise defects in enzymes, ion channels, and transporters essential for cartilage metabolism and homeostasis. An example is the diastrophic dysplasia group (see Table 105-1), a spectrum of disorders (lethal to mild short stature) resulting from mutations in the DTDST (SLC26A2) gene. These disorders result from a varying defect in the degree of sulfate uptake or transport into chondrocytes. Lack of adequate intracellular sulfate affects the normal post-translational modification of proteoglycans and leads to abnormal chondrogenesis that is proportional to the degree of transporter compromise. Affected individuals suffer from severe degenerative joint disease.

![Diagram of the cartilage oligomeric matrix protein delineating the domains—NH2, amino terminus, epidermal growth factor-like (EGF-like), calmodulin-like, carboxy-terminus (COOH), pseudoachondroplasia (PSACH), and multiple epiphyseal dysplasia (MED). Amino acid substitutions are listed below the molecule.](image-url)
Defects in Intracellular Structural Proteins

Intracellular proteins are ubiquitously expressed proteins; the finding that mutations in the genes encoding filamin A and filamin B produced primarily skeletal disorders was surprising.56-58 The filamins are cytoskeleton proteins involved in multicellular processes including providing structure to the cell, facilitating signal transduction and transport of small solutes, allowing communication between the intracellular and extracellular environment, and participating in cell division and motility. Defects in these genes have a profound effect on the skeleton ranging from absence of bone formation to significant joint dislocations. The mechanisms by which these mutations produce disease are unclear, though alterations in the cellular and organelle functions are beginning to be elucidated.59

Defects in Membrane Channels

Calcium homeostasis is critical for cartilage and bone.59,60 TRPV4, or transient receptor potential cation channel subfamily V member 4, is a cation channel that mediates calcium influx in response to numerous stimuli. The importance of this channel has been demonstrated because it produces a vast spectrum of autosomal dominant skeletal disorders including lethal metatropic dysplasia, non-lethal metatropic dysplasia, spondyloepiphyseal dysplasia, Koizlowski type, and brachyolmia.61-63 In addition, heterozygosity for mutations in TRPV4 also causes neuromuscular diseases without notable bony manifestation that include hereditary motor and sensory neuropathy type IIC, congenital spinal muscular atrophy, and scapuloperoneal spinal muscular atrophy.64-66 The mechanism by which these mutations scattered throughout the molecule produce such divergent phenotypes is unclear but supports some common pathway in tissues of mesenchymal origin.

Summary

Although these osteochondrodysplasias are rare disorders, affected individuals have significant skeletal complications throughout their lives, first owing to patterning defects, then effects on linear growth, and finally loss of normal structural cartilage as a cushion later in life. The explosion in delineating the molecular defects has shown the complexity of cartilage as a tissue and the large number of cellular processes necessary for a normal skeleton.

OSTEOGENESIS IMPERFECTA

OI is a heritable disorder of bone and was one of the first disorders hypothesized to be a defect in collagen by McKusick.1 Although an osteochondrodysplasia, OI is discussed separately from the chondrodysplasias delineated previously. OI is a generalized disorder of connective tissue that predominantly affects the skeletal system65 and affects numerous individuals (estimates at about 1 in 20,000 individuals).

Initially, there were four types of recognized OI in the clinical classification of Sillence.68 There are now seven well-recognized forms of OI, and through recent gene discoveries it is apparent that a clinical classification system is no longer useful. Because there is enormous clinical variability in these types, the subtypes are discussed separately using historical classifications, but many experts advocate using the terms mild, moderate, and severe (Table 105-2). These disorders all share the same phenotypic finding of hypomineralization of the skeleton.

Mild Osteogenesis Imperfecta (Type I)

Affected individuals with OI type I disease have mild disease in terms of clinical course, the extent of skeletal deformity, and the radiologic appearance of the skeleton (see Figure 105-4A and Table 105-2). They also account for most individuals with OI. Individuals are usually short for their age or their unaffected family members. Many of these individuals experience numerous fractures, especially in childhood; children with OI type I may have 20 fractures by the age of 5.

The disorder is autosomal dominant, and in many cases the individual is the first affected in the family. There is mild facial dysmorphism in OI type I with a mild triangular facial shape. The blue sclerae become gray-to-pale blue in adulthood. Arcus senilis not related to lipid abnormalities may occur in some patients. Other reported ocular defects include scleromalacia, keratoconus, and retinal detachment.69 Teeth frequently show dentinogenesis imperfecta owing to the effects of mutation on the tooth dentin. The deciduous and permanent teeth have an opalescent and translucent appearance, which tends to darken with age. The enamel is normal, but the dentin is dysplastic; chipping of enamel occurs, and the teeth are subjected to erosion and breakage. Teeth of affected individuals appear discolored or gray. This finding varies in the disorder but does co-segregate in families with OI. During the second and third decades of life, a characteristic high-frequency sensorineural or mixed hearing loss can be detected.70 The incidence of mitral valve prolapse is not increased in these patients compared with the population at large, but individual kindreds with increased diameter of the aortic root or patients with aortic regurgitation have been reported.71 Many patients complain of easy bruising, and this may result from the effects of mutation on skin and the vessels below.

Mildly affected patients may not have fractures at birth, although occasionally a fracture of a clavicle or extremity occurs during delivery. Radiographically, affected newborns have wormian bones seen on lateral views of the skull, with significant osteopenia seen through the skeleton, especially the spine. After birth, the frequency of fracture depends on the child’s activity, the need for immobilization after lower extremity fractures, and the attitude of the family toward independent activity. Generally, these patients may experience 5 to 15 major fractures before puberty and several minor traumatic fractures of the digits or the small bones of the feet. Characteristically, the fracture rate declines dramatically after puberty, only to increase during later life. Mild scoliosis approximating 20 degrees is common. Osteopenia is observed in vertebral bodies and the peripheral skeleton and progresses with age. In mild OI, the long bones usually heal with no significant deformity. Compared with more severe phenotypes, children with mild OI only infrequently require the insertion of intramedullary rods and almost never experience nonunion at a fracture site.
Although osteopenia with rarefaction of the medullary space and cortical thinning are observed in radiographs, many mild OI cases can be missed on routine radiographic examination and present later in life as individuals with significant osteoporosis. Measurement of bone mineral density by dual-energy x-ray absorptiometry at any age discloses a significant decrease in bone mass. T scores (i.e., standard deviation from the young-adult mean bone mineral density) are frequently in the range of −2.5 to −4.0 at the lumbar spine or proximal femur, consistent with the diagnosis of osteoporosis as defined by the World Health Organization. Low bone mineral density in children with recurrent fractures may assist in identifying children with OI.

**Molecular Pathology**

As in many other OI phenotypes, OI type I or mild OI is the result of mutations affecting the COL1A1(I) and COL1A2(I) polypeptide chains of type I collagen. Cultured fibroblasts from individuals with mild OI synthesize low amounts (approximately one-half) of the expected amounts of type I collagen. The molecular basis for the low production of type I collagen seems to be diminished activity of one of the COL1A1(I) or COL1A2(I) collagen alleles. Many of the reported mutations in OI type I are nonsense and frameshift mutations and are predicted to lead to premature termination codons, although there are some exceptions.

**Lethal Osteogenesis Imperfecta (Type II)**

Approximately 10% of OI patients have the severe neonatal form of the disease, lethal OI. Most cases result from sporadic mutations; however, more recently, recessive forms of the disease have been documented. These infants present with severe bone fragility, multiple intrauterine fractures at various stages of healing, deformed extremities, and occasionally hydrops fetalis (Figure 105-7). Radiographic features include wormian bones, multiple fractures, crumbled bones, and characteristic beading of the ribs owing to healing callus formation. There is a subtype of the lethal form, OI type IIC, which is autosomal recessive and is differentiated by the absence of beaded ribs (thin ribs) and a different molecular basis of disease.

**Molecular Pathology**

Most cases occur de novo, as new dominant mutations; however, autosomal recessive forms have been established, as has recurrence based on germline mosaicism. The biochemical abnormality in lethal OI is the inability to synthesize, modify, and secrete normal type I collagen. As a result, the amount of type I collagen in bone is low, much of the secreted collagen is abnormally overmodified, and the quantity of the minor collagen types III and V is high. Bone collagen fibers are thinner than normal, and at the intracellular level, type I collagen is retained within dilated endoplasmic reticulum.

Similar to other forms of OI, mutations in the genes encoding COL1A1 and COL1A2 lead to the dominant form or de novo form of lethal OI. Single glycine substitutions with the Gly-X-Y triplet of either COL1A1 or COL1A2 lead to this form of OI, as do some small deletions, all producing severe effects on the triple helix. The recessive form accounts for a few of these cases and results from mutations in the genes encoding either CRTAP (cartilage-associated protein), P3H1 (prolyl-3-hydroxylase 1), and cyclophilin B (PPBI). These molecules form a complex that hydroxylates (add an -OH group) to a third position

---

**Table 105-2 Classification and Molecular Basis of Osteogenesis Imperfecta**

<table>
<thead>
<tr>
<th>OI</th>
<th>Clinical Features</th>
<th>Inheritance</th>
<th>Biochemical Abnormality</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild (type I)</td>
<td>Normal stature, little or no deformity, blue sclerae, hearing loss, dentinogenesis imperfecta</td>
<td>AD (new mutations are common)</td>
<td>50% reduction in type I collagen synthesis</td>
<td>COL1A1, COL1A2</td>
</tr>
<tr>
<td>Lethal (type II)</td>
<td>Lethal; minimal calvarial mineralization, beaded ribs, compressed femurs, long bone deformity</td>
<td>AD (new mutations; gonadal mosaicism)</td>
<td>Structural alterations of type I collagen chains—overmodification of type I collagen</td>
<td>COL1A1, COL1A2, CRTAP, P3H1, PPBI, FKBP10, SERPINH1, SP7</td>
</tr>
<tr>
<td>Severe (types III and IV)</td>
<td>Progressively deforming bones, dentinogenesis imperfecta, hearing loss, short stature</td>
<td>AD</td>
<td>Structural alterations of type I collagen chains—overmodification of type I collagen</td>
<td>COL1A1, COL1A2, CRTAP, P3H1, PPBI, FKBP10, SERPINH1, SP7</td>
</tr>
<tr>
<td>V</td>
<td>Similar to severe OI plus calcification of interosseous membrane of forearm, hyperplastic callus formation</td>
<td>AD</td>
<td>None described</td>
<td>Unknown</td>
</tr>
<tr>
<td>VI</td>
<td>Similar to type IV with vertebral compression; mineralization defect</td>
<td>AR</td>
<td>None described</td>
<td>SERPINF1</td>
</tr>
<tr>
<td>VII</td>
<td>Moderate to severe, with fractures at birth, early deformity and rhizohemia</td>
<td>AR</td>
<td>None described</td>
<td>CRTAP</td>
</tr>
</tbody>
</table>

AD, autosomal dominant; AR, autosomal recessive; CRTAP, cartilage-associated protein; P3H1, prolyl-3-hydroxylase 1; PPBI, cyclophilin B, FKBP10, FK506-binding protein 10; SERPINH1, Serpin Peptidase Inhibitor, Clade H, Member1; SP7, osterix; SERPINF1, Serpin Peptidase Inhibitor, Clade F, Member1.
radiograph of lethal osteogenesis imperfecta (type II) patients with scoliosis greater than 60 degrees develop respiratory compromise and need pulmonary investigations. Many of these individuals need supplemental oxygen.

Platybasia secondary to soft bone at the base of the skull may cause the external ear canals to slant upward as the base of the skull sinks on the cervical vertebrae; this may lead to communicating or obstructive hydrocephalus, cranial nerve palsies, and upper and lower motor neuron lesions. Headache, diplopia, nystagmus, cranial nerve neuralgia, decline in motor function, urinary dysfunction, and respiratory compromise are complications of basilar invagination. As opposed to OI type I, most affected OI type III patients have white sclerae as adults. Approximately 25% of patients with autosomal dominant type III OI have dentinogenesis imperfecta, necessitating constant dental care throughout childhood, though this is not true of the recessive forms of this severe form of OI. Severe hearing impairment occurs in 10% of patients, although milder degrees of hearing loss are more common.

The skeleton in these patients has significant osteopenia, leading to multiple fractures in the upper and lower extremities and vertebral bodies, particularly before puberty. In contrast to OI type I, in which fractures tend to heal without deformity, fractures in OI type III frequently lead to skeletal deformity. Radiographs of the skeleton reveal marked osteopenia, thinning of cortical bone, narrowing of the diaphysis, and widening of the metaphysis, which merges into a dysplastic epiphyseal zone filled with whorls of partially calcified cartilage (i.e., popcorn deformity) (see Figure 105-4C). Osteoporosis leads to collapse of vertebral end plates contributing to worsening kyphoscoliosis. Pectus excavatum or pectus carinatum adds to thoracic deformity. In addition, lack of weight bearing increases the severity of osteoporosis and increases the risk of fracture. Many individuals become wheelchair bound at an early age or walk with mechanical assistance.

Clinically, the phenotype of patients with moderately severe OI (OI type IV) falls between the milder and severe forms of OI. In most cases, this form of OI is inherited in an autosomal dominant fashion. Fractures occur rarely at birth, and some patients may not have an initial fracture until later in the first decade. The extent of skeletal deformity involving the spine, thorax, and extremities is usually intermediate between mild and severe, but these patients have short stature and frequently these patients have scoliosis. Patients may have some mild facial dysmorphisms and hearing loss. Most fractures occur during childhood and may reoccur during the postmenopausal period in women or in men older than age 50 years. Long bone deformity tends to develop after fractures, which may lead to a difficulty in ambulation. Radiographs of the long bones and vertebral bodies show marked osteopenia with vertebral collapse. Although there is marked cortical thinning, bowing, and coarsening of trabeculae, the overall architecture of the bone is normal (see Figure 105-4B).

**Molecular Pathology**

The molecular basis of OI type III and OI type IV is similar to OI type II. Most cases result from heterozygosity for mutations in COL1A1 and COL1A2. These mutations are glycine substitutions scattered throughout the triple helix and in-frame deletions. As in OI type II, familial...
reurrences result from mutations in CRTAP, P3H1, and PPBP, which cause autosomal recessive forms of OI. Recently other genes producing autosomal recessive forms of severe OI have been identified including FKBP10, HSP47, and SP7. There may be subtle clinical distinctions between the recessive forms of the disease, though radiographic abnormalities are quite similar. Clinical suspicion of the recessive form of the disease should be considered if there is a family history of recurrence.

Osteogenesis Imperfecta Type V (Moderate to Severe)

OI type V was reported in 2000 as a variant within the heterogeneous group classified under OI type IV. In the initial report of seven OI patients, the phenotype was distinguished by the following criteria: moderate fracture history, hyperplastic callus formation, limitation in forearm pronation and supination as a result of intramembranous bone formation at the joint, normal sclerae, and no dentinogenesis imperfecta. Bone biopsy specimens showed a meshlike appearance of irregularly spaced lamellae, different from the woven bone seen in OI types II, III, and IV. The etiology of this rare form has not been established.

Osteogenesis Imperfecta Type VI (Moderate to Severe)

The brittle bone phenotype OI type VI was also reported among the heterogeneous OI type IV group of patients. Characteristic among the eight subjects was the occurrence of a first fracture at an early age (4 to 18 months old). The bone is severely brittle, and affected patients have white sclerae. All patients had vertebral compression fractures, and patients showed elevated serum alkaline phosphatase levels. The gene for this form of OI has been recently identified, pigment epithelium-derived factor (PEDF), also known as serpin F1 (SERPINF1), an antangiogenic protein (unpublished data).

Osteogenesis Imperfecta Type VII (Moderate to Severe)

In addition to OI types V and VI, Glorieux reported on an autosomal recessive form of OI and used the designation OI type VII. This form occurred with a small genetic isolate among the First Nations community in northern Quebec, Canada (S89). The phenotype includes fractures at birth, blue sclerae, osteopenia, rhizomelia, and deformities of the lower extremities. The disorder has been localized to chromosome 3p22-24 and has been shown to result from a hypomorphic allele in CRTAP. The identification of the molecular basis of OI type VII changed the molecular view of the basis of disease with the identification of recessively inherited gene defects.

Histopathology of Bone in Osteogenesis Imperfecta

The range of histologic appearances of bone in the different OI phenotypes is as variable as the clinical phenotypes. Undermineralization and overmineralization of bone have been recognized within the same specimen. Bone histomorphology appears relatively normal in OI type I, but osteopenia secondary to thin lamellar plates and diminished cortical width is evident. Immature woven bone and lamellar disarray are characteristic of more severe OI phenotypes.

Treatment

Over the years, there have been multiple attempts to treat OI with a variety of vitamins, hormones, and drugs, none of which has been successful. The list includes administration of mineral supplements, fluoride, androgenic steroids, ascorbic acid, and vitamin D. During the past decade, bisphosphonates administered parenterally or orally to children and adults have shown favorable results. The bisphosphonate pamidronate administered intravenously increased bone mass, decreased skeletal pain, and decreased fracture incidence in children with severe OI. Similar results involving cyclic administration of pamidronate have been reported by other investigators. Dosage regimens in different series for children and adults range from 1 to 3 mg/kg, administered intravenously at 2- to 4-month intervals; lower dosage regimens also have been reported. Generally, reports indicate a significant increase in bone mass in children and a decrease in fracture rate. The effect is most marked in the spine, where vertebral remodeling may improve vertebral height. Metabolic studies have shown a decrease in serum ionized calcium and increase in serum parathyroid hormone.

Urinary excretion of N-telopeptide as an index of bone resorption decreased from 61% to 73%. The major side effects of intravenous bisphosphonate treatment include the acute-phase response (24 hours after infusion) and the occurrence of otitis and vestibular imbalance in a few patients. The currently recommended treatment regimen includes the use of a bisphosphonate, with adequate calcium and vitamin D supplementation to avoid the occurrence of hypercalcemia and to maintain normal serum vitamin D levels. Newer treatments for osteoporosis such as the RANK ligand inhibitor, denosunab, and antiscalcerostin antibody may hold promise for treatment of patients with OI.

The use of surgery to correct deformities and to facilitate weight bearing has been the subject of several reviews. Multiple osteotomies and realignment of a deformed bone over intramedullary rods is an option for many children with severe bowing. Indications include frequent fractures at the apex of the bow, impaired standing, and limb-length inequality owing to bowing. Expanding (telescoping) rods are best for growing children because they require fewer revisions. Spinal deformities are common and usually progressive. Surgical stabilization is most advisable in the teen years or early adulthood when patients can tolerate these complex reconstructions. Early basilar invagination may be halted with prophylactic posterior fusion of the occipital-cervical junction with plate fixation. Patients with severe brain stem compression may require anterior transoral decompression and posterior instrumented fusion. Patients with various types of OI seem to be at increased risk of premature osteoarthritis, the reasons for which are unclear. Total joint arthroplasty is usually successful in these patients, and referral is appropriate if arthroplasty is indicated.
Every child with OI benefits from appropriate rehabilitative therapy.\textsuperscript{102,103} Bracing with lightweight plastics as the child begins to walk can minimize microfracture and bowing of the upper femurs. Muscle-strengthening exercises are essential as primary care and after immobilization for fracture. Perhaps the most beneficial programs have been developed around swimming, preferably in heated pools, and as part of continuous rehabilitative medical care.

**EHLERS-DANLOS SYNDROME**

The heterogeneous group of disorders grouped together as EDS illustrates the genetic and clinical variability characteristic of the heritable disorders of connective tissue. The most cardinal feature of these disorders is the presence of joint hypermobility, associated with an increase in skin elasticity and skin fragility. In 1997, a simplified classification was proposed dividing EDS into six major clinical types. The classification includes the classic, hypermobility, vascular, kyphoscoliosis, arthrochalasia, and dermatosparaxis types, as well as several rarer EDS types grouped into “other forms.”\textsuperscript{104} Clinically, EDS can be difficult to separate, however, because of considerable overlap in phenotype findings.

**Classic Type**

The classic type of EDS accounts for about 80% of reported cases,\textsuperscript{105} and is inherited as an autosomal dominant trait. Originally, EDS was classified as types I and II, and now these types are classified as the classic form, although these subclassifications are still in use. Previously, types I and II EDS were distinguished from each other on the basis of joint laxity and skin fragility, which are less severe in type I than in type II EDS. Most prototypic forms of EDS (Figure 105-8) are characterized by various degrees of hyperextensibility of large and small joints, which are classic findings in EDS. It is crucial that hyperextensibility be defined, and differentiating mild “normal” laxity from hyperextensibility can be challenging. Beighton and colleagues\textsuperscript{104} have presented a clinically useful classification of joint laxity (Figure 105-9), as follows:

1. Passive dorsiflexion of the fifth digit beyond 90 degrees = 1 point for each hand
2. Passive apposition of the thumbs to the flexor surface of the radius = 1 point for each hand
3. Hyperextension of the elbows beyond 10 degrees = 1 point for each side
4. Hyperextension of the knees beyond 10 degrees = 1 point for each knee
5. Flexion of the trunk forward so that the palms can be placed flat on the ground = 1 point

A score of 5 or more points is defined as joint hypermobility.

Large joint hyperextensibility is seen in varying degrees in the classic form and decreases with age. Recurrent joint dislocations, periodic joint effusion related to trauma, and the eventual appearance of osteoarthritis pose significant management problems. Bilateral synovial thickening has been observed in EDS, along with the accumulation of small masses of crystalline material in synovial villi. It has been observed that EDS patients constituted 5% of cases in a pediatric arthritis clinic population.\textsuperscript{106} There is debate about whether affected infants may be born prematurely to affected mothers because of early rupture of amniotic membranes. Patients with EDS have characteristic facies, with a broad nasal root and epicanthal folds. They may have large, lax ears, and traction on the ears or elbows reveals skin hyperextensibility. Another sign of hypermobility is the ability to touch the tip of the tongue to the nose (Gorlin’s sign). In addition, absence of the lingual frenulum is characteristic for this disorder.

In EDS, the skin has a characteristically pleasant soft or “velvety” feel that can be appreciated by stroking the forearms. Thin, atrophic corrugated and hyperpigmented scars are found on the forehead, under the chin, and on the lower extremities (known as cigarette paper or papyraceous scars), although this is not a uniform finding. Typically, skin lesions heal slowly after injury or surgery. Molluscoid pseudotumors (violaceous subcutaneous tumors ranging in size from 0.5 to 3 cm) may be palpated in tissue over pressure points on the forearms and lower extremities and may be seen on radiographs. Although many patients claim to bruise easily, ecchymoses distributed on the extremities are found only in patients with the more severe forms of the disorder. Severe bilateral varicose veins are a common problem.

Associated pulmonary complications of EDS include spontaneous pneumothorax, pneumomediastinum, and...
It is associated with arterial dilatation and dislocations in the 20s and 30s with rheumatologic symptoms that can pose significant difficulties related to intubation, spontaneous arterial bleed ing during surgery, and ligation of vessels that tear under pressure complicate surgical maneuvers. Similarly, arteriography may be dangerous in these individuals. These patients are susceptible to rupture of internal viscera and may experience repeated rupture of diverticula on the antimesenteric border of the large bowel. Problems with pregnancy vary from preterm delivery to uterine or vascular rupture, although delivery is uneventful in many instances. Typical causes of death in EDS families have included gastrointestinal rupture, peripartum uterine rupture, rupture of the hepatic artery, and vascular ruptures.

In contrast to the other forms of EDS, EDS type IV is not associated with hyperextensibility of large joints, although small joints may be minimally hypermobile. These patients have thin, soft, transparent skin, through which a prominent venous pattern is seen, especially on their chest walls. Their skin is not velvety as in the classic form. Excessive bruising may occur. Vascular EDS includes, as a subgroup, patients who have been described as acrogeric—having characteristic thin faces, prominent eyes, and extremities that lack subcutaneous fat, giving the appearance of premature aging. Peripheral joint contractures and acro-osteolysis have been described.

Spontaneous hemopneumothorax associated with hemoptysis and mitral valve prolapse occurs frequently. Surgical repair of ruptured vessels or internal viscera is extremely difficult because of friable tissues. Anesthetic and surgical difficulties related to intubation, spontaneous arterial bleeding during surgery, and ligation of vessels that tear under pressure complicate surgical maneuvers. Similarly, arteriography may be dangerous in these individuals. These patients

**Hypermobility Type**

The hypermobile type of EDS is a dominantly inherited disorder that manifests as marked joint and spine hypermobility, recurrent joint dislocations, and the typical soft skin that is neither hyperextensible nor velvety. Individuals with EDS type III may have virtually normal skin. Because of the extent of joint laxity affecting large and small joints, these patients experience multiple dislocations and may require surgical repair. The shoulders, patellae, and temporomandibular joints are frequently sites of dislocation. Musculoskeletal pain may mimic that of fibromyalgia syndrome, and patients frequently seek medical attention for symptoms consistent with chronic pain.

One difficulty in this subtype is differentiating it from benign hypermobility syndrome. Benign hypermobility syndrome is used to describe patients with generalized joint laxity, associated musculoskeletal complaints, but normal skin. They do not have the classic stigmata of either EDS or Marfan syndrome. Many of these patients present in their 20s and 30s with rheumatologic symptoms that can pose problems in diagnosis and treatment. The precise approach and treatment for these patients are unclear.

**Structural and Molecular Pathology of the Classic and Hypermobile Types of Ehlers-Danlos Syndrome**

Abnormally large, small, or frayed dermal collagen fibrils and disordered elastic fibers have been observed in the classic and hypermobile forms of EDS by electron microscopy. Type V collagen is a heterotrimeric collagen composed of the products of three genes: COL5A1(V), COL5A2(V), and COL5A3(V). Type V collagen may stabilize type I collagen by co-assembling with that protein. Initially, linkage analysis was used to show that some families with the classic form of EDS (originally types I and II) were linked to COL5A1. Subsequently, it has been established that about 50% of patients with either the classic or hypermobility type of EDS have mutations in COL5A1(V) or COL5A2(V). There seems to be no genotype-phenotype correlation in these disorders, and no mutations have been identified in COL5A3(V).

In some cases of EDS classic type, heterozygosity for mutations in COL1A1(I) has been shown.

**Vascular Type**

The vascular type of EDS, an autosomal dominant disorder, is one of the most severe forms of EDS and was formerly referred to as EDS type IV. It is associated with arterial rupture, commonly involving iliac, splenic, or renal arteries or the aorta and resulting in either massive hematomas or death. Arterial rupture may lead to stroke or intracompartmental bleeding in a limb. Patients with vascular EDS are susceptible to rupture of internal viscera and may experience repeated rupture of diverticula on the antimesenteric border of the large bowel. Problems with pregnancy vary from preterm delivery to uterine or vascular rupture, although delivery is uneventful in many instances. Typical causes of death in EDS families have included gastrointestinal rupture, peripartum uterine rupture, rupture of the hepatic artery, and vascular ruptures.

EDS Vascular Type was formerly referred to as EDS type IV. It is associated with arterial rupture, commonly involving iliac, splenic, or renal arteries or the aorta and resulting in either massive hematomas or death. Arterial rupture may lead to stroke or intracompartmental bleeding in a limb. Patients with vascular EDS are susceptible to rupture of internal viscera and may experience repeated rupture of diverticula on the antimesenteric border of the large bowel. Problems with pregnancy vary from preterm delivery to uterine or vascular rupture, although delivery is uneventful in many instances. Typical causes of death in EDS families have included gastrointestinal rupture, peripartum uterine rupture, rupture of the hepatic artery, and vascular ruptures.
can be quite difficult to manage. Imaging studies may reveal normal-appearing aorta or other large vessels that rupture shortly after a “normal study.”

**Molecular Pathology**

Although EDS type IV was clinically recognized as a disorder distinct from the other forms of EDS, the finding that tissues from these individuals were deficient in type III collagen clearly distinguished this as a separate form of EDS. Type III collagen is a homotrimer [1(III)3] found in skin, blood vessels, and the walls of hollow viscera. Heterozygosity for mutations in the gene encoding COL3A1 leads to EDS vascular type and affects the synthesis and secretion of type III collagen. Various types of mutations have been identified including missense, nonsense, and deletions, and there is no correlation between the clinical phenotype and type III collagen mutation. In this disorder, the biochemical abnormalities include decreased or absent type III collagen or production of an abnormal homotrimer that is retained in the endoplasmic reticulum and, if secreted, contributes to abnormal matrix. Biochemical and mutational analysis for this disorder is available (GeneTests) and should be considered because this is dominantly inherited.

**Therapy in Classic, Hypermobility, and Vascular Types of Ehlers-Danlos Syndrome**

There are no specific treatments for the classic, hypermobility, and vascular forms of EDS. Supportive therapy is essential, however, for preservation of normal joint function and alleviation of joint pain. Planned exercise programs and muscle strengthening exercises are useful and do much to maintain a positive outlook in these individuals, who may have a poor prognosis if joint stability and articular surfaces are compromised by excessive activity or chronic trauma. Many children and young adults with large joint hypermobility are attracted to activities such as gymnastics and dance, and these activities promote hypermobility and joint damage. The presence of multiple ecchymoses raises concern about a bleeding diathesis, particularly at the time of elective surgery. Although there is no consistent basis for the hemorrhagic tendency in the classic and hyperextensibility forms of EDS, anecdotal, these patients tend to have greater blood losses than expected at surgery. In our center, we discourage pregnancy in patients with the vascular form because the mortality rate is increased.

**Arthrochalasia Type**

Formerly known as EDS types VIIA and VIIB, the arthrochalasia type of EDS is another autosomal dominant form resulting from mutations that cause faulty processing of type I collagen at the N-terminus. The arthrochalasia type of EDS is characterized by pronounced and generalized joint hypermobility, moderate cutaneous elasticity, moderate bruising, a characteristic round facies with midface hypoplasia, and significant short stature. The skin has a doughy feel and is fragile and hyperelastic. Kyphoscoliosis and muscle hypotonia are frequently present. These patients experience multiple dislocations, particularly involving large joints including the hips, knees, and ankles. These dislocations manifest in the newborn period, especially hip and ankle dislocations. Patients frequently need orthopedic surgery for joint dislocation, and their tissues are highly friable, which complicates orthopedic procedures.

**Molecular Pathology**

The two disorders EDS types VIIA and VIIB, now termed arthrochalasia type, result from mutations involving the N-terminal propeptide cleavage site of type I collagen. The arthrochalasia type of EDS has provided insight into the process of normal type I collagen fiber formation. The initial observation was of an accumulation of unprocessed procollagen within the dermis of affected individuals. With subsequent recognition that procollagen had N-terminal and C-terminal extension propeptides, and that separate enzymes were responsible for their removal, the syndrome became more sharply defined as an accumulation of procollagen with the N-terminal peptides still attached (pN collagen). Of the two distinctly different genetic abnormalities resulting in procollagen accumulation, the more frequent form is the mutational resistance of a procollagen cleavage site to the action of the N-terminal procollagen peptidase. The resistance results from an amino acid substitution or deletion in the proCOL1A1 (EDS type VIIA) or pro2COL2A1 (EDS type VIIB) chain, leading to a portion of the collagen chains containing an abnormal N-terminal extension; this results from mutations in COL1A1 or COL1A2 in exon 6 of the molecule, which alters the propeptidase cleavage site. Individuals with mutations in exon 6 of COL1A1 are more severely affected than individuals with similar mutations in COL1A2.

**Dermatosparaxis Type**

The dermatosparaxis type of EDS was formerly known as EDS type VIIC and is an autosomal recessive form of EDS. In this type, the skin is extremely fragile, soft, and doughy with easy bruising. The phenotype includes blue sclerae, marked joint hypermobility, micrognathia, large umbilical hernia, epiphyseal delay, and mild hirsutism. The dermatosparaxis type results from a deficiency of the procollagen N-propeptidase, in contrast to the arthrochalasia form, which involves the enzyme cleavage site, and individuals have been identified who are homozygous for mutations in the gene. This defect is homologous to the dermatosparaxis defect in sheep and cattle.

**Kyphoscoliosis Type**

The kyphoscoliosis type of EDS, formerly known as EDS type VI, is inherited as an autosomal recessive disease. The findings in this disorder include severe kyphoscoliosis noted at birth, recurrent joint dislocations, hyperextensible skin and joints, poor tone, and reduced muscle mass. The skin is grossly abnormal and has been described as pale, translucent, and velvety; on trauma, the skin shows gaping wounds that heal poorly. One difference in this form of EDS is that there is significant ocular involvement. Affected individuals have microcornea, retinal detachment, and glaucoma leading to blindness in some individuals. In addition, patients with severe kyphoscoliosis may develop respiratory
and cardiac compromise and ultimately cardiorespiratory failure.

**Molecular Pathology**

The kyphoscoliosis type of EDS results from lysyl hydroxylase deficiency. A variety of mutations within the lysyl hydroxylase gene have been defined and include premature stop codons, amino acid substitutions, internal deletions, and compound heterozygotes. Defective lysyl hydroxylase impairs the conversion of lysyl residues to hydroxylysine on procollagen peptides. The consequence of deficient hydroxylysine content of collagen is the effect it has on cross-linking, which helps stabilize the mature collagen molecule.

**Other Ehlers-Danlos Syndrome Types**

Numerous other rare forms of EDS have some overlap with other disorders or have been reported only in a small cohort of individuals, and these are not discussed in this chapter.

**MARFAN SYNDROME**

One of the most common inherited disorders of connective tissue, Marfan syndrome is an autosomal dominant disorder with a reported incidence of 1 in 10,000 to 20,000 individuals. Clinical presentations range from the severe infantile form to individuals who are only mildly affected. Although the most impressive findings in Marfan syndrome are relative to the musculoskeletal, cardiac, and ocular findings, affected individuals also have pulmonary, neurologic, and psychological complications. Marfan syndrome also has become one of the few genetic disorders for which there has been advocacy for treatment to slow the progression of the disease, and physicians need to recognize the phenotype because many affected individuals present with life-threatening emergencies.

**Clinical Features**

Marfan syndrome can be difficult to diagnose in some individuals and families, and it has been recognized that it has also been overdiagnosed. Stringent criteria for this diagnosis were proposed in 1996. The 1996 criteria rely on the recognition of “major” and “minor” clinical manifestations involving the skeletal, cardiovascular, dura, and ocular systems (excellent review in GeneReviews, Marfan syndrome). Major criteria include four of eight typical skeletal manifestations, ectopia lentis, aortic root dilation involving the sinuses of Valsalva or aortic dissection, and lumbosacral dural ectasia by computed tomography or MRI. Major criteria for establishing the diagnosis in a family member include having a parent, child, or sibling who meets major criteria independently, and the presence of a fibrillin-1 mutation known to cause the syndrome identified in a familial Marfan syndrome patient.

Establishing the diagnosis unequivocally in the absence of a family history requires a major manifestation from two systems and involvement of a third system. If a mutation known to cause Marfan syndrome is identified, the diagnosis requires one major criterion and involvement of a second organ system. The reason is that there is a great deal of intrafamilial variability in this disorder, and there are individuals who harbor heterozygosity for mutations but do not meet criteria for Marfan syndrome and may have different prognoses. Similar to other connective tissue disorders, there is wide variability in phenotypic expression.

Aortic disease leading to the formation of aneurysmal dilation and dissection is the main cause of morbidity and mortality in Marfan syndrome. Dilation of the aorta is found in 50% of children and progresses over time. Echocardiography shows that 60% to 80% of adult patients have dilation of the aortic root that may involve other segments of the thoracic aorta, the abdominal aorta, or even the carotid and intracranial arteries. Dissection usually begins above the coronary ostia and extends the entire length of the aorta. Of Marfan syndrome patients, 60% to 70% have mitral valve prolapse with regurgitation. Heart failure and myocardial infarction may complicate the course of Marfan syndrome patients. Pregnant women are at particular risk for aortic dissection, particularly women who already have aortic root dilation, and this should be taken into consideration when treating a woman of reproductive age with Marfan syndrome.

Arachnodactyly occurs in 90% of patients. Following are techniques that aid in determining arachnodactyly (Figure 105-10):

1. The thumb: The Steinberg test is positive when the thumb, enclosed in the clenched fist, extends beyond the hypothenar border.
2. The wrist: The Walker-Murdoch sign is positive when there is overlap of the thumb and fifth digit as they encircle the opposite wrist.
3. The metacarpal: The metacarpal index is done by radiographic determination and is the mean value of the lengths divided by the midpoint widths of the second, third, and fourth metacarpals. In normal subjects, the metacarpal index ranges from 5.4 to 7.9, whereas this range is 8.4 to 10.4 in patients with Marfan syndrome.

Thoracic kyphosis may be associated with reduced lung capacity and residual volume that may lead to pulmonary insufficiency. Dural ectasia, which may occur in 40% of patients, results from enlargement of the spinal canal owing to progressive ectasia of the dura and neural foramina and erosion of vertebral bone; this usually involves the lower spine. Diminished bone mineral density has been reported in several patients with Marfan syndrome. Ectopia lentis occurs in 50% to 80% of patients with Marfan syndrome. Subluxation of the lens is usually bilateral and appears by age 5 years. Although the lens is typically displaced upward, displacement into any quadrant may occur. Visual acuity is diminished in many patients because of lens subluxation or secondary acute glaucoma. Secondary myopia, retinal detachment, and iritis with loss of vision contribute to most of the ocular-related morbidity.

Marfan syndrome patients have been found to develop large epidural venous plexuses in the lumbar and cervical regions, a major diagnostic criterion for the syndrome. These engorged venous plexuses, which are visualized by MRI myelography, have been associated with the syndrome of spontaneous intracranial hypotension, which is also
associated with dural tears. Clinical signs are severe headache, back and leg pain, radiculopathies, and incontinence secondary to cerebral displacement. Spinal abnormalities in Marfan syndrome include increased interpedicle distance of nonrotated vertebrae, vertebral inversion (flattening of the normal kyphosis at the dorsal level and kyphosis or disappearance of the physiologic lordosis at the lumbar level), and vertebral dysplasia (dolichospondyllic, elongated vertebral bodies with increased concavity). Scoliosis constitutes one of the major management problems in Marfan syndrome. In one series, the average age of onset was 10.5 years (range, 3 to 15 years), with rapid progression during adolescence. If mechanical bracing or physical therapy fails to halt progression, spinal fusion should be considered, particularly when the curvature exceeds 45 to 50 degrees.

**Differential Diagnosis: Homocystinuria**

Homocystinuria, which shares several skeletal and ocular features with Marfan syndrome, is the prime diagnostic consideration. Homocystinuria is an autosomal recessive disease. The characteristic features of this metabolic disorder of sulfur metabolism are marfanoid phenotype with joint laxity, scoliosis, lens dislocation, early-onset osteoporosis, vascular thrombosis affecting arteries and veins owing to increased clotting activity and the cytotoxic effect of homocysteine on vascular endothelial cells, and mild mental retardation.

Cystathionine β-synthase deficiency is the most common cause of homocystinuria. Affected individuals have elevated levels of homocysteine and methionine, whereas cystathionine and cysteine levels in blood are decreased. This disorder is differentiated from Marfan syndrome because the direction of ectopia lentis is different than in Marfan syndrome, and there is no progressive aortic root dilation.

**Molecular Biology of Marfan Syndrome**

Fibrillin-1 protein is an important component of elastic and nonelastic connective tissues throughout the body. It is the main protein of a group of connective tissue microfibrils that are essential for normal elastic fibrillogenesis. In nonelastic tissues, the fibrillin-1-containing microfibril functions as an anchoring fiber. FBN-1 is a large gene (65 exons) located at chromosome 15q21.1. Since the first report of an FBN-1 mutation in Marfan syndrome in 1991, more than 500 different FBN-1 mutations have been described in Marfan syndrome and related disorders. FBN-1 mutations occur across a wide range of milder phenotypes that overlap the classic Marfan phenotype including dominantly inherited ectopia lentis, Shprintzen-Goldberg syndrome, and familial or isolated forms of aortic aneurysms. Most of these are private mutations (occur genetically independent with no “hot spot” in the molecule). The one exception is the rare infantile Marfan syndrome mutations that cluster between exons 24 and 26 and exon 32. Heterozygosity for missense, frameshifts, deletions and insertions, splice site alterations, and nonsense mutations all have been seen. Robinson and colleagues stated that at least 337 mainly unique mutations in the FBN-1 gene had been reported in Marfan syndrome up to that time. The clinical presentation of the fibrillinopathies caused by FBN-1 mutations ranged from isolated ectopia lentis to neonatal Marfan syndrome, which generally leads to death within the first 2 years of life.
Treatment

In 1972, the life span of untreated patients with classic Marfan syndrome was about 32 years. The early mortality in Marfan syndrome results primarily from complications associated with aortic dilation. This symmetric dilation of the sinuses of Valsalva is progressive throughout life and is often detectable in infancy. In the early 1970s, there was discussion on attempting to reduce the risk of aortic dissection in patients with Marfan syndrome. Shores and colleagues reported on a 10-year open-label trial of propranolol in 70 patients with Marfan syndrome. When compared with the control group, the treated individuals had a significantly lower rate of dilation of the aortic root, improved survival, and fewer treated patients reaching a clinical endpoint (death, congestive heart failure, aortic regurgitation, aortic dissection, or cardiovascular surgery).

More recent data generated from a mouse model of Marfan syndrome suggest excessive signaling by the transforming growth factor transforming growth factor (TGF)-β family of cytokines. There is evidence that aortic aneurysm in the mouse model of Marfan syndrome is associated with increased TGF-β signaling and TGF-β antagonists such as TGF-β-neutralizing antibody or the angiotensin II type 1 receptor blocker, losartan. In this mouse model, losartan (angiotensin II type 1 blockade) fully corrected the abnormalities in the aortic wall. There was some evidence that alveolar septation, which contributes to pulmonary problems in Marfan syndrome, was partially reversed with losartan treatment. Because this drug is in clinical use for hypertension, it could merit further investigation as a preventive treatment in Marfan syndrome. Clinical trials are now under way testing the use of losartan in Marfan syndrome and many individuals with Marfan syndrome are using losartan outside of clinical trials.

Electrocardiogram monitoring is done yearly until the aortic root diameter exceeds 45 mm, at which time monitoring is done every 6 months. Elective repair of aortic root disease before enlargement to 6 cm has occurred is preferable to emergency repair required for marked dilation or dissection. Surgical intervention is considered when the aortic root diameter approaches twice the upper limit of normal for body surface area, or the absolute measurement exceeds 50 to 55 mm. Total aortic root replacement with a composite valve graft (Bentall procedure) and coronary artery implantation have become the surgical procedures of choice and are associated with an 81% 10-year survival rate and a 75% 20-year survival rate. Mitral valve replacement and coronary artery implantation may be accomplished during the same procedure. Most importantly, repeated trials have shown that patients who undergo elective repair, as opposed to emergent repair, do substantially better.

Correction of scoliosis may be attempted with bracing; however, surgical repair should be considered when the curve exceeds 40 degrees. Progressive scoliosis in Marfan syndrome may require fixation with rods, and complications of joint laxity may require orthopedic correction. Arthropathy associated with excessive joint mobility may require orthopedic intervention. Dislocated lenses should not be removed surgically, unless more conventional means of correcting vision are ineffective.

LOEYS-DIETZ SYNDROME

In 2005, Loes and colleagues described individuals with a previously undescribed autosomal dominant aortic aneurysm syndrome. This disorder, now referred to as Loes-Dietz syndrome, is also characterized by hypertelorism, bifid uvula or cleft palate or both, and generalized arterial tortuosity with ascending aortic aneurysm and dissection. Other abnormal findings include craniosynostosis, structural brain abnormalities, mental retardation, congenital heart disease, and aneurysms with dissection throughout the arterial tree.

Some individuals with Loes-Dietz syndrome had a clinical phenotype that overlapped with Marfan syndrome, but none met diagnostic criteria set forth in 1996. Although Marfan syndrome is associated with progressive arterial disease, in Loes-Dietz syndrome the aneurysms tended to be particularly aggressive and rupture at an earlier stage and size than seen in Marfan syndrome. Heterozygosity for mutations in TGFβRI and TGFβRII has been identified.

From a management perspective, it is important to recognize these individuals because they are managed more aggressively than patients with Marfan syndrome. Aortic aneurysms are corrected at smaller sizes (4 cm), and complaints such as abdominal pain and headache should be thoroughly investigated because they may be associated with aneurysms.

CONGENITAL CONTRACTURAL ARACHNODACTYLY

Congenital contractural arachnodactyly is an autosomal dominant condition that includes tall stature, arachnodactyly, dolichostenomelia, and multiple contractures involving large joints. There is a characteristic “crumpled ear” deformity as a result of a flattened helix with partial obliteration of the concha. Marked deformity of the chest cage also occurs, and scoliosis may be progressive and severe. For unknown reasons, the contractures tend to become less severe with age. Radiographically, osteopenia can be seen. The ocular and typical cardiac lesions of classic Marfan syndrome are absent. This disorder results from heterozygosity for mutation in fibrillin-2 (FBNB-2). There are many other extremely rare disorders of connective tissue, especially with profound effects on the skin including the group of disorders termed cutis laxa and pseudoxanthoma elasticum.

SUMMARY

Heritable disorders of connective tissues are a heterogeneous group of disorders characterized by abnormalities in skeletal tissues including cartilage, bone, tendon, ligament, muscle, and skin. The clinical spectrum ranges from extreme short stature to excessively tall individuals, and the types of altered genes span all of the numerous gene families and pathways. Affected individuals usually need medical attention their entire lives and have been victims of appearing different because they cannot mask their abnormalities. Understanding and appreciation for the unique set of medical issues in each disorder would improve these individuals’ quality of life and their life span.
Selected References


Full references for this chapter can be found on www.expertconsult.com.
References


activated T cells, monocytes, and neutrophils are attracted to joint and secrete mediators that perpetuate inflammation and also attenuate immune regulation. The relative importance of individual cytokines varies between disease subtypes. Results from therapeutic trials support a role for tumor necrosis factor (TNF) in the pathology of polyarthritides forms of JIA and for interleukin (IL)-1β and IL-6 in systemic JIA (sJIA) (Table 106-1). The past decade has seen a step change in the quality of treatments in JIA. To build on this success, researchers need to be able to interrogate the vast array of biologic data that is emerging in the field of JIA and translate this knowledge into novel therapies and a more tailored treatment approach for our patients.

The term juvenile idiopathic arthritis (JIA) refers to a group of conditions, defined under the International League of Associations for Rheumatology (ILAR) classification as conditions starting before the sixteenth birthday that are characterized by arthritis of at least one joint that persists for 6 weeks or more, as a common feature. Although this classification has proven highly valuable for both clinical and basic research and enables more precisely defined subgroups and comparison of data from different studies, it does not encompass all aspects of the heterogeneity of childhood arthritis (e.g., use of the classification criteria by strict exclusion rules may lead to up to 30% of cases being designated as "unclassified"). In addition, certain common features may occur across several subtypes (e.g., factors associated with risk of autoimmune uveitis of JIA include positivity of the antinuclear antibody [ANA], early age at onset, and female sex). Girls with young-age onset of arthritis and ANA positivity represent a group of patients who have been proposed to represent a relatively homogeneous group, yet they currently fall into several subtypes within the ILAR classification. Similarly, genes and immunologic processes that influence the likelihood of mild oligoarthritis to extend to more severe arthritis may well overlap with etiopathologic factors involved in polyarticular JIA. Thus in the future a full molecular and genetic analysis of the heterogeneity of childhood arthritis may permit the development of a more mechanism-driven classification, which could help inform
on likely disease course, complications, or response to treatment.

Both genetic and environmental factors, as well as the manifestations of their interactions through inflammatory and immunologic mechanisms, are clearly implicated in the pathogenesis of JIA. Recent advances in the understanding of some clinical subtypes, most notably sJIA, have led to recognition of specific and distinct features and novel treatment options. However, it is also increasingly clear that some underlying genetic risk associations and immune regulators play a crucial role.
abnormalities may have effects across several types of childhood arthritis and indeed across many forms of autoimmune disease.

**HISTOLOGIC FEATURES OF JUVENILE IDIOPATHIC ARTHRITIS INFLAMED SYNOVIAU**

The pathologic hallmark of juvenile idiopathic arthritis is the inflamed synovium. Histology of this tissue shows thickened synovium that is highly vascular and shows marked hyperplasia of synoviocytes in the lining layer, as well as a dense infiltrate of inflammatory cells, comprising T cells, macrophages, dendritic cells, and in some cases B cells and natural killer (NK) cells. The hyperplastic synovial layer is highly vascular, with endothelium expressing markers of activation such as human leukocyte antigen (HLA)-DR and intracellular adhesion molecule 1 (ICAM-1). The vascularity is likely related to the increased production of proangiogenic factors such as vascular endothelial growth factor (VEGF), osteopontin, and the angiogenic chemokines, while recruitment of this inflammatory infiltrate is likely mediated by multiple chemokines shown to be increased in JIA including CCL5, CXCL10, CCL20, IL-8, and MCP-1 among others. The proteolytic enzymes matrix metalloproteinases MMP-1 and MMP-3 are abundantly expressed in synovial lining, and their levels correlate with degree of infiltration, in particular by myeloid CD68+ cells. The T cells that infiltrate the inflamed synovium, like those in the excess synovial fluid, are highly activated memory cells of both CD4 and CD8 populations.

**GENETICS OF JUVENILE IDIOPATHIC ARTHRITIS**

**KEY POINTS**

- There is strong evidence for a genetic component to the etiology of JIA.
- The strongest genetic associations of JIA are with genes of the MHC.
- Specific MHC class I alleles are associated with enthesitis-related arthritis (ERA).
- Distinct MHC class II alleles are associated with oligoarthritis and polyarthritis subtypes.
- Several genes that alter the threshold of T cell activation, or relate to cytokines and their receptors, are linked with JIA.

There is strong evidence for a genetic component to the etiology of the JIA diseases. Twin studies have shown concordance rates in monozygotic twins of between 20% and 40%. A study of 164 affected sibling pairs (ASPs) with JIA showed a 70% concordance for gender, 73% for age at disease onset, and 66% for disease course. Estimates suggest that the recurrence risk for siblings of a proband with JIA are between 15 and 30 times that of the general pediatric population, a figure that is as high as for insulin-dependent diabetes mellitus (IDDM) or multiple sclerosis. In addition, first-degree relatives of children with JIA have a higher rate of other autoimmune disease than controls. Despite this evidence, families of multiple affected siblings are still relatively rare. One caveat in comparing early studies is that some were performed using the American College of Rheumatology (ACR) classification criteria (at that time called juvenile rheumatoid arthritis), so they may not be directly comparable with studies that have used the current ILAR classification of JIA. However, there is now a growing body of data that has been analyzed using the ILAR criteria, making comparisons feasible.

Genetic influences on JIA susceptibility and phenotype are polygenic, such that JIA is thought of as a complex genetic trait. Although some studies have analyzed all forms of JIA together, the heterogeneity of childhood arthritis would lead to the prediction of different genetic risk associations for different disease phenotypes, and this has been confirmed in recent adequately powered studies. The unique features of different subtypes of JIA mean that it is perhaps unsurprising that there are also genetic associations with specific subtypes.

The strongest genetic associations with JIA are of genes that lie within the MHC, or HLA; these were the first to be documented as associated with childhood arthritis. The MHC region, located on chromosome 6 of the human genome, includes more than 200 genes, many of which are central to functions of the immune system. Numerous associations between HLA genes and JIA have been reported, and these involve many different populations (reviewed by Prahalad and Glass). The best characterized associations are with genes of the so-called MHC class I and class II loci, genes that code for heterodimeric proteins that present peptide antigen to the specific antigen receptor of T cells (the TCR).

Class I loci include HLA-A, HLA-B, and HLA-C, of which the earliest report was of the association between HLA-B27 and the form of JIA known as enthesitis-related arthritis (ERA), which includes children whose disease has parallels with adult ankylosing spondylitis (see Chapters 74 and 75). The HLA-A2 allele HLA-A*0201 is increased in several types of JIA, especially oligoarthritis and children with early-age onset. A recent study has also implicated HLA-C*0201 in association with persistent oligoarticular JIA.

Multiple studies have revealed associations of class II MHC loci (HLA-DR, HLA-DP, and HLA-DQ) with JIA, of which the strongest are DRB1*0801 and 11 with oligoarthritic JIA, as well as DRB1*1301, in particular in ANA-positive cases. Because of inheritance of genes in the MHC region together in so-called haplotypes due to linkage disequilibrium (LD), observed associations may in fact be due to genes within the haplotype, distinct from the locus first implicated. Several haplotypes across the MHC confer an increased risk for all types of JIA such as DRB1*08-DQA1*0401-DQB1*0402, which confers an odds ratio of 6.1 and 10.3 for persistent and extended oligoarticular JIA, respectively. Frequency of the DRB1*1301-DQA1*01-DQB1*06 haplotype distinguishes persistent from extended oligoarticular JIA, whereas DRB1*0801 and DRB1*1401 are associated with polyarticular JIA. Together, these effects may be large: In one study the presence of the combination of the HLA-DRB1*0801, HLA-DRB1*1101, and DPB1*0201 alleles conferred a relative risk of 236.
associations closely mirror those of the corresponding adult disease such as the strong association of the HLA-B27 allele with spondyloarthropathy and the HLA-DRB1*0401 with rheumatoid factor (RF)-positive polyarticular JIA. Although sJIA shows less strong associations with HLA alleles, even in this subtype, specific haplotypes (such as DRB1*11-DQA1*05-DQB1*03) are increased compared with control subjects. A recent study that used fine allele-specific typing across eight HLA loci in a large cohort and analysis by haplotype has confirmed that within class II loci, HLA-DR is the driving association, rather than HLA-DP. Remarkably, some HLA allele/JIA subtype associations show age-specific effects, in that they confer risk over a specific age range only. Another recent study confirmed both age- and gender-specific effects. Thus, for example, in polyarticular JIA the HLA-DRB1*0801 allele has risk effect in those whose arthritis starts after the age of 6, whereas in younger-onset polyarticular JIA the risk alleles are more closely related to those of patients with oligoarticular JIA, HLAB1*1103/1104. In addition to these risk alleles, some alleles, notably DRB1*0401, *0701 and *1501, are consistently reduced in frequency in JIA cases compared with controls, suggesting a protective effect, for some or even all subtypes. These associations between genes that code for proteins whose central function is the presentation of antigenic peptides to T cells implicate a T cell–driven process in at least part of the etiopathogenesis of JIA.

In addition to the genes coding for MHC proteins, a large number of other candidate genes (now totaling more than 100) have been the focus in different studies of JIA. Inflammatory cytokines have been an important target for drug development in JIA over the past decade, and similarly their gene polymorphisms have been a key area of scrutiny. Genes studied have included cytokine and chemokine genes such as IL-1, IL-6, TNF, MIF, and IL-10; their receptors (including IL-1R, IL-2RA, and CCR5); and key signaling molecules including CTLA4 and PTPN22. However, only a few of these loci have been independently validated: PTPN22, MIF, SLC11A6, WISP3, TNF, and in a recent meta-analysis, CCR5.

Genetic association at the TNF locus is complex to study given the position of the TNF gene within the MHC, but some data suggest that several HLA-independent TNF haplotypes are significantly associated with JIA, although the functional consequences of these alleles are not yet clear. The hypothesis suggesting a link between sJIA and IL-6 was proposed many years ago because several clinical features in sJIA resembled the phenotype of IL-6 overexpression (e.g., fevers, stunted growth, anemia). A polymorphism (−174G/C) in the regulatory region of the IL-6 gene alters transcription of IL-6 in response to IL-1 and LPS; sJIA patients have significantly lower frequency of the protective CC genotype, and the IL-6 −174G allele has been confirmed as a susceptibility gene for sJIA. Macrophage inhibitory factor (MIF) gene is associated with JIA. This polymorphism (MIF −173C) results in higher MIF production in the serum and synovium of JIA patients and has been suggested to be predictive of outcome of intra-articular steroid injections in sJIA.

The apparently conflicting studies on several of these candidate gene loci reflect two problems that have hindered progress: the heterogeneity of JIA, which means that combining all JIA into one study may lead to loss of detection of effects specific for a subtype, and opposing this, the difficulty of reaching adequate power for genetic studies, if stratification by subtype is preferred. Recently, new insights have come from the application of new knowledge from other diseases to the understanding of JIA genetics. Thus the highly successful Wellcome Trust case control Consortium performed genome-wide association studies in seven major common diseases including rheumatoid arthritis (RA). Extrapolation of new loci implicated in RA to JIA has been fruitful and has suggested that other key loci that show association with JIA are IL-2/IL-21 and IL-2RA, the α chain of the IL-2 receptor, also known as CD25. A recent study focused on specific subtypes of JIA and comparison with their adult disease counterparts. The genetic loci IL-23R and endoplasmic reticulum (ER) aminopeptidase-1 (ERAP1) have been identified as carrying risk associations with ankylosing spondylitis (AS) and psoriatic arthritis in adults. Genotyping of these genes in a large cohort of JIA cases has shown associations of both with the ERA subtype of JIA but no association with JIA as a whole, again emphasizing subtype-specific genetic features and likely pathogenesis. ERAP1 is of particular interest given its functional role in trimming of peptides within the ER, for presentation of peptide on, and folding of, MHC class I molecules, and given the strong evidence for MHC class I folding abnormalities in patients with the HLA-B27 risk alleles (HLA-B*2705 and *2702 among others); see also section on enthesitis-related arthritis (and Chapter 74). ERAP1 is also thought to have a role in trimming of cyto- kine receptors at the cell surface. IL-23 is of great functional interest because of the central role of IL-23 in the Th17 pathways, as well as the demonstration that Th17 cells may play a role in both JIA and adult AS/psoriatic arthritis.

New insights into the pathogenesis of childhood arthritis are eagerly awaited from large international efforts to perform genome-wide association studies (GWASs) in JIA, of which several are in progress, some targeting specific subtypes or specific research questions such as response to medication.

**ADAPTIVE IMMUNE SYSTEM**

### KEY POINTS

**Effector T cells secreting proinflammatory cytokines are enriched in the joint membrane and synovial fluid.**

**T cells with a regulatory phenotype have been detected within the joints of JIA patients and are enriched in persistent oligoarthritis patients.**

**A B cell gene expression signature has been detected in the peripheral blood of patients with early-age onset of JIA.**

**T Cells**

The strong association of many JIA subtypes with genetic variants at HLA loci, as well as the central role of HLA proteins in presenting peptide antigens to T cells for recognition, which is central to T cell function, led to intense investigation of the role of T cells in the pathology of JIA. Highly activated memory T cells make up a significant proportion of the inflammatory infiltrate in JIA and express an
“oligoclonal” or restricted set of T cell receptors (TCRs).\textsuperscript{14,16,17} Specific T cell clones can be long-lived and are detectable in different inflamed joints.\textsuperscript{11} Nevertheless, it is still unclear whether these clones represent autoreactive T cells specific for an “arthritogenic” epitope, akin to islet cell antigens in type I diabetes. Although immune activation leads to tissue damage in JIA, the role of self-antigen recognition in this process remains unclear. However, the association of JIA with genetic loci that influence the threshold of T cell activation such as PTPN22, as well as those central to recognition of antigen by T cells, supports the concept of JIA as a disease of dysregulated adaptive immunity.

Early work examining animal models of arthritis led to the hypothesis that cells of the T helper 1 (Th1) lineage, secreting interferon-γ (IFN-γ), were central to pathogenesis. Th1 cells are recruited to the joint by high levels of chemokines CCL5, macrophage inflammatory protein (MIP)-1α, and IP-10\textsuperscript{43} and make up the majority of T cells in the JIA joint.\textsuperscript{12} However, in both mouse models of arthritis and early adult clinical trials, blocking IFN-γ has offered little clinical benefit, which suggests that other players may be important to pathogenicity. TNF, the prototypic inflammatory cytokine in arthritis, is secreted by T cells and macrophages within the joint and is detectable within inflamed synovial tissue,\textsuperscript{44} and to a lesser extent in synovial fluid.\textsuperscript{45} The success of TNF blockade in polyarthritis and extended oligoarthritis subtypes implicates TNF in JIA pathology.\textsuperscript{46}

Still, up to a third of patients fail to respond adequately or relapse on anti-TNF therapy and a recently discovered T cell population, “Th17” secreting IL-17 and IL-22, may account for this recalcitrant disease.\textsuperscript{47} IL-17 causes significant bony and cartilage damage in the joint by promoting neutrophil influx via the secretion of IL-8 and synergizing with IL-1β and TNF to drive metalloproteinase secretion and osteoclast activation.\textsuperscript{48} Th17 cells are enriched in the joints of JIA patients, and Th17 numbers correlate with the severity of disease course in oligoarthritis.\textsuperscript{11} Recent evidence suggests that a significant proportion of the inflammatory T cells in the joint have a Th17 ancestry, raising the hope that Th17 blockade will be an effective treatment in some subtypes of JIA.\textsuperscript{49}

In addition to proinflammatory processes, there is strong evidence for ongoing immune regulation in JIA. Children with persistent oligoarticular JIA have high numbers of a regulatory subset of T cells (Treg) within the joint that express CD25 and Foxp3, and the number of Tregs is significantly higher in the children with persistent oligoarticular JIA than those with the more severe extended oligoarticular disease.\textsuperscript{50,51} In addition, T cells specific for the conserved self-antigen heat shock proteins (HSPs) play a similar regulatory role.\textsuperscript{52} Although synovial Treg suppress effector T cell functions in vitro, inflammatory cytokines such as IL-6 and TNF, present in the arthritic joint,\textsuperscript{10,53} may attenuate Treg function in vivo. Treatment strategies that expand Treg numbers and augment function are currently under investigation.\textsuperscript{54,56}

**Antigen-Presenting Cells**

Dendritic cells are specialized antigen-presenting cells (APCs) that activate T cells, as well as promote the differentiation of effector T cell functions. Comparatively little is known about their role in JIA. In the inflamed joint, myeloid dendritic cells (mDCs) are localized to the synovial lining layer, whereas interferon-α-secreting plasmacytoid dendritic cells (pDCs) are in T and B cell–rich aggregates.\textsuperscript{55} Both subsets are found in high numbers within synovial fluid when compared with peripheral blood. Synovial mDCs express high levels of co-stimulatory molecules including CD80, CD86, and RANK,\textsuperscript{7,28} which augment T cell activation and lead to the secretion of proinflammatory cytokines and chemokines. A triggering receptor expressed on myeloid cells (Trem-1) may play a role in activating mDCs in response to local hypoxia.\textsuperscript{59} In contrast to mDCs, some studies suggest that synovial pDCs may provide a regulatory function. pDCs within the joint also do express markers of activation\textsuperscript{50} but secrete high levels of Granzyme B, which limits T cell proliferation.\textsuperscript{61}

**B Cells**

There has been a renewed interest in B cell effector function in adult arthritis following the efficacy of the anti-CD20 agent, rituximab, in RA. In JIA, the exact role of B cells as effector or regulatory players is still unclear. B cell–derived autoantibodies, antinuclear antibodies (ANAs), RF, and anticitrullinated protein antibodies (ACPAs) are common in some JIA subtypes, but none are considered to be directly pathogenic (see Chapter 107). Interestingly, patients that are ANA positive are more likely to have T–B cell aggregates in their synovial membrane, suggesting a link between lymphoid neogenesis and autoantibody production.\textsuperscript{52} Although lymphoid aggregates are seen in JIA synovium, mature germinal centers are rare when compared with RA.\textsuperscript{63} Total B cell numbers do not vary between JIA patients and healthy controls, but oligoarthritis patients have higher numbers of transitional type B cells in the synovial fluid exudate (CD38high, CD24high) in peripheral blood than controls.\textsuperscript{61} Patients with an early onset of disease also have expression of B cell– and immunoglobulin-related genes in peripheral blood.\textsuperscript{64} The B cell signature, as well as frequent ANA positivity in oligoarthritis patients, has led some authors to suggest a major role for B cells in the pathogenesis of this subtype.\textsuperscript{64}

**INNATE IMMUNE SYSTEM**

**KEY POINTS**

**Defects in the innate immune system have been most closely associated with systemic JIA.**

**Neutrophils and monocytes secrete myeloid-related protein 8/14, IL-6, IL-1β, and TNF, which drive systemic inflammation.**

Many cytokines detectable in the inflamed joint are the products of the innate immune system. The successful targeting of these cytokines with biologic agents (e.g., IL-1β and IL-6) heavily implicates the innate immune system in the pathophysiology of JIA, particularly the sJIA subtype.

**Macrophages/Monocytes**

Monocytes and their tissue counterparts, macrophages, are key effector cells of the innate immune system and have
been linked to the pathogenesis of autoimmune arthritis for several decades. In JIA, analysis of gene expression has detected an activated macrophage gene expression signature in the cells within synovial fluid of early oligoarthritis patients at risk of extension and a monocyte signature in the peripheral blood of patients with an older-onset oligoarthritis, as well as RF-positive polyarthritis patients. Monocytes typically make up equivalent proportions of synovial fluid and peripheral blood mononuclear cells (¿10%), but they have a highly activated phenotype, secreting high levels of IL-6, IL-1¿, and TNF, within the joint. Synovial monocytes also secrete vascular endothelial growth factor (VEGF) and osteopontin (OPN), which contribute to the vascularity of inflamed pannus, and the chemokine CCL20, which drives Th17 cell recruitment to the joint.

Neutrophils

Neutrophils may make up the major fraction of the synovial infiltrate but have remained an infrequent research interest in JIA. SJIA patients have high levels of the heterodimeric myeloid related protein (MRP) 8/14 secreted by activated monocytes and neutrophils. Roth and colleagues have shown that MRP 8/14, although secreted by cells of the innate immune system, induces autoreactive CD8+ T cells, illustrating how cross-talk between the innate and adaptive immune system can lead to chronic inflammation.

Examination of neutrophils from JIA has found more than 700 genes that are differentially expressed in patients’ blood compared to controls. Genes linked to IL-8 and IFN-¿ were prominent among these differences but failed to return to baseline levels after clinical remission. Rather, disease quiescence was associated with an increase in regulatory genes, such as transforming growth factor-¿ (TGF-¿) and retinoic acid, suggesting that remission reflects a balanced state of inflammation and regulation rather than true immunologic resetting.

Stromal Cells

Resident tissue stromal cells are important targets of both the innate and adaptive immune system and may play an important role in defining the anatomic location of inflammation after systemic immune dysregulation. Synovial fibroblasts from JIA patients secrete metalloproteinases and chemokines in response to locally secreted proinflammatory cytokines and may be an important therapeutic target of the future.

DISEASE SUBTYPE–SPECIFIC PATHOGENESIS

**KEY POINTS**

- Excessive secretion of IL-1¿ and IL-6 contributes to many of the clinical features seen in patients with SJIA.
- The balance between inflammatory and regulatory T cell populations may determine the severity of clinical course in oligoarticular forms of JIA.

Systemic Juvenile Idiopathic Arthritis

A key role for IL-1¿, IL-6, and IL-18, cytokines of the innate immune system, as well as the absence of autoantibodies or a strong association with MHC, have led many to propose SJIA as an autoinflammatory disease. This hypothesis is supported by the close mirroring of these cytokines with the characteristic fever of SJIA, the high levels of IL-6 and IL-18 in serum from active SJIA patients, and enrichment of monocytes in SJIA blood, secreting high levels of IL-1¿ after activation. Early studies reported a significantly higher ratio of IL-1 receptor antagonist (IL-1ra) to IL-1 in these SJIA patients, although this may be difficult to interpret because IL-1¿ and IL-1¿ are highly labile in serum. Early evidence of successful treatment with a soluble IL-1, anakinra, at least for some patients with SJIA supports a role for IL-1.

Interestingly, IL-1¿ secretion in SJIA may be mediated by MRP 8/14 because once this protein is removed from SJIA serum, the potential to induce IL-1¿ is almost completely attenuated. This study confirmed that levels of serum MRP 8/14 are high in active SJIA and that this measure is both sensitive and specific in distinguishing SJIA from other important diagnoses such as infection or hematopoietic malignancy.

IL-18, a macrophage-derived proinflammatory cytokine with a similar signal transduction pathway to IL-1¿, is also grossly elevated in JIA with serum levels correlating with disease activity. High IL-18 levels may account for the defective NK cytotoxic function found in SJIA. Follow-up studies of anakinra have cast doubt on the primacy of IL-1¿ in SJIA because more than 50% of patients are either non-responders or relapse on treatment. Even in responders, it is likely that other inflammatory mediators are involved because treatment response to anakinra fails to correlate with either pretreatment IL-1¿ or IL-18 secretion or changes in gene expression of the IL-1¿ pathway.

IL-6 is another key player in the pathogenesis of SJIA, secreted by a range of cells including monocytes. IL-6 levels are elevated in the serum and synovial fluid of SJIA patients, and levels correlate with disease activity. Spikes and falls in fever are secondary to the circadian rhythm of IL-6, and many clinical features including growth failure and osteoporosis can be explained by high levels of IL-6. Trial data show efficacy of IL-6 blockade with the monoclonal antibody to soluble IL-6 receptor, tocilizumab. A trial of tocilizumab demonstrated impressive results with 86% of patients achieving an ACR50 response (see Chapter 107). IL-6 expression is upregulated by IL-1¿, so the success of IL-6 blockade may represent its role as the final common pathway for inflammation, both IL-1 dependent and independent.
Gene expression profiling from sJIA patients has confirmed pathways that involve IL-6 and IL-1 and that can distinguish active from inactive patients. In addition to these roles for monocytes in sJIA, defects of NK cell function, in particular perforin function, are well recognized. The defect in perforin function in active sJIA has been seen to reverse on successful treatment with autologous stem cell transplantation.

**Macrophage Activation Syndrome**

MAS is a potentially life-threatening complication of sJIA (as well as sometimes complicating other rheumatologic conditions such as JSLE) that results from immune activation of pathogenic T cells and hemophagocytic macrophages. These macrophages express CD163, a scavenger receptor that recognizes haptoglobin-hemoglobin (HP-Hb) complexes. Increased uptake of these complexes within the macrophage leads to production of ferritin, explaining the hyperferritinemia associated with MAS. Soluble forms of CD25, the α chain of the IL-2 receptor expressed on T cells, and CD163 are useful biomarkers for MAS in JIA.

MAS shares many similarities with a group of inherited disorders called hemophagocytic lymphohistiocytosis (HLH). Defects in the perforin gene and related genes involved its cytosolic secretory pathway (MUNC13-4, Rab27a, and SH2D1A) have been identified as causes of HLH. When examined in JIA, polymorphisms in MUNC13-4 and perforin have been associated with MAS in some studies, but not others. How defects in cytotoxic function lead to MAS is still unclear, but it has been proposed that appropriate clearance of microbial antigens or activated macrophages is prevented, leading to chronic immune stimulation.

A recent gene expression profiling study of active sJIA showed that patients with active sJIA, and in particular those with MAS, have a signature in PBMC that is similar to patients with familial HLH and characteristic of immature erythropoiesis.

**Oligoarthritis**

There is increasing evidence that the balance between inflammation and regulation plays a role in driving the clinical phenotype of oligoarthritis patients. The dominance of Treg within the joints of patients with persistent oligoarthritis may explain the relatively benign prognosis of this subgroup, in contrast to the extended subgroup that has a low ratio of Treg to Th17 cells. Those who remain in the mild prognosis group have higher numbers of regulatory T cells in the joint, both as defined by CD25 or Foxp3 expression, as well as of CD30 Treg cells, which are specific to the human self-antigen hsp60, and these CD30 regulatory cells have recently been shown to reside within the Foxp3 population.

Although oligoarthritis JIA patients all present with few joints involved and apparently mild disease, evidence suggests that underlying differences between the groups that remain mild and those that go on to extend are present even early in disease. Thus genetic differences between persistent and extended oligoarticular phenotypes include variation in the MHC class I allele associations and IL-10. A recent study using gene expression, cellular and cytokine profiling showed that those children destined to go on to extend had a more heavily IFN-γ-driven and activated macrophage signature in synovial cells, as well as differences in levels of the chemokine RANTES (CCL5), from the start of disease even before extension occurs. Similarly, proteomic profiles have been shown to be distinct, early in disease, in those who go on to extended disease. These data have led to the concept of a separate group of patients with oligoarthritis who are destined to progress to more severe disease, the so-called extended-to-be group. There is strong evidence to link Th17 cells with autoimmune uveitis, and so agents targeting IL-17, currently under trial in RA, may ameliorate both joint and eye disease in extended oligoarticular JIA.

**Rheumatoid Factor–Positive Polyarthritis**

The presence of RF and a severe erosive disease course suggests that children in this subgroup represent an early presentation of adult-onset RA. Indeed, this subtype of JIA shares genetic associations with adult rheumatoid arthritis, including the HLA DRB1 locus, in particular the association with the DRB1*0401 allele at this locus. However, few studies have directly compared immunology between adult and childhood onset of disease. Studies of ACPA positivity in children with JIA have shown association with RF-positive polyarticular disease, and there is evidence to support a pathogenic role for RF and associated autoantibodies against citrullinated peptides by fixation of complement on synovial endothelium. In animals, transfer of ACPAs does not lead to arthritis but enhances tissue injury when there is a background of low-grade joint inflammation. Children with RF-positive polyarthritis may share histologic features with adult RA patients, with lymphoid follicles and germinal centers that may be more abundant than in RF-negative patients.

**Rheumatoid Factor–Negative Polyarthritis**

RF-negative polyarthritis represents a more heterogeneous group than RF-positive polyarthritis, with several studies suggesting clinically distinct subgroups (see Chapter 107). Analysis of gene expression data from polyarthritis patients showed three separate signatures. The first was associated with monocyte-related genes and had the highest proportion of RF-positive patients. The second signature found in RF-negative patients was associated with reduced CD8 T cells and increased plasmacytoid DC. Interestingly, patients did not have overlap between signatures 1 and 3, suggesting distinct immunopathology. It is noteworthy that a significant proportion of patients (24 out of a total of 61) who were younger and had a higher rate of ANA positivity did not have overlap between signatures 1 and 3, suggesting distinct immunopathology. A further gene expression study suggested these patients may overlap with ANA-positive oligoarthritis patients and that age of onset of disease is a more important distinguishing factor than the number of involved joints. These data are consistent with results of cytokine levels in plasma and synovial fluid from oligoarthritis and polyarthritis. Both subtypes clustered together and were distinct from sJIA and RA patients. As
discussed, there are similar immune pathways in RF-negative polyarthritis and oligoarthritis patients, and overall disease expression may depend on the balance between immune regulation and inflammation.

**Enthesitis-Related Arthritis**

The HLA-B27 allele, recognized many years ago for its association with autoimmune arthritis, appears to play a particular role in pathogenesis through its molecular properties. The presence of a cysteine residue at position 67 of the HLA-B27 α1 helix heavy chain makes possible a disulfide bond that promotes the formation of homodimers of the B27 chain. These homodimers, which can form in the absence of antigenic peptide and can induce ER stress, have been shown to be proinflammatory and to be ligands for NK cell receptors such as KIR3DL1 and KIR3DL2 (see also Chapter 74).

Cells expressing HLA-B27, which is prone to this abnormal folding of the HLA molecule, and which in the context of microbial antigens drives high expression of IL-23, may contribute to factors promoting Th17 cells. This hypothesis may also explain the association between ERA and inflammatory bowel disease because Th17 cells also drive pathology in the latter condition and pathogenic cells readily recirculate from gut to joint. Patients with ERA may be particularly prone to disease induction by bacterial pathogens because cells from patients have high expression of pathogen recognition molecules TLR2 and TLR4 when compared with healthy controls. Along with putative misfolding in the ER by HLA-B27, this allele may alter the antigenic peptides available for T cell recognition. It is interesting that ERAP1, which is strongly associated with adult ankylosing spondylitis (AS), is also associated with pediatric-onset ERA. This protein is thought to affect the repertoire of peptides available to bind class I MHC, by cleavage of N-terminal amino acids from peptide precursors in the ER.

**Psoriatic Arthritis**

The etiology of psoriatic arthritis shares features with enthesitis-related arthritis, having a genetic association with MHC class I loci, a clonal expansion of CD8 T cells in inflamed joints, and a putative role for Th17 cells. Certainly in adult psoriasis and psoriatic arthritis there is strong evidence that Th17 cytokines IL-22 and IL-17 contribute to the disease process. The receptor for IL-23, a key cytokine that induces Th17 cells, is linked to the juvenile form of the disease and IL-23 blockade appears promising in trials of skin and joint disease in adult patients. The innate immune system may also play a role because a recent study found genetic associations with autoinflammatory genes in juvenile psoriatic arthritis patients.

**TRANSLATION FROM UNDERSTANDING PATHOGENESIS TO CLINICAL PRACTICE**

The heterogeneity of childhood arthritis is complex but is gradually being characterized and harnessed, through the application of novel methods including gene expression profiling, proteomics, and high-throughput genetics. Major progress in the understanding of immunologic processes involved has occurred. In this chapter, cellular processes at play in JIA have been considered. The immune system acts in concert to create vastly complex networks. The challenge is to understand the functional hierarchy of these networks and discover checkpoints that will be amenable to therapeutic targeting in the future.

In addition to new treatments, better biomarkers are required for use in the clinic. For example, in the case of MAS, as knowledge of pathogenesis improves, more biomarkers will become available to predict adverse outcomes. The task in coming years will be to integrate the vast body of data that will be generated through these novel approaches in order to allow the development of more precise classification definitions and, perhaps more importantly, predictive tools and algorithms with which to drive treatment choices for patients and so allow accurate stratification for those who should receive novel biologic agents (e.g., TNF or IL-6 blockade) early in their disease, to achieve early remission.

**References**


74. Agarwal S, Misra R, Aggarwal A: Interleukin 17 levels are increased in juvenile idiopathic arthritis synovial fluid and induce synovial fibroblasts to produce proinflammatory cytokines and matrix metalloproteinases, J Rheumatol 35:515–519, 2008.


110. Myles A, Aggarwal A: Expression of Toll-like receptors 2 and 4 is increased in peripheral blood and synovial fluid monocytes of patients with enthesitis-related arthritis subtype of juvenile idiopathic arthriti, Rheumatology (Oxford) 50:481–488, 2011.


The references for this chapter can also be found on www.expertconsult.com.
Treatment of Juvenile Idiopathic Arthritis

JOYCE J. HSU • TZIELAN CHANG LEE • CHRISTY I. SANDBORG

KEY POINTS

Juvenile idiopathic arthritis (JIA) is the umbrella term for the family of childhood arthritides of unknown cause.

JIA affects at least 1 in 1000 children.

Before wide use of methotrexate and biologic therapies, 50% of children with JIA would reach adulthood with significant disabilities.

The goals of treatment in JIA are complete suppression of inflammation on medications and remission when possible.

Oligoarticular JIA, characterized by early age of onset, female predominance, antinuclear antibody positivity, and frequent subacute anterior uveitis, occurs only in childhood.

Systemic JIA is distinct from other subgroups of JIA in its equal sex distribution, lack of autoantibodies and human leukocyte antigen associations, and increased responsiveness to interleukin (IL)-1 and IL-6 inhibition compared with tumor necrosis factor inhibition.

Enthesitis-related arthritis typically occurs in children greater than 6 years of age, but sacroillitis may not develop until adolescence.

Arthritis in a child can result in overgrowth (in the knee causing length discrepancy) and undergrowth (in the temporomandibular joint causing micrognathia) of joints.

Because bone erosions on conventional radiographs are late radiographic findings in growing children, early joint damage may require different imaging modalities.

Considerable progress in understanding the genetics and pathogenesis of JIA has revealed subtype-specific associations, as well as some common mechanisms of disease that will translate to more effective and targeted therapies.

Juvenile idiopathic arthritis (JIA) is the most common rheumatic disease in childhood, but actual estimates of prevalence and incidence vary remarkably in different geographic regions, ranging from 7 to 400 per 100,000 children, reflecting variations in disease reporting, classification, and ethnic and environmental differences in disease expression.1 Reasonable working estimates are 150 per 100,000 children, which makes JIA one of the most common chronic diseases of childhood. There are significant differences in the disease manifestations in children compared with adults, with some types occurring exclusively in children. This chapter discusses the current understanding of the key clinical features of the various forms of JIA, differential diagnoses, treatment approaches, prognosis, and outcomes. The rapid growth in understanding the biologic basis for JIA and the ongoing development of targeted therapies for rheumatic diseases are likely to lead to enhancements in these recommendations. Data concerning long-term disease course and outcome in children with JIA in this modern treatment era remain limited, and there are international prospective inception cohorts of children with arthritis designed to address these gaps in knowledge.2,3

CLASSIFICATION CRITERIA FOR JUVENILE IDIOPATHIC ARTHRITIS

Differential Diagnosis

In the past, different groups had used various types of nomenclature to classify children with persistent arthritis including “juvenile rheumatoid arthritis” (American College of Rheumatology [ACR]) and “juvenile chronic arthritis” (European League Against Rheumatism), which created problems in comparing research studies and outcomes. The goal of the International League of Associations for Rheumatology (ILAR) is to identify subtypes of JIA for research purposes that are homogeneous and mutually exclusive. JIA classification is currently based on predominant clinical and laboratory features and the number of involved joints at disease onset.4 There is a continual renewal process, with the second revision occurring in Edmonton in 2001, which is presented in Table 107-1. However, classification systems are ever evolving, and categorization may evolve to more biologically and genetically similar subgrouping, especially with recent advances in etiology and pathogenesis. For example, age of onset may be a more biologically relevant parameter to distinguish between subtypes of JIA than classification based on number of involved joints. PBMC gene expression analysis reveals biologic differences between patients with early-onset (<6 years) and late-onset (>6 years) JIA, which was independent of oligoarthritis or polyarthritis subtype.5 Ravelli and colleagues6 provided clinical support for this approach, showing that antinuclear antibody (ANA)-positive patients with oligoarthritis and rheumatoid factor (RF)-negative polyarthritis were similar in terms of early age at onset, female predilection, increased frequency of asymmetric arthritis, and increased frequency of uveitis.

Pattern recognition is perhaps the most significant skill needed by clinicians in the diagnostic evaluation of patients. The usual patterns in children with rheumatic diseases often overlap with malignancies, infection, and trauma (especially nonaccidental). Therefore it is crucial to evaluate these diagnostic possibilities first before accepting the diagnosis of JIA. The ILAR categories are meant to simplify classification and are useful for typical presentations of
### Table 107-1  International League of Associations for Rheumatology Classification Criteria for Juvenile Idiopathic Arthritis (JIA)†

<table>
<thead>
<tr>
<th>Subcategory</th>
<th>Definition</th>
<th>Exclusions</th>
</tr>
</thead>
</table>
| Oligoarthritis                     | Arthritis affecting 1-4 joints during the first 6 mo of disease            | a. Psoriasis or a history of psoriasis in the patient or first-degree relative  
|                                    | 1. Persistent oligoarthritis: Affecting ≤4 joints throughout the disease course | b. Arthritis in an HLA-B27-positive male beginning after the sixth birthday  
|                                    | 2. Extended oligoarthritis: affecting a total of >4 joints after the first 6 mo of disease | c. Ankylosing spondylitis, ERA, sacroiliitis with inflammatory bowel disease, reactive arthritis, acute anterior uveitis, or a history of one of these disorders in a first-degree relative  
|                                    |                                                                          | d. The presence of IgM RF on at least 2 occasions at least 3 mo apart  
|                                    |                                                                          | e. The presence of systemic JIA in the patient                          |
| RF-Negative Polyarthritis          | 1. Arthritis affecting ≥5 joints during the first 6 mo of disease and     | a, b, c, d, e                                                             |
|                                    | 2. Test for RF is negative                                               |                                                                          |
| RF-Positive Polyarthritis          | 1. Arthritis affecting ≥5 joints during the first 6 mo of disease, and    |                                                                          |
|                                    | 2. ≥2 positive RF tests (as routinely defined in an accredited laboratory), at least 3 mo apart during the first 6 mo of disease |                                                                          |
| Psoriatic Arthritis                | 1. Arthritis and psoriasis, or                                            | b, c, d, e                                                               |
|                                    | 2. Arthritis and at least 2 of the following:  
|                                    | 1. Dactylitis                                                           |                                                                          |
|                                    | 2. Nail pitting (minimum of 2 pits on 1 or more nails at any time) or onycholysis |                                                                          |
|                                    | 3. Psoriasis in a first-degree relative                                  |                                                                          |
| Enthesitis-Related Arthritis       | 1. Arthritis and enthesis, or                                            | a, d, e                                                                  |
|                                    | 2. Arthritis or enthesis, with at least 2 of the following:  
|                                    | a. The presence of or a history of sacroiliac joint tenderness and/or inflammatory lumbosacral pain |                                                                          |
|                                    | b. The presence of HLA-B27                                              |                                                                          |
|                                    | c. Onset of arthritis in a male ≥6 yr of age  
|                                    | d. Acute (symptomatic) anterior uveitis                                  |                                                                          |
|                                    | e. History of ankylosing spondylitis, ERA, sacroiliitis with inflammatory bowel disease, reactive arthritis, acute anterior uveitis in a first-degree relative |                                                                          |
| Systemic JIA                       | Arthritis in 1 or more joints with, or preceded by, fever of at least 2 weeks’ duration that is documented to be daily and quotidian (fever that rises to ≥39° C once a day and returns to ≤37° C between fever peaks) for at least 3 days, and accompanied by 1 or more of the following:  
|                                    | 1. Evanescent (nonfixed) erythematous rash                             | a, b, c, d                                                               |
|                                    | 2. Generalized lymph node enlargement                                   |                                                                          |
|                                    | 3. Hepatomegaly and/or splenomegaly                                    |                                                                          |
|                                    | 4. Serositis                                                            |                                                                          |
| Undifferentiated Arthritis         | Arthritis that fulfills criteria in no category or in ≥2 of the above categories |                                                                          |

ERA, enthesitis-related arthritis; RF, rheumatoid factor.

disease. However, if a patient does not easily fit into the ILAR classification system, clinicians must carefully consider all other possibilities, rheumatologic and nonrheumatologic. To this end, ILAR recommends the following descriptors in order to obtain more clinical information: age at onset; characteristics of articular involvement; serologies (ANA, RF, anticitrullinated protein antibody [ACPA]; uveitis; and the human leukocyte antigen (HLA) allelic associations. For example, a 2-year-old girl who wakes up crying at night from back or hip pain should immediately trigger alarms of infection or malignancy rather than a rheumatic condition. This case is different from a 12-year-old boy with back stiffness and hip entheseal pain, which would be a classic presentation of enthesitis-related arthritis (ERA) in this age group.

Among patients with acute lymphocytic leukemia (ALL), 15% to 30% present with musculoskeletal symptoms and may be misdiagnosed as JIA.† Up to 75% of children ultimately diagnosed with ALL presenting with musculoskeletal complaints did not have blasts in the peripheral blood at the time of evaluation by pediatric rheumatologists, although low white blood cell count, mild
thrombocytopenia, and nighttime pain were early indicators of ALL. ANA status, rash, radiographic abnormalities, and objective signs of arthritis were not helpful in distinguishing between ALL and JIA because they occurred at similar rates in both groups. Each subcategory is discussed later, focusing on clinical manifestations, diagnostic features, treatment, outcome, and prognosis. Treatment recommendations are discussed later for each of the subtypes of JIA on the basis of current available evidence including the recently developed ACR recommendations for the treatment of JIA.

RHEUMATOID FACTOR–NEGATIVE POLYARTHRITIS

Clinical Manifestations and Diagnostic Features

RF-negative polyarthritis makes up 10% to 30% of all JIA cases, with a bimodal distribution of age of onset with the first peak at 1 to 4 years of age and the second peak at 10 to 12 years. Girls are more commonly affected than boys with a ratio of 3.2:1, and subacute anterior uveitis occurs in 4% to 25%. Any joint may be affected in RF-negative polyarthritis JIA, with more involvement of hip, shoulder, cervical spine, and distal interphalangeal joints than in adults. The arthritis is often usually insidious and can be symmetric or asymmetric, affecting both large and small joints. Some authors distinguish between two clinical subgroups on the basis of ANA status: (1) an ANA-positive form that resembles oligoarthritis, except for the number of joints affected in the first 6 months of disease, consisting of young girls (younger than age 6) with an asymmetric-onset arthritis and at a high risk of uveitis, and (2) an ANA-negative form that is similar to adult-onset RF-negative rheumatoid arthritis (RA), characterized by symmetric synovitis of large and small joints, with onset in a slightly older age group (ages 7 to 9). The similarities between ANA-positive, RF-negative polyarthritis and oligoarthritis have led to the hypothesis that these two entities are actually in the same disease spectrum.

RF-negative polyarthritis may be associated with elevated acute phase reactants, mild anemia, and ANA positivity in up to 40%. Even though RF is negative, 50% to 80% of patients are ACPA positive, using high-sensitivity (but low-specificity) testing methods.

Differential Diagnosis

The differential diagnosis of RF-negative polyarthritis includes the other JIA subtypes including ERA, which should be considered particularly in boys older than 6 years of age because sacroiliac involvement may not occur until adolescence. Other major diagnostic considerations include other rheumatic conditions such as lupus, especially in an older girl who is ANA positive; lymphoma; and leukemia. Neisseria gonorrhoeae and Lyme disease can present as an acute polyarthritis.

Treatment

With the current development of increasingly more effective biologic treatment for arthritis, pediatric rheumatologists now aim to achieve complete disease remission as early as possible in the disease course. Several studies support the paradigm of treating aggressively to reach inactive disease as early as possible, which may ultimately lead to better outcome such as improved quality of life, shorter periods of time spent in active disease, and less long-term joint damage. Active disease in the first 2 years was significantly associated with the duration of active disease in the following 3 years, and conversely, improved disease control with disease-modifying antirheumatic drugs (DMARDs) and/or biologics was associated with improved outcomes. To this end, children with polyarthritis require a disease-modifying agent as soon as practically possible after the diagnosis has been confirmed. If nonsteroidal anti-inflammatory drugs (NSAIDs) are initially used as monotherapy, with or without intra-articular steroids (IASs), continued disease activity at no more than 2 months should prompt escalation of therapy. As shown in Figure 107-1, methotrexate (MTX) is the first DMARD of choice, but if there is an inadequate response to MTX by 2 months, anti–tumor necrosis factor (TNF) agents should be initiated. The different TNF inhibitors have not been directly compared against each other, so it is not possible to determine which might be most effective and safe in a given patient. Switching between TNF inhibitors in children has not been well studied. Sulfasalazine and leflunomide can still be used before an anti-TNF agent in mild disease, although evidence suggests that leflunomide may be slightly less effective than MTX. Polyarticular disease course of any type, which does not respond well to MTX or anti-TNF agents, is now increasingly being treated with a range of newer biologics. Physiotherapy (PT) is important for all children with JIA, for stretching, muscle building, and consequent joint protection. Children with hand involvement need occupational therapy (OT) assessment and input regarding writing and school accommodations. The following paragraphs summarize specific information about different therapies used in children and may be applied to all types of JIA as appropriate. (See Table 107-2 for medications, dosing, route, and safety monitoring recommendations.)

Nonsteroidal Anti-inflammatory Drug Use in Children

NSAIDs may help control symptoms but do not alter the natural history of JIA. In general, NSAIDs should only be considered as monotherapy in initial therapy in low disease activity. If control is not achieved in 1 to 2 months, additional therapy should be considered. NSAIDs are frequently used for symptom control as an adjunctive therapy to more definitive therapies. Gastric protection with H2 blockers or proton pump inhibitors may be required.

Intra-articular Steroid Injections

The use of intra-articular triamcinolone hexacetonide (THA), 1 mg/kg in large joints such as the knee and 0.5 mg/kg in smaller joints such as the ankle, has been found to be superior to triamcinolone acetonide, betamethasone, and methylprednisolone acetate in randomized controlled trials. Early treatment is associated with better outcome, and IASs are expected to result in clinical improvement
arthritis for at least 4 months. Therefore if arthritis recurs, joint injections can be repeated up to three times in a 12-month period. Difficult-to-reach joints such as the hip, sacroiliac (SI) joint, temporomandibular joint (TMJ), and subtalar joint may be injected using ultrasound or fluoroscopy. IASs should be administered under sedation or general anesthesia for the young. The use of IASs for active arthritis has been recommended by the ACR guidelines regardless of concurrent therapy, JIA subtype, disease activity, prognostic features, or joint contracture. If the duration of clinical improvement is shorter than 4 months, systemic treatment (e.g., MTX) may be indicated. 

**Corticosteroid Use in Children with Juvenile Idiopathic Arthritis**

In general, systemic corticosteroids should be used sparingly in the treatment of any subgroup of JIA because of the severe morbidity associated with chronic use, even at low doses. Newer therapeutic modalities such as biologics reduce the dependence for any corticosteroids and/or limit the doses needed. The evidence for use of corticosteroids for the synovitis of JIA is controversial and conflicting, and the ACR recommendations for JIA treatment were unable to make any recommendations for corticosteroid use, except in systemic juvenile idiopathic arthritis (sJIA) for severe systemic features. 

However, in some cases, low doses of corticosteroids (<0.1 mg prednisone equivalent/kg/day) or brief high-dose regimens (intravenous methylprednisolone 30 mg/kg/day for 1 to 3 days) may be used in polyarticular JIA, in order to bridge constitutional features of pain and fatigue while waiting for DMARDs or biologic therapies to reach their therapeutic effect. Local use of corticosteroids for uveitis is discussed later in “Uveitis.”

**Methotrexate**

MTX is the most commonly used DMARD in JIA. In a retrospective cohort study involving all JIA subtypes, the strongest predictor of response to MTX at 6 months of treatment was the time from diagnosis to start of MTX, suggesting that starting MTX early will lead to a better response. The ACR guidelines support a maximum dose of 0.6 mg/kg once weekly (equivalent to 15 mg/m²/week, maximal 25 mg/week) of parenteral MTX. In patients with lower disease severity, lower doses (8 to 12.5 mg/m²/week oral or parenteral) may be effective, and these doses are similar in safety. Most pediatric rheumatologists start folic acid at 1 mg/day, but daily folate supplementation remains controversial. Routine monitoring of liver enzyme tests should be done as noted in Table 107-2, but liver biopsies are not indicated except in unusual circumstances. Some children develop an intolerance to MTX, and leflunomide may be used as an alternative.

**Tumor Necrosis Factor Inhibitors**

Anti-TNF agents are effective in many children with polyarticular-course JIA of any onset type who fail to respond fully to MTX. The first TNF inhibitor to be studied in JIA, etanercept, demonstrated efficacy and safety in a novel randomized withdrawal trial design in patients with polyarticular JIA, in which the study design was based on an open-label period, followed by randomization of responders to placebo or study drug with the primary endpoint being time to flare. Since that initial randomized control study in 2000, long-term studies and registries have continued to demonstrate the safety and efficacy of etanercept in children.

---

**Figure 107-1**  Polyarticular-course juvenile idiopathic arthritis treatment algorithm including rheumatoid factor (RF)-negative, RF-positive, extended oligoarticular, and psoriatic arthritis treatment. The treatment goal is remission of disease activity and is stratified by severity of disease. More disease activity or high-risk features should prompt disease-modifying antirheumatic drug use as initial therapy, whereas nonsteroidal anti-inflammatory drug (NSAID) monotherapy could be used initially for 1 to 2 months. If NSAID monotherapy is ineffective, treatment should be escalated. ACPA, anticitrullinated protein antibody; IAS, intra-articular steroid; IL, interleukin; MTX, methotrexate; TNF, tumor necrosis factor.
with polyarthritis, with and without MTX.\textsuperscript{27,32} The frequency of major serious adverse events (SAEs) have been low, with adjusted rates of 0.12 events per patient year and 0.03 medically important infections per patient year in open-label extensions.\textsuperscript{32} Importantly, no cases of lupus, demyelinating disorders, malignancies, opportunistic infections, or tuberculosis were seen. Although there have been several long-term studies of etanercept with international registries, the total number of patient-years available for evaluation is still too small to be able to detect rare long-term SAEs such as malignancy or demyelinating diseases.

A 1-year prospective, observational study to compare combination etanercept and MTX with etanercept monotherapy demonstrated that the likelihood of achieving 70% disease control (ACR Pedi 70) was increased with combination therapy (odds ratio of 2.1, 95% CI 1.2 to 3.5) as compared with etanercept monotherapy.\textsuperscript{27} In this cohort, there were 24 infectious SAEs and 23 noninfectious SAEs including three malignancies in 496 patients over 12 months. Growth delay in JIA was improved and reversed with the use of etanercept.\textsuperscript{13}

An international, multicenter, randomized, double-blind, placebo-controlled trial of infliximab provided

### Table 107-2  Commonly Used Medications in Juvenile Idiopathic Arthritis\textsuperscript{9,176}

<table>
<thead>
<tr>
<th>Medication</th>
<th>Typical Maximum Dose</th>
<th>Typical Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abatacept</td>
<td>10 mg/kg (max 1000 mg) IV</td>
<td>Load at 0, 2, 4 wk, then every 4 wk</td>
</tr>
<tr>
<td>Adalimumab</td>
<td>24 mg/m$^2$</td>
<td>Every 2 wk</td>
</tr>
<tr>
<td>Anakinra</td>
<td>2 mg/kg (max 100 mg) SQ</td>
<td>Daily</td>
</tr>
<tr>
<td>Cyclosporine</td>
<td>6 mg/kg/day orally</td>
<td>Divided twice a day</td>
</tr>
<tr>
<td>Diclofenac (SR preparation available)</td>
<td>0.8 mg/kg/wk (max 50 mg) SQ</td>
<td>Divided 1-3 times a day</td>
</tr>
<tr>
<td>Etanercept</td>
<td>0.4 mg/kg/dose (max 25 mg) SQ</td>
<td>Once weekly</td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>≥6 mo of age: 30-40 mg/day</td>
<td>2x/wk</td>
</tr>
<tr>
<td>Indomethacin (SR preparation available)</td>
<td>10 mg/kg/dose IV</td>
<td>Divided 3-4 times a day</td>
</tr>
<tr>
<td>Infliximab</td>
<td>&gt;2 yr of age: 20-30 mg/kg/day (max 1 g/day)</td>
<td>Load at 0, 2, 6 wk, then every 4 wk</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>≥40 kg: 20 mg orally</td>
<td>Every 2 wk</td>
</tr>
<tr>
<td>Naproxen</td>
<td>&lt;15 kg: 5 mg orally</td>
<td>Daily</td>
</tr>
<tr>
<td>Piroxicam</td>
<td>16-25 kg: 10 mg orally</td>
<td>Daily</td>
</tr>
<tr>
<td>Rilonacept</td>
<td>2.2-4.4 mg/kg SQ</td>
<td>Once weekly</td>
</tr>
<tr>
<td>Rituximab</td>
<td>750 mg/m$^2$ (max 1000 mg) IV</td>
<td>Twice 2 wk apart</td>
</tr>
<tr>
<td>Sulfasalazine</td>
<td>50 mg/kg/day (max 2 g) orally</td>
<td>Divided twice a day</td>
</tr>
<tr>
<td>Tacrolimus</td>
<td>0.2 mg/kg/day orally</td>
<td>Divided twice a day</td>
</tr>
<tr>
<td>Thalidomide</td>
<td>5 mg/kg/dose orally</td>
<td>Daily</td>
</tr>
<tr>
<td>Tocilizumab</td>
<td>&lt;30 kg: 12 mg/kg IV</td>
<td>Every 2 wk</td>
</tr>
<tr>
<td></td>
<td>&gt;30 kg: 8 mg/kg IV</td>
<td></td>
</tr>
</tbody>
</table>

*The effectiveness of the 0.8 mg/kg/wk dose has been evaluated in JIA patients.\textsuperscript{177,178} IV, intravenous; SQ, subcutaneous; SR, sustained release.

### Summary of Recommendations for Medication Safety Monitoring

#### Nonsteroidal Anti-inflammatory Drugs

- Complete blood count, liver enzymes, serum creatinine
  - Before or soon after initiation of routine use
  - Repeat approximately twice yearly for chronic daily use
  - Repeat approximately once yearly for routine use (3-4 days/wk)

#### Methotrexate

- Complete blood count, liver enzymes, serum creatinine
  - Before initiation
  - Approximately 1 mo after initiation
  - Approximately 1-2 mo after increase in dose
  - Repeat approximately every 3-4 mo if prior results normal and dose stable

#### Tumor Necrosis Factor Inhibitors

- Complete blood count, liver enzymes, serum creatinine
  - Before initiation
  - Repeat approximately every 3-6 mo
  - Tuberculosis screening
  - Before initiation
  - Repeat approximately once yearly
important lessons regarding dosing in children. The initial
dose of 3 mg/kg/infusion based on adult studies did not
demonstrate efficacy at 3 months compared with placebo;
however, a separate arm suggested that 6 mg/kg/infusion
had better pharmacokinetics leading to better effectiveness.
In a 4-year, long-term, open-label extension, 14% of patients
discontinued infliximab due to SAEs including six infec-
tious SAEs in 120 patients over 52 weeks. Significant
infusion reactions associated with infliximab antibodies and
asymptomatic development of antinuclear antibodies were
observed.

Adalimumab was shown to be efficacious and safe in a
randomized placebo-controlled withdrawal trial in patients
with moderately to severely active polyarticular JIA with or
without MTX, and there was a trend toward more improve-
ment with combination therapy, although the study was not
statistically powered to measure this difference. The dose
used was adalimumab 24 mg/m² (maximum dose, 40 mg)
subcutaneously every other week. SAEs possibly related to
adalimumab occurred in 14 patients (total 177), seven of
which were serious infections.

In August 2009, the U.S. Food and Drug Administration
issued a black box warning on the increased risk of cancer,
particularly lymphoma, in children and adolescents receiving
TNF inhibitors for arthritis or inflammatory bowel
disease (IBD). The other adverse events involved with
TNF inhibitors are similar to that of adults and have been
reported in children as well including serious infections,
demyelinating processes, optic neuritis, injection site
reactions or infusion reactions, and development of auto-
immune conditions. Because of the voluntary nature of
reporting rare adverse drug events in the United States, the
actual risk of any of these rare events is not known.

Abatacept

The T cell co-stimulatory inhibitor, CTLA4-Ig (abatacept),
was studied in an international placebo-controlled random-
ized withdrawal trial in 190 patients with active polyarthritis
regardless of onset type who had inadequate response or
intolerance to one DMARD in the past including TNF
inhibitors. Concurrent MTX was allowed. Thirty-percent
improvement (ACR Pedi 30) response rates were 76% in
biologic-naïve patients and 39% in patients with prior bio-
logic therapy. Abatacept continued to be clinically signifi-
cant with durable efficacy in patients with JIA, and some
patients require longer periods for optimal response (>3
to 4 months) compared with TNF inhibitors. No cases of
tuberculosis or malignancy were detected, but patient
numbers were small and follow-up time was limited.

Other Biologics

The ACR guidelines recommend consideration of ritux-
imab as a treatment option for patients who have received
TNF inhibitor and abatacept sequentially and have high
disease activity. IL-6 and IL-1 inhibition are discussed in
more detail later in “Systemic Juvenile Idiopathic Arthritis”
but have not been specifically studied in polyarticular
disease. The use of nonbiologic DMARD combinations
(such as MTX plus sulfasalazine and/or hydroxychloro-
quine) has not been studied in children.

Outcome

Early response to treatment was an important predictor
of long-term outcome. Symmetric arthritis and early
hand involvement predicted future disability and poorer
overall well-being. The ACR guidelines for JIA treatment
also consider the following as poor prognostic factors in
patients with RF-negative polyarthritis: arthritis of hip or
cervical spine; positive ACPAs; and radiographic damage
(erosions or joint space narrowing by radiograph).

RF-negative polyarthritis has a variable outcome,
which shows the heterogeneity of the subtype, but the
overall prognosis appears to be better than RF-positive
polyarticular JIA. Approximately 30% of children will
go into long-term remission off medication, with the
chance of remission being highest in the first 5 years of
disease. However, flare of disease 2 years after reaching
clinical remission off medication had occurred in 69% of
patients.

RHEUMATOID FACTOR–POSITIVE
POLYARTHRITIS

Clinical Manifestations and Diagnostic Features

RF-positive polyarthritis is a well-characterized JIA subcat-
cegory and is part of the same disease spectrum as adult
RF-positive RA, sharing immunogenetic and serologic
factors. This subcategory makes up 5% to 10% of cases of
JIA and is more common in girls, with reported female-
to-male ratios between 5.7 and 12.8:1.0. Age of onset
occurs in late childhood or adolescence and typically is an
aggressive, symmetric polyarthritis affecting the small joints
of the hands, as well as large joint involvement in a pattern
that resembles RA. These children frequently have more
than 30 joints with arthritis. Hip involvement is common
and may be debilitating. The arthritis can be quite severe,
often resulting in bony erosions and joint destruction.
Radiologic changes tend to take place early, especially in
hands and feet. With active disease, patients may occasion-
ally have mild systemic signs and symptoms such as weight
loss, low-grade fever, malaise, mild hepatosplenomegaly, or
lymphadenopathy. Rheumatoid nodules occur in up to 10%
of cases, most frequently around the elbow. Other extra-
articular manifestations are reported less often than in
adults. Uveitis is an unusual feature of this subtype, occur-
ing in only about 0% to 2% of patients.

Polyarthritis may be associated with mild to moderate
inflammation such as elevated acute-phase reactants and
a normocytic, normochromic anemia. By definition, all
patients have IgM-anti-IgG RF. The ANA test is positive
in about 55% of patients, and ACPAs have been reported
in 57% to 73%.

Differential Diagnosis

The differential diagnosis of RF-positive polyarthritis
includes other JIA subcategories, especially when there is
no confirmed RF-positive test on two occasions. Such cases
are frequently unclassified in the JIA system; however, man-
agement and therapy remain similar.
Treatment

Treatment algorithms and ACR guidelines for patients with all types of polyarthritis are discussed in Figure 107-1. However, because children with RF-positive polyarthritis are at higher risk of prolonged erosive arthritis compared with other types of JIA, these children should be considered to be in the more severe disease category, requiring rapid escalation of treatment if even mild disease activity persists. Rather than an initial period of NSAID monotherapy, RF-positive polyarthritis patients should receive MTX, at the time of diagnosis, with rapid addition of a TNF inhibitor if response is not adequate. Some children benefit from multiple joint injections to maintain control of the arthritis.

Outcome

Children with RF-positive polyarthritis have a poorer long-term prognosis than the other JIA subcategories. Inactive disease is difficult to achieve with 84% of the disease course consisting of active disease, and only 5% of patients were able to maintain remission after cessation of therapy.4,5

OLIGOARTICULAR JUVENILE IDIOPATHIC ARTHRITIS

Clinical and Diagnostic Features

Oligoarticular JIA accounts for up to 20% of all new rheumatic diagnoses in the general pediatric rheumatology clinic and is the most prevalent of all the JIA subcategories, comprising 30% to 60% of all JIA patients in North America and Europe.41 Oligoarticular JIA has no adult equivalent. The peak age of onset occurs in Caucasian children ages 2 to 4 years from the United States and Europe.40 Females are affected more commonly than males, 3:1. Two general subgroups are recognized within oligoarticular JIA: extended oligoarticular involvement, in which many additional joints develop arthritis after the initial 6 months, as contrasted to persistent oligoarticular JIA, in which the number of joints affected remains less than 5. Currently, there is no single reliable predictor of extension, but symmetric disease, ankle and/or wrist involvement, and an elevated erythrocyte sedimentation rate (ESR) in the first 6 months of disease may indicate likelihood of extension.51,52 Disease extension to the extended subtype has been reported to be 30% to 50% at 4 to 6 years after disease onset.51,53

Oligoarticular JIA usually presents as an asymmetric arthritis affecting one or two large joints, especially of the lower extremities, with the knee being the most commonly affected, followed by the ankle, wrist, and digits. Hand involvement is the third most commonly affected location, but this pattern may portend the later onset of psoriatic arthritis.54 Involvement of the hip and back, especially in young children, is so unusual that extensive evaluation is warranted to rule out other conditions such as infection or tumors.

Significant constitutional and systemic symptoms are unusual in oligoarticular JIA and, if present, should raise concern regarding the accuracy of the initial diagnosis. Pain in an obviously inflamed joint is surprisingly minimal compared with septic arthritis, and in up to 25% of cases the symptoms may be subtle with parents only noticing a limp and joint swelling. In a young child there is reluctance to walk and bear weight with a return to crawling. There is a high risk for developing a relatively asymptomatic chronic uveitis, especially in ANA-positive individuals, requiring regular ophthalmology examinations to detect early changes. (See Table 107-3 for guidelines and “Uveitis” later for treatment.) Other complications that can be prevented in most patients with proper treatment include growth discrepancies in muscle tone and bulk, leg length, and the development of micrognathia and joint contractures.

Among children with oligoarthritis, 75% to 85% have a positive ANA (70% to 80% in persistent oligoarticular JIA and 80% to 95% in extended oligoarticular JIA),42 with low to moderate titers (1:40 to 1:320). The rate of ANA positivity is even higher in girls with an early onset.5 Patients are RF negative, although ACPAs have been detected in some patients with oligoarticular JIA, depending on the enzyme-linked immunoreceptor assay (ELISA) method used for screening.5 Some suggest that ANA serology should delineate a homogeneous group of arthritis patients, independent of the course of arthritis and number of joints involved.6

Table 107-3  American Academy of Pediatrics Recommended Ophthalmologic Screening Frequency for Asymptomatic Uveitis in Juvenile Idiopathic Arthritis (JIA) Patients*67

<table>
<thead>
<tr>
<th>JIA Subtype</th>
<th>Antinuclear Antibody Status</th>
<th>Age of JIA Onset (yr)</th>
<th>Duration of Disease (yr)</th>
<th>Uveitis Risk Category</th>
<th>Eye Examination Frequency (mo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligoarthritis or</td>
<td>+</td>
<td>≤6</td>
<td>≤4</td>
<td>High</td>
<td>3</td>
</tr>
<tr>
<td>polyarthritis</td>
<td>+</td>
<td>≤6</td>
<td>&gt;4</td>
<td>Moderate</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>≤6</td>
<td>&gt;7</td>
<td>Low</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>&gt;6</td>
<td>≤4</td>
<td>Moderate</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>&gt;6</td>
<td>≤4</td>
<td>Low</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>−</td>
<td>≤6</td>
<td>&gt;4</td>
<td>Moderate</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>−</td>
<td>≥6</td>
<td>NA</td>
<td>Low</td>
<td>12</td>
</tr>
<tr>
<td>Systemic JIA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Low</td>
<td>12</td>
</tr>
</tbody>
</table>

*Several investigators have recommended the most intensive screening ophthalmologic exams in ANA-positive girls with JIA onset <7 yr of age independent of JIA subtype. Patients with JIA onset ≤5 yr should have eye exams every 3 mo until 7 yr after JIA diagnosis. On the other hand, in a large, German, population-based study using the International League of Associations for Rheumatology classification system of JIA, certain modifications were proposed that were less conservative than the American Academy of Pediatrics in terms of frequency of eye exams, so the timing of screening exams remains unclear.66 NA, not applicable.
The typical child with oligoarticular JIA will have normal white blood counts, normal or mild to moderately elevated acute-phase reactants, and in some cases, mild anemia. Elevated acute-phase reactants may suggest other conditions with inflammatory features such as IBD or malignancy.

**Differential Diagnosis**

The differential diagnosis of oligoarticular JIA includes other JIA subtypes, especially ERA and psoriatic JIA; other rheumatic diseases of childhood; and nonrheumatic causes of joint pain and swelling such as septic arthritis, benign or malignant tumors, reactive arthritis, foreign body synovitis, pigmented villonodular synovitis, arterial-venous malformation, bleeding disorders (such as hemophilia), or severe trauma including nonaccidental injury. Mild trauma such as from a fall does not cause persistent joint swelling, and trauma is rarely a cause of joint swelling unless there is an internal derangement seen in older, not younger, children. Children with hypermobility can develop transient joint effusions after exercise, but this should not be long-lasting swelling. Lyme disease (in an endemic area) can cause recurrent monoarticular arthritis (typically involving the knee and often popliteal cysts), usually for less than 6 weeks. As described earlier, ALL may present with bone and joint pain and swelling, often monoarticular. If there is any concern of malignancy or infection, a complete blood count with manual differential and peripheral smear is crucial and bone marrow examination should be performed if indicated.

A common initial response by general pediatricians when evaluating children with joint pain or swelling is to believe the problem is mechanical. However, in young children and toddlers, orthopedic problems such as meniscal tears or ligamentous injury are exceedingly rare because of the nature of pediatric musculoskeletal development and anatomy. In general, ligaments and tendons are stronger than growing bone in children, and the bone-ligament or bone-tendon junction is the weakest link. Therefore fractures are more common in children than adults, but nonaccidental trauma should always be considered, especially in children younger than 3 years.

**Treatment**

The treatment strategies for extended versus persistent oligoarticular JIA differ in both the approach to therapy and the intensity of escalating treatment on the basis of the number of joints at risk for significant damage. For extended oligoarticular JIA, treatment approaches are similar to RF-positive or RF-negative polyarticular JIA because these subtypes all have polyarticular involvement (see Figure 107-1). Treatment of persistent oligoarticular JIA is usually approached in a stepwise fashion (Figure 107-2). For these patients with a history of arthritis in four or fewer joints, the initial treatment is an NSAID, with or without an adjunctive IAS, followed by MTX if there is an inadequate response to one or more IASs. Initiation of MTX was recommended in the ACR guidelines for JIA as initial treatment for children with oligoarticular JIA with high disease activity and poor prognostic features, defined as involvement of hip, cervical spine, ankle or wrist, high inflammatory markers, or radiologic changes. Partial or complete remission on MTX can be induced in 60% to 70% with extended oligoarticular JIA. TNF inhibition should be considered in resistant cases, often in combination with MTX.

Concurrent sulfasalazine and hydroxychloroquine have been used with variable success. Sulfasalazine is more typically used with HLA-B27–associated arthritis and ERA.

---

**Figure 107-2**  Persistent oligoarticular juvenile idiopathic arthritis (JIA) treatment algorithm including oligoarticular course psoriatic JIA. The treatment goal is remission of disease activity and is stratified by severity of disease. More severe disease activity or high-risk features should prompt disease-modifying antirheumatic drug use as initial therapy, whereas nonsteroidal anti-inflammatory drug (NSAID) monotherapy could be used initially for 2 to 3 months. If NSAID monotherapy is ineffective, therapy should be escalated. IAS, intra-articular steroid; MTX, methotrexate; TNF, tumor necrosis factor.
Special Considerations: Knee Monoarthritis

Because knee monoarthritis is the most common presentation of oligoarticular JIA, specific management is discussed here. The two treatments to be considered are NSAIDs and IASs, with evidence that IASs may be more effective even though most pediatric rheumatologists use NSAIDs before IASs. An initial trial of NSAIDs may be conducted with the hopes of avoiding IASs in some patients, but the risk of IASs must be weighed against the cost of continued active arthritis. Choosing among treatment strategies involves a tradeoff between several different outcomes including duration of active arthritis, potential for long-term complications, adverse effects of therapies, discomfort of daily medications or potentially painful procedures and anesthesia, as well as parental preferences. Synovectomy is not indicated in oligoarticular JIA.

Outcome

Long-term studies of adults treated before the use of biologic agents have shown that up to 50% of adults who had oligoarticular JIA may have ongoing active disease or functional problems in adulthood, and the rate of remission after 6 to 10 years from onset of disease ranges only from 23% to 47%. Ongoing disease activity and extension of joint involvement is related to a poor outcome and radiographic damage, and therefore emphasis on control of disease activity is critical. Morbidity from long-term inflammation can cause problems as well such as leg-length discrepancy with knee arthritis, muscle atrophy, bony overgrowth, and joint contractions (Figure 107-3), as well as other growth abnormalities such as micrognathia.

Patients with persistent oligoarthritis generally have the best outcome with 68% of patients achieving clinical remission off medication. Patients with an extended oligoarticular JIA have higher cumulative duration of active arthritis, more erosive disease, and a higher risk of chronic disability. Only 31% of children with extended oligoarticular JIA achieved remission after discontinuing their medication. Relapses occur, and within 2 years after clinical remission off medication, flares occurred in 47% of patients with persistent oligoarthritis and 67% of patients with extended oligoarthritis. Diligence regarding ophthalmologic screening for asymptomatic uveitis needs to continue as described in “Uveitis.”

UVEITIS

Clinical Manifestations and Diagnostic Features

Chronic anterior uveitis, defined as inflammation involving the anterior uveal tract including the iris and ciliary body, is the most frequent extra-articular manifestation of JIA. Uveitis is most common in oligoarticular JIA and RF-negative polyarticular JIA, with a prevalence ranging from 17% to 26% and 4% to 25%, respectively. It is rare in RF-positive polyarticular JIA and sJIA. Risk factors for uveitis include ANA positivity, being younger than 6 years of age at JIA onset, female sex, and oligoarticular subtype. These risk factors of developing uveitis are larger influences in girls but not boys. The interval from diagnosis of JIA to the development of uveitis is longer the younger the age at onset of JIA, especially in ANA-positive patients. Overall, uveitis is observed in 30% of ANA-positive patients with JIA.

Chronic anterior uveitis is typically nongranulomatous and asymptomatic at onset, and if unrecognized it can lead to serious visual deficits. Because of its insidious onset, regularly scheduled screening by an experienced ophthalmologist with slit lamp examination is required for early diagnosis and treatment. Newly diagnosed patients are ideally screened within 6 weeks of diagnosis because in 5% of cases, uveitis occurs before diagnosis of JIA. The highest risk period for uveitis development is the first 4 years after arthritis onset, although the risk is never completely eliminated. The frequency of ophthalmologic screening according to the American Academy of Pediatrics guidelines is determined by the degree of risk such as ANA status, age of JIA onset, and duration of disease shown in Table 107-3. In children with HLA-B27–related disease, anterior uveitis occurs in 10% to 15% but is usually highly symptomatic and therefore does not require routine screening. The severity of chronic anterior uveitis associated with JIA is unrelated to the severity of the underlying joint disease, and the clinical course of the uveitis and arthritis may not parallel one another. Disease is eventually bilateral in nearly two-thirds of patients, but both eyes are not always inflamed at the same time.

Treatment

Differential Diagnosis

When evaluating the etiology of uveitis, it is important to consider infectious causes, such as tuberculosis,
toxoplasmosis, cytomegalovirus, herpes simplex virus, syphilis, human immunodeficiency virus, Lyme disease, cat scratch disease, and fungus. Uveitis can be confined primarily to the eye or occur secondary to a systemic illness, such as in Behçet’s disease, sarcoidosis, autoimmune inflammatory syndromes, multiple sclerosis, TINU (tubulointerstitial nephritis and uveitis syndrome), and Vogt-Koyanagi-Harada syndrome. Although the cause of uveitis often remains idiopathic, it is important to not miss a malignancy such as lymphoma or retinoblastoma.

Uveitis is initially treated with topical corticosteroids and mydriatics, although in approximately 30% of patients the uveitis remains active even with topical or local treatment with subtenon corticosteroid injections, and immunosuppressive medications are indicated to aggressively control the inflammation and prevent poor visual outcomes. Although there are no prospective randomized controlled studies on the use of immunosuppressive medications in children with uveitis, several observational studies suggest the effectiveness of these treatments. MTX, either oral or parenteral, is usually the first choice of treatment and is also used as a steroid-sparing agent. Other effective immunosuppressives include mycophenolate mofetil and azathioprine, while cyclosporine has limited value. Treatment of uveitis requires close collaboration between the affected child’s rheumatologist and ophthalmologist.

In cases of refractory inflammation, biologic agents should be considered. TNF inhibitors are the most commonly used, specifically infliximab and adalimumab, which appear to be more effective than etanercept. The dosing range and frequency of infliximab varies widely when used for uveitis, and up to 10 mg/kg/month are often required for control and should be used with MTX to prevent development of anti-infliximab antibodies. Adalimumab is given 20 to 40 mg subcutaneously every 2 weeks, and when ineffective, it can be administered weekly. A case series reported remission with the administration of abatacept 10 mg/kg intravenously monthly in six of seven patients refractory to immunosuppressives and TNF inhibition. High-dose intravenous daclizumab, a humanized monoclonal antibody against IL-2 receptor, has been reported to be effective but is no longer available in the United States.

In the past, ophthalmologists were concerned about primary placement of intraocular lens in JIA patients with history of uveitis with subsequent formation of cataracts. Now, with the more widespread practice of strict control of uveitis, good visual outcomes with cataract surgery and intraocular lens placement can be achieved using aggressive systemic immunomodulatory therapy perioperatively. The general expert opinion among uveitis specialists is to try to taper immunomodulatory therapy after 12 to 24 months of quiescence of uveitis; however, this has not been studied.

**Outcome**

Complications resulting from uveitis include posterior synchiae, cataract, band keratopathy, glaucoma, papillitis, or cystoid macular edema. Posterior synchiae (fibrous bands adhering the iris to the lens, Figure 107-4) result in a distorted papillary border. Band keratopathy (a layer of calcium deposited in Bowman’s membrane of the cornea), is not uncommon. Cataracts and glaucoma may develop as a complication of the uveitis or its treatment, so chronic monotherapy with topical corticosteroids should be avoided.

The reported rate of visual loss due to JIA-related uveitis has decreased from 22% to 66% in studies before 1990 to 3% to 25% in newer studies, suggesting that newer and more aggressive approaches are effective. Ongoing active intraocular inflammation with greater than or equal to 0.5+ cells was associated with increased risk of visual impairment and blindness.

Prognostic factors in JIA-associated uveitis are not clearly identified, and the results of the studies are often controversial. Different studies have revealed the following as potential factors associated with poor prognosis: short intervals between the diagnosis of arthritis and uveitis, severity of uveitis at first examination, signs of anterior chamber involvement (cells and flare), and male gender.

**JUVENILE PSORIATIC ARTHRITIS**

**Clinical Manifestations and Diagnostic Features**

Juvenile psoriatic arthritis (JPsA) was initially defined as juvenile-onset arthritis associated with psoriasis occurring at some point during the disease course. This definition has been expanded to include not only patients with overt psoriatic lesions but also arthritis who only have nail abnormalities or a first-degree relative with overt psoriasis. JPsA represents 2% to 15% of all JIA with slight female preponderance and a bimodal pattern of onset (2 to 4 years and 7 to 10 years).

The psoriasis typically occurs within
2 years of the onset of arthritis, and for the majority of children the skin symptoms follow the arthritis symptoms. Up to 80% of children have a classic psoriatic vulgaris or plaque psoriasis characterized by well-demarcated erythematosus scaly lesions occurring over extensor surfaces (elbows and knees), scalp, and trunk. However, in small children younger than the age of 2, the most common finding is psoriatic diaper rash. Additional areas that should be evaluated include the hairline behind the ears, the navel, the groin region, and superior to the gluteal cleft.

Because JPsA shares similar manifestations with multiple subtypes of juvenile arthritis, there is not one specific articular presentation. Patterns of joint involvement can be similar to both oligoarticular or polyarticular JIA with 60% to 70% of JPsA patients having an oligoarticular onset (<5 joints). There were no differences in ANA or HLA-B27 positivity, or frequency of uveitis between oligoarticular JIA and JPsA with fewer than 5 joints at onset, although dactylitis was more frequent, reported in 15% to 37% of JPsA. When JPsA subjects who had a polyarticular type course were compared with polyarticular JIA subjects, there was also no difference seen in ANA positivity, HLA-B27 positivity, and dactylitis. JPsA is associated with subacute anterior uveitis and ANA positivity in 15% to 20% of children with JPsA. A recent study of childhood psoriatic arthritis argued for two distinct subpopulations: (1) a younger group (median age, 2.7 years), ANA positive, with a female preponderance, all of whom had dactylitis and more persistent disease, and (2) an older group (median age 9.5 years), more likely to be oligoarticular, have axial disease and enthesis, and have higher remission rates.

Children with psoriatic arthritis may have mild elevated acute-phase reactants (ESR, C-reactive protein [CRP], and platelets) and a mild anemia of chronic disease. However, up to one-third have no laboratory evidence of inflammation. Serologies and HLA associations are noted earlier.

**Treatment**

To date there have been no controlled studies examining the efficacy of antirheumatic medications in JPsA. A recent study compared treatment regimens of JPsA patients versus oligoarticular and polyarticular JIA patients and found no difference in use of NSAIDs, MTX, and TNF inhibitors. In general, the treatment of JPsA should follow the treatment approaches for oligoarticular or RF-negative polyarticular JIA (see Figures 107-1 and 107-2) depending on the patient’s pattern of joint involvement. MTX is of benefit for both the skin psoriasis and arthritis and, when used in children, is recommended as a single weekly dose rather than split doses. In children with more aggressive disease, TNF inhibitors (which have also been successful in treating psoriasis) are indicated and may significantly limit bony destruction. Regular screening for uveitis is required, especially in ANA-positive patients with both oligoarticular and polyarticular courses, with monitoring and management as described later in “Uveitis.”

**Outcome**

Because of the heterogeneity in the patterns of arthritis in JPsA, the outcomes are variable and tend to track with the pattern of joint involvement. Although in one study erosions were less common in JPsA compared with polyarticular JIA (23% vs. 46%, respectively), there appears to be more persistent disease activity and more evidence of physical limitations and ongoing disease activity continuing into adulthood in patients with JPsA compared with oligoarticular and polyarticular JIA patients. The uveitis of psoriatic arthritis, like that of oligoarticular disease, can lead to blindness if untreated.

**ENTHESIS-RELATED ARTHRITIS/ JUVENILE SPONDYLOARTHRITOPATHY**

**Clinical Manifestations and Diagnostic Features**

The current ILAR category of ERA is a spectrum that includes spondyloarthropathies, juvenile ankylosing spondylitis, and SEA (spondylitis enthesisis and arthritis). ERA accounts for about 20% of JIA in one U.S. study, but previous classification schemas have included JPsA within spondyloarthropathies resulting in variability in incidence and prevalence data. In some cases, patients are first diagnosed as oligoarticular JIA, but characteristics of ERA later evolve. ERA occurs after age 6 and is more common in boys (male-to-female ratio of 7:1), although it may be under-recognized in symptomatic girls who can have milder disease with less axial skeleton involvement.

The typical feature of ERA is the presence of enthesis, or inflammation of the tendons and ligaments where they attach to bone (or entheses). It is an early manifestation and occurs more frequently in children than in adult-onset ankylosing spondylitis (AS). The typical enthesitis sites are around the patella (at the 2, 6, and 10 o’clock positions); Achilles’ tendon; plantar fascia insertions into the calcaneus and metatarsal heads, greater trochanter, tibial tuberosity, and the base of the fifth metatarsal. Children may report vague buttck pain, groin pain, or heel pain, and a classic finding is of tarsitis with inflammation of the subtalar joint and surrounding tendon sheaths. At onset spinal symptoms are rare, but a subgroup of children with ERA will progress to features more typical of adult AS with SI joint and spinal inflammation during adolescence. This progression is more likely in boys who are HLA-B27 positive and have spinal or sacroiliac pain within 1 year of diagnosis. The modified Schober’s test can be used to evaluate lower lumbar flexibility with a change less than 6 cm being abnormal. Because thoracic excursion varies greatly in a growing child, only performing sequential measurements are helpful. In the adolescent, any thoracic excursion less than 5 cm should be regarded as abnormal.

ERA is associated with an acute anterior uveitis (AAU) in 6% to 27% of children, which typically presents as an acutely red, painful eye that needs immediate medical attention to avoid blindness. Cardiovascular disease, although uncommon, can be severe in patients with ERA that has evolved into juvenile ankylosing spondylitis, with inflammatory aortic regurgitation reported in up to 10% of patients. Restrictive pulmonary disease has been seen in juvenile spondyloarthritis without clinical or radiologic findings.

Reactive arthritis is categorized under ERA, following infection with organisms from the gastrointestinal tract.
(e.g., Salmonella, Shigella, Yersinia, Campylobacter, Clostridium difficile) or genitourinary tract (i.e., Chlamydia or Ureaplasma) without actual infection in the joint tissues. Usually self-limited, the arthritis can persist and transform into a more chronic arthropathy. In addition to the arthritis, uveitis and enthesitis have been documented in children to complete the classic triad for reactive arthritis. The clinical infection usually precedes the arthritis, enthesitis, or extra-articular disease by 1 to 4 weeks. Arthritis associated with IBD (Crohn’s disease or ulcerative colitis) is seen in children, and two distinct patterns of joint involvement are noted: (1) peripheral arthritis and (2) sacroiliitis and spondylitis, with the former more common than the latter. Additional clinical manifestations unique to arthritis associated with IBD include clubbing, periostitis, erythema nodosum, pyoderma gangrenosum, osteoporosis, and rarely hypertrophic osteoarthropathy. The peripheral arthritis activity has been shown to correlate with gut disease activity unlike the SI joint activity, which tends not to correlate.

Compared with adults, in whom radiologic evidence of sacroilitis is the diagnostic hallmark of ankylosing spondylitis, radiologic evaluation in children with ERA is less often diagnostic because sacroilitis is rare as a presenting symptom. Magnetic resonance imaging (MRI) with contrast is being increasingly used to detect evidence of acute sacroilitis without chronic changes in children. MRI has also been used to evaluate enthesitis. HLA-B27 is present in 80% to 90% of cases depending on ethnicity. The ESR may be mildly or markedly raised, and there may be a mild anemia, but these should also raise the suspicion that the child may have subclinical IBD. RF is negative, and ANA may be positive.

### Differential Diagnosis

Children with true infectious arthropides (viral, rheumatic fever, poststreptococcal, Lyme) can present with similar manifestations as ERA patients. More benign entities such as toxic synovitis and benign limb pains of childhood (growing pains), as well as more concerning entities such as malignancies and solid tumors, should be considered. Orthopedic diagnoses such as Legg-Calvé-Perthes disease, slipped capital femoral epiphysis, and Osgood-Schlatter disease and less common rheumatologic diseases such as SAPHO (synovitis, acne, pustulosis, hyperostosis, and osteomyelitis), Kawasaki’s syndrome, and vasculitis can share similar articular and extra-articular manifestations with ERA. Lastly, children with widespread amplified musculoskeletal pain may have tender entheses that can be mistaken for enthesitis.

### Treatment

Treatment of ERA ranges from NSAIDs to DMARDs with the use of biologics in the most aggressive cases. To date there have been few randomized controlled trials in this population. One study of sulfasalazine showed no significant effects in juvenile spondyloarthopathy, whereas another showed good response in oligoarticular- and polyarticular-onset JIA patients. Most respond well to intra-articular corticosteroid injections, but many may need a DMARD. For severe symptoms or evidence of potential joint damage (such as erosive sacroilitis), the use of TNF inhibitors has been effective on the basis of small pediatric studies.

### Outcome

Long-term outcome of ERA is unknown, but a proportion of these children may progress to the adult form of AS. Predictors of sacroilitis were HLA-B27 positivity, absence of DPB1*02, hip joint involvement within the first 6 months, and disease onset after age 8 years. When compared with adult-onset ankyllosing spondylitis, juvenile-onset ankylosing spondylitis had less severe axial involvement but worse hip involvement, and in some studies, worse functional impairment. The probability of remission remains low with remission rates reported ranging from 17% to 44%. Predictors for failure to achieve remission include family history of AS in first-degree relative, female sex, younger age at disease onset, arthritis in ankle joint within 6 months of disease onset, and HLA-DRB1*08. Minden and colleagues showed that if ERA persisted for more than 5 years, the chance of remission was only 17%.

### Systemic Juvenile Idiopathic Arthritis

**Clinical Manifestations and Diagnostic Features**

sJIA, which makes up 10% of all JIA, has unique characteristics that include fever, specific rash, and significantly elevated inflammatory markers in addition to arthritis. Recent research has identified biologic differences between sJIA and the other subcategories including prominent involvement of components of the innate immune system (in particular, inflammatory cytokines IL-1, IL-6, IL-18, neutrophils, and monocytes/macrophages), suggesting that sJIA may be in the autoinflammatory disease spectrum.

sJIA occurs at any age with a tendency toward younger than 5 years peaking around 2 years of age. Rarely it occurs in adults and is called adult-onset Still’s disease. The gender predilection has stayed neutral, and some studies have suggested that sJIA may be relatively more frequent in Asian countries. The systemic features of this disease are striking and are always present at onset, often predominating the clinical presentation. The arthritis may not be clinically present at onset, although arthralgias and myalgias are almost universally present at onset. The fever is typically spiking in character with a peak of at least 39°C daily, occurring once or twice a day. Interestingly, the temperatures can become subnormal between fevers. The child is usually unwell and irritable during the fever but often recovers in between. In 80% of patients, the fevers are accompanied by an evanescent, migratory, salmon pink, and sometimes urticarial macular rash (Figure 107-5). The rash typically spares the face and can occur without specific pattern on any part of the trunk or extremities. It can be subtle and scattered or diffuse and almost confluent. A helpful diagnostic feature is that the rash can be elicited by the Koebner phenomenon in which rubbing or scratching the skin elicits the rash.

Other systemic features include lymphadenopathy, hepatosplenomegaly, searsitis (pleural/pericardial involvement/
abdominal pain), headaches, and sore throat. sJIA has the most common and severe cardiac presentations of all the subtypes of JIA. Pericarditis and pericardial effusions are the most common organ system manifestations, occurring in up to 10% of the presentations. A serious life-threatening systemic manifestation called macrophage activation syndrome (MAS) is discussed in more detail later. The arthritis in sJIA can be varied from minimal to oligoarticular to polyarticular presentation and typically does not correlate with the severity of the systemic manifestations. Tenosynovitis and synovial cyst formation can also be seen. Although the systemic manifestations can be serious initially, the long-term morbidity is due to articular disease and adverse effects of medications, especially chronic corticosteroids.

There are no diagnostic tests for sJIA, but there are characteristic patterns of laboratory abnormalities including high inflammatory markers (CRP, ESR), significant leukocytosis with neutrophilia and bandemia, thrombocytosis, and anemia. Liver transaminases, aldolase, ferritin, fibrin split products, and coagulation screen may be abnormal in severe cases and can be signs of early or impending macrophage activation syndrome (MAS). ANA or other autoantibodies are rarely present. Inflammatory cytokine gene expression profiles have been shown to distinguish sJIA from other febrile inflammatory diseases seen in children and may lead to more specific tests in the future.

Differential Diagnosis

Due to the nonspecific nature of the characteristics of sJIA, the diagnosis can be difficult and should be approached as a diagnosis of exclusion. Infections, Kawasaki’s syndrome, malignancy, and other autoimmune diseases can present with similar symptoms; therefore it is necessary to screen for infectious agents and neoplasia, especially leukemia and neuroblastoma with appropriate tests including cultures, bone marrow aspiration and biopsy, and urinary vanillylmandelic acids. Some physicians do these tests routinely because malignancies are often close mimics to the early stages of sJIA. The recurrent fever syndromes are often mistaken for sJIA, but the character of the fevers and the fixed rashes associated with these syndromes should alert the clinician to a different diagnosis. Other childhood rheumatic diseases should be considered such as systemic lupus erythematosus, Behçet’s syndrome, and others.

Special Considerations: Macrophage Activation Syndrome

About 10% of the sJIA patients will develop overt life-threatening MAS, and up to 30% will develop a milder form, which if inadequately treated could lead to full-blown MAS. MAS is a form of secondary or acquired hemophagocytic lymphohistiocytosis (HLH) seen within rheumatic disease. Two-thirds of the mortality seen in all patients with JIA are due to this entity. The main manifestations include unrelenting high fevers, hepatosplenomegaly, lymphadenopathy, severe cytopenias, liver dysfunction, central nervous system (CNS) involvement (seizures/coma), and coagulopathy (Table 107-4). The ferritin often exceeds 10,000 ng/mL, and coagulopathy can be impressive with elevated prothrombin and partial thrombin times, hypofibrinogenemia, petechiae, mucosal bleeding, epistaxis,
and hematemesis. Often MAS is heralded by a decrease of the ESR, leukocyte and platelet counts, and liver dysfunction. Histologically, patients with MAS show expansion of well-differentiated macrophages exhibiting hemophagocytosis in the bone marrow, lymph nodes, and other organs such as the liver and lungs. There is no gender, age, or race predilection for MAS in sJIA. It often occurs during active systemic disease but has also been seen in the quiescence phase of the disease (no clinical symptoms but still on medications). Triggers for MAS in sJIA include bacterial, fungal, and parasitic infections. Epstein-Barr virus, varicella, coxsackie, parvovirus B19, hepatitis A, Salmonella, and Pneumocystis infections have been implicated. Drugs have also been implicated including aspirin, NSAIDs, sulfasalazine, MTX, etanercept, anakinra, and gold salts. One must be cautious to quickly lay blame because many patients who are receiving these medications have active disease and may have been developing MAS despite the medications. Most of the time, the trigger is unknown. Poor natural killer cell cytolytic activity leading to low levels of perforin expression has been reported in sJIA, a unique finding compared with the other subtypes of JIA. Soluble IL-2Rα receptors and soluble CD163 are increased severalfold in MAS.

**Treatment**

sJIA is an active area of research with increasing understanding of the autoinflammatory biology and the potential therapeutic roles for targeted biologics. Therapies such as MTX and TNF inhibition are less helpful in sJIA compared with other forms of JIA. An approach to treating sJIA is shown in Figure 107-6, separating this complex disease into different severities and manifestations. This algorithm is consistent with the recently published ACR guidelines for JIA. Mild sJIA, defined as mild systemic symptoms without organ system involvement and synovitis, can be successfully treated with anti-inflammatory doses of NSAIDs. Certain NSAIDs are helpful for different aspects of the disease (e.g., indomethacin can be used for fever control and serositis symptoms). In addition, if there are only a few large joints involved, intra-articular steroid injections are an additional option.

In the more severe cases with persistent fevers, cardiopulmonary symptoms, significant anemia, and significantly elevated inflammatory markers, corticosteroids may be used, often given as pulsed intravenous high-dose methylprednisolone (30 mg/kg/dose) daily for 3 days in a row followed by tapering doses of oral prednisolone. Increasingly, as shown in Figure 107-6, IL-1 or IL-6 inhibition is instituted to avoid steroid toxicity and in some recent studies is being used as initial therapy before corticosteroids. Use of DMARDs and TNF inhibitors may be more helpful for more significant articular disease. TNF inhibition has been shown to be effective in some patients with sJIA, though to a lesser extent than in other types of JIA. Similarly, MTX may be helpful in sJIA with prominence of articular symptoms over systemic features. Cyclosporine as

---

**Figure 107-6** Systemic juvenile idiopathic arthritis treatment algorithm. The treatment goal is remission of disease activity, both systemic and articular, and is stratified by severity of disease. Algorithm is divided into severe systemic disease manifestations (macrophage activation syndrome [MAS], serositis) or synovitis with milder systemic disease. Currently there is significant variability in practice regarding using corticosteroid as initial systemic therapy or moving directly to inflammatory cytokine inhibitors. At the time of this writing, interleukin (IL)-1 inhibition and IL-6 inhibition are currently in trials, and more information is likely to be available in the future. CSA, cyclosporine A; CYC, cyclophosphamide; IAS, intra-articular steroid; MP, methylprednisolone; NSAID, nonsteroidal anti-inflammatory drug; Pred., prednisone; TNF, tumor necrosis factor.
remission. Early experience demonstrated cases of induction with its potential serious bone marrow suppression. The newer biologics that block IL-6 and IL-1 signaling are promising according to the recent elucidations in the pathophysiology of sJIA, as well as recent clinical trials and studies. Initial reports on IL-1 blockade using anakinra showed rapid and sustained remission within a few days in sJIA patients with chronic disease activity resistant to conventional therapy.142 However, more recently reports have called into question the sustainability of the response using these medications.143,144 Newer IL-1 inhibitors such as rilonacept (currently being studied in a clinical trial, NCT00534495) and anti-IL-1β monoclonal antibody (canakinumab) may shed more light on the effectiveness of these therapies and the biology of the responses. Two phase III placebo-controlled, double-blind trials using the IL-6 receptor antagonist tocilizumab showed significant improvement of ACR Pedi 30, 50, and 70 responses.145,146 In the TENDER trial, the tocilizumab dose for patients less than 30 kg was 12 mg/kg/dose and for greater than 30 kg was 8 mg/kg/dose every 2 weeks.145 Thalidomide has also been reported to improve symptoms in refractory sJIA patients.147 The teratogenic potential of thalidomide and its newer analogue lenalidomide limits the usefulness of these medications.148

Abatacept has been used in sJIA patients who have failed previous treatments. In the open-label portion of a double-blind placebo-controlled trial of abatacept therapy in JIA patients, 24 of 37 (65%) of sJIA patients had an ACR Pedi 30 response. This was comparable with the other JIA subtypes evaluated in that study.18 The 2010 ACR guidelines for JIA treatment divided sJIA treatment into two groups: active systemic features and active arthritis. But the guidelines did not include patients with MAS, implying MAS, or life-threatening manifestations (e.g., cardiac tamponade). The algorithm shown in Figure 107-6 incorporates some of the principles from the guidelines, expanding the area’s severe systemic disease and MAS. One difference is that the guideline supports earlier use of methotrexate and TNF inhibition in patients with active sJIA with primarily articular manifestations.

The optimal treatment for MAS continues to be controversial. Therapy should be aggressive and usually starts with high-dose steroids and disease-modifying agents such as cyclosporine and IL-1 blockade.149,150 Almost 50% of patients in one study showed response to steroids alone.151 A consensus treatment protocol for HLH has been published and uses a combination of VP-16 (etoposide), dexamethasone, with or without intrathecal methotrexate, followed by a maintenance with cyclosporine A.151 This protocol or parts of it have been used with MAS in sJIA patients with success.152 However, the new use of the powerful IL-1 and IL-6 inhibitors may preclude the need for the HLH approach with its potential serious bone marrow suppression.151

For the severe refractory cases, stem cell transplantation has been done. A recent review of the experience showed favorable response with more than 50% sustained drug-free remission. Early experience demonstrated cases of induction of MAS during transplant conditioning and immediately post-transplant. This early morbidity has improved with less intense conditioning regimen.153 A follow-up study reported that late relapses were noted with lower percentages for drug-free long-term outcome. The late relapses were often less severe and treated successfully with conventional drugs.154-156

**Outcome**

sJIA is heterogeneous in severity, disease course, and outcome. Studies of the natural history of sJIA before biologic therapies demonstrate that the course can be monoinflamatory, with remission within 2 to 4 years; relapsing, characterized by flares of systemic features with mild arthritis; or continuous with persistent destructive arthritis, often more prominent after the regression of systemic features.125,157 Patients with severe disease can have flares of extra-articular features at any time and may have active arthritis into adult life despite standard therapies.147 Emerging evidence suggests that IL-1 and IL-6 inhibitors have permitted significant disease control with no or much lower doses of corticosteroids, which were one of the major causes of morbidity in sJIA in the past.

Predictors of poor articular outcome in sJIA include the systemic features 6 months after onset, thrombocytosis, and radiographs. Unfortunately, effectiveness of these DMARDS is much less than in other subtypes of JIA with cyclosporine less than MTX.

**IMAGING**

Imaging is important in JIA to confirm the diagnosis; to exclude other diseases such as infection, malignancy, osteoid osteoma, or avascular necrosis; and to monitor therapy. Determining the presence of cartilage loss and erosions are complicated by the anatomically changing joints during the normal growth process. Because children have a large amount of cartilage in their joints, a significant amount of destruction can occur before erosions are identified on plain radiographs (Figure 107-7). Local growth disturbances can also be seen in children including bony overgrowth, particularly in the small joints of the hands and feet and in the knees, likely due to chronic hyperemia. Bony hypertrophy around the knee joints can lead to leg-length discrepancies,
MRI is the only technique available to visualize bone marrow edema, which has been shown to be an important predictor of erosive joint damage in RA patients, but the connection is not established in JIA patients. For evaluating bony erosions, MRI was found to be the most sensitive imaging modality in wrists of JIA patients, revealing more than twice as many erosions compared with plain radiography and US. Pilot MRI grading scores of hip and knee disease activity and damage in JIA have been proposed but not validated. A recent systematic review of the literature assessed that the overall quality of the reporting of methods in studies on the MRI assessment of JIA is heterogeneous and fair overall.

**SPECIAL CONSIDERATIONS: REHABILITATION IN CHILDREN**

PT and OT play a crucial role in JIA patients. PT and OT can consult for stretches, range-of-motion evaluation, joint protection, muscle building and strengthening, as well as endurance and graduated aerobic activity. Children who present late in the course of oligoarthritis may already have flexion contractures, which require splinting and serial casting. Serial casting is done two or three times a week for up to a month if necessary and is probably most effective when started just after joint injections. Some children with marked leg-length discrepancy (resulting from overgrowth of the affected knee) may require a shoe lift/raise. Strict limitations from physical education at school and sports are generally not encouraged because physical activity plays an important role in rehabilitation and therapy; however, impact sports should be limited in the presence of joint effusions or when there is a concern for joint instability (e.g., C1/C2 atlantoaxial subluxation). A modified physical education program and sports activities to the patients’ own tolerance are recommended instead. Orthotics can be helpful in foot involvement in ERA, as well as other types of JIA.

Although there is a wide within-patient and between-patient variability, children with early-disease onset and a greater number of restricted joints have the highest risk of developing long-term functional physical disability. In addition, health-related quality of life was significantly lower in patients with JIA than in healthy children, and patients with persistent oligoarthritis were less severely affected compared with the other JIA subcategories. Total energy and activity-related energy expenditure, as well as physical activity levels, were significantly lower in JIA patients compared with controls, and only 23% of the JIA patients met the public health guidelines on physical activity compared with 66% in controls.

**SUMMARY**

The modern treatment of JIA has significantly decreased the long-term burden of disease but requires early recognition and diagnosis. It also requires aggressive treatment to extinguish disease activity and multidisciplinary approaches to maintain and improve function and quality of life. Pediatric rheumatologists have in their favor the remarkable capacity for childhood growth and development to allow repair and restoration of function, in contrast to adults with permanent gait disturbances, and secondary scoliosis (see Figure 107-3). In contrast, premature fusion of the epiphyses can lead to shortening of certain joints, most notably the TMJs, leading to micrognathia. Interestingly, the articular cartilage of growing children has been shown to have unique regenerative qualities and studies have shown that children with JIA can have improvement in their radiographic joint damage.

Plain radiographs in patients with JIA show soft tissue swelling as an early but nonspecific finding. The most common form of radiographic damage in JIA is joint space narrowing due to erosion, thinning, and loss of articular cartilage. Periarticular osteoporosis due to hyperemia of inflamed joints is frequently seen. Erosions and ankylosis can be seen as late findings in radiographs. Lastly, periostitis along the shafts of the phalanges, metacarpals, and metatarsals and calcifications from steroid joint injections can be seen. The trend toward early aggressive treatment to prevent erosive disease shifts the imaging need away from plain radiographs and toward other imaging modalities that are more sensitive in detecting early disease activity.

Therefore MRI and ultrasound (US) are becoming more popular.

US is reliable, safe, and relatively inexpensive but not as well standardized in children as in adults. It has been shown to be more sensitive than plain radiography in the detection of effusion, synovial thickening, and synovial cysts. In more experienced hands, cartilage thinning and bone erosions can also be seen. Inflammatory involvement of the hip, shoulder, and elbow is more frequently accurately detected by US compared with clinical examination. Color and power Doppler US can facilitate the evaluation of hyperemia and vascular abnormalities in affected joints and tendon involvement.

MRI is the most sensitive imaging modality currently available for detecting synovial inflammation; however, its use is limited by expense and the need for anesthesia in young children. With the use of contrast, accurate differentiation between active and inactive states can be established. Synovial enhancement and thickening can easily be seen, and pannus volume measurements can be tracked. MRI is the only technique available to visualize bone

---

**Figure 107-7** Plain radiograph of a young child with active polyarticular juvenile idiopathic arthritis, demonstrating the challenge in evaluating joint damage in the growing child. Periarticular osteoporosis and early periostitis of the metacarpals are noted.
inflammatory arthritis. The advent of new biologic therapies and rapid translation of basic research into therapeutic strategies and an increased willingness on the part of regulatory bodies to make new therapies available to children should combine to continue to improve the outlook for children with arthritis.

Selected References


References


CHAPTER 107  |  TREATMENT OF JUVENILE IDIOPATHIC ARTHRITIS 1770e8


146. Ravielli A: The time has come to include assessment of radiographic progression in juvenile idiopathic arthritis clinical trials, J Rheumatol 35(4):553–557, 2008.


Pediatric Systemic Lupus Erythematosus, Dermatomyositis, Scleroderma, and Vasculitis

RONALD M. LAXER • SUSANNE M. BENSELER

understanding of these diseases and enabled us to recognize novel inflammatory diseases and their mimics.

SYSTEMIC LUPUS ERYTHEMATOSUS

Definition and Classification

Pediatric SLE (pSLE) is a chronic multisystem autoimmune disease with remitting, relapsing course and onset of symptoms before age 18 years, accounting for approximately 20% of all SLE. This clinically heterogeneous disease is characterized by a distinct spectrum of autoantibodies including antinuclear antibody (ANA), double-stranded deoxyribonucleic acid (dsDNA), and antibodies against extractable nuclear antigens (ENAs). In genetically susceptible hosts B cell–mediated autoimmune processes lead to a variable combination and severity of clinical symptoms including antibody-mediated vasculitis, direct antibody binding to target cells, and thrombotic organ dysfunction.

The American College of Rheumatology (ACR) classification criteria for adults with SLE are commonly applied to children with pSLE. The classification of neuropsychiatric SLE (NPSLE) in children and adolescents remains a challenge. The 1990 ACR NPSLE nomenclature and case definitions appear to have limited applicability for children in certain domains.

Epidemiology

Pediatric SLE affects children and adolescents around the world. On average, 60% of patients develop pSLE after age 10, 35% between 5 and 10 years, and only 5% before age 5. In studies from Asia, the mean ages at diagnosis were reported to be 8.6 to 13.5 years. The incidence and prevalence of pSLE varies between populations. Similar to adults, SLE more commonly affects non-Caucasian populations. Overall, incidence rates of pSLE have been reported to be 0.28 to 0.48 per 100,000 children with prevalence rates of 6.3 to 24 per 100,000 depending on the ethnic background of the population. The prevalence of pSLE is consistently higher in girls than boys: Canadian and Taiwan studies estimated a factor of 6. Female predominance increases with age. This observation strongly supports the suggested role of female hormones. Although pSLE is less common than adult SLE, childhood onset was recently...
found to be a strong, independent predictor of overall lupus mortality.\textsuperscript{15}

Children with pSLE are likely to have relatives with SLE. The pattern of familial aggregation for siblings with SLE suggests a polygenic inheritance.\textsuperscript{16} In addition, pSLE patients often have asymptomatic relatives with evidence of autoantibodies.\textsuperscript{17}

**Causes**

**Genes**

Enabled by large international collaborations, susceptibility genes have been identified suggesting a dysregulated immune phenotype in lupus patients partially overlapping with other autoimmune diseases.\textsuperscript{18} These include genetic variants in the cytokine interferon-\(\alpha\) pathway and their functional impact,\textsuperscript{19} the contribution of signal transduction \(\text{STAT4}\) gene variations on lupus susceptibility,\textsuperscript{20} and the association of the interleukin-1 receptor-associated kinase-1 (\(\text{IRAK1}\)), an X chromosome gene, and disease susceptibility in pSLE.\textsuperscript{21}

Beyond the type of gene variant (mutation, polymorphisms), altered copy number variations reflecting “gene dose” and epigenetic modifications of key lupus genes are reported in pediatric lupus patients: Garcia-Ortiz demonstrated an association of the lupus gene Toll-like receptor 7 (\(\text{TLR7}\)) copy number variation with susceptibility to pSLE in Mexican populations.\textsuperscript{22} Drug-induced SLE and incomplete monogygotic twin concordance rates suggest an important role of epigenetic factors such as histone modification or altered DNA methylation in pSLE.\textsuperscript{23} Finally, mitochondrial DNA polymorphisms may also be important in the pathogenesis of SLE.\textsuperscript{24} Children with inherited complement deficiencies including C2, C4A/C4B, C1q, and C1s can present with pSLE.\textsuperscript{25} These single gene defects are more commonly seen in familial cases of lupus.

**Environment**

Beyond genes and epigenetics, other potential contributing factors have been intensely studied in pSLE including environmental factors such as parental smoking and organic dust exposure.\textsuperscript{27} As a link between genetic and environmental factors, endogenous and external viruses are currently being studied in pSLE.\textsuperscript{28} The presence of an interferon signature pattern had raised suspicion for a potential viral contribution and discovery of the genetic basis of chilblain lupus.\textsuperscript{29} This focused the researchers’ attention to DNA repair of endogenous virus as a potential pathogenetic factor in pSLE.

**Pathology**

Similar to adult SLE, the inflammatory processes leading to organ dysfunction are heterogeneous between and within organs (see Chapter 79). Immunoglobulin deposition in small vessels such as glomerulus or lung capillaries, complement activation, antibody-binding to single cells, and microthrombotic or macrothrombotic vessel disease are the hallmarks of lupus. Hematologic manifestations of lupus are often related to direct antibody-binding and complement activation.\textsuperscript{30} In contrast, the histology of pediatric lupus nephritis is more variable and includes the distinct subtypes of mesangial, focal, or diffuse proliferative and membranous nephritis, which can coexist.\textsuperscript{31} Renal biopsies are required to distinguish subtypes and define treatment regimens. In children, severity of the glomerulonephritis on renal biopsy was shown to be associated with treatment choice and response and long-term outcome.\textsuperscript{32-34} Confounding risk factors for severe lupus nephritis and adverse renal outcome include evidence of thrombosis microangiopathy, anti-phospholipid antibodies, tubulointerstitial disease, hypertension, nephrotic syndrome, and access to health care.\textsuperscript{35}

**Clinical Features**

The clinical presentation of pSLE was reported in large series from many countries, allowing a better understanding of the clinical diversity, impact of ethnicity, confounding factors including infections, and access to health care (Table 108-1).

Systemic features including fever and fatigue are found in more than 90% of children at diagnosis of pSLE. Arthritis is the most frequently reported organ manifestation in pSLE. Typically pSLE arthritis is a nonerosive, painful polyarthritis. Mucocutaneous involvement includes the typical “butterfly” or malar rash, an erythematous rash in the malar distribution sparing the nasolabial folds (Figure 108-1). Diffuse hair loss is commonly seen in children with active disease. In addition, children with lupus can present with a photosensitive rash, exacerbated in sun-exposed areas including upper arms, neck, and face; Raynaud’s phenomenon of fingers and toes; a vasculitic skin rash, which is often raised and painful affecting the fingers and toes; and oral and/or nasal ulcers. Oral ulcers are typically located on the hard palate and are painless (Figure 108-2). Uncommon skin manifestations include discoid lupus lesions, which heal with scarring.\textsuperscript{48}

Nephritis is the most common major organ manifestation and occurs in more than 50% of children with pSLE, most commonly at diagnosis. Children with lupus nephritis present with peripheral edema, proteinuria, active urine sediment, and hypertension. Focal or diffuse proliferative
Table 108-1  Clinical Characteristics of Children at Diagnosis of Pediatric Systemic Lupus Erythematosus (pSLE): Comparison of Four Recent Cohorts

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of pediatric SLE patients</td>
<td>213</td>
<td>70</td>
<td>258</td>
<td>230</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>11.6 (±2.6)</td>
<td>10.5 (4-15)</td>
<td>13.1 (±3.2)</td>
<td>15.3 (13.2-16.7)</td>
</tr>
<tr>
<td>Female-to-male ratio</td>
<td>4.2:1</td>
<td>6:1</td>
<td>4.7:1</td>
<td>9:01</td>
</tr>
<tr>
<td>Median follow-up</td>
<td>3.6 yr</td>
<td>NR</td>
<td>3.5 yr</td>
<td>1.7 yr</td>
</tr>
<tr>
<td>Generalized symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatigue</td>
<td>NR</td>
<td>NR</td>
<td>50%</td>
<td>NR</td>
</tr>
<tr>
<td>Fever</td>
<td>NR</td>
<td>92%</td>
<td>39%</td>
<td>NR</td>
</tr>
<tr>
<td>Weight loss</td>
<td>NR</td>
<td>30%</td>
<td>29%</td>
<td>63%</td>
</tr>
<tr>
<td>Lymphadenopathy/hepatosplenomegaly</td>
<td>NR</td>
<td>42%/47%</td>
<td>19%</td>
<td>NR</td>
</tr>
<tr>
<td>Organ manifestations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hematologic disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thrombocytopenia</td>
<td>NR</td>
<td>24%</td>
<td>29%</td>
<td>25%</td>
</tr>
<tr>
<td>Lymphopenia</td>
<td>NR</td>
<td>NR</td>
<td>29%</td>
<td>60%</td>
</tr>
<tr>
<td>Coombs-positive hemolytic anemia</td>
<td>NR</td>
<td>58%</td>
<td>23%</td>
<td>16%</td>
</tr>
<tr>
<td>Mucocutaneous disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malar rash</td>
<td>NR</td>
<td>57%</td>
<td>61%</td>
<td>70%</td>
</tr>
<tr>
<td>Photosensitivity</td>
<td>NR</td>
<td>51%</td>
<td>17%</td>
<td>53%</td>
</tr>
<tr>
<td>Oral ulcers</td>
<td>NR</td>
<td>NR</td>
<td>21%</td>
<td>49%</td>
</tr>
<tr>
<td>Nasal ulcers</td>
<td>NR</td>
<td>NR</td>
<td>8%</td>
<td>NR</td>
</tr>
<tr>
<td>Arthritis</td>
<td>NR</td>
<td>66%</td>
<td>61%</td>
<td>83%</td>
</tr>
<tr>
<td>Nephritis</td>
<td>81% total</td>
<td>77% total</td>
<td>37% total</td>
<td>49% total</td>
</tr>
<tr>
<td>Mesangial—WHO Class II</td>
<td>47%</td>
<td>NR</td>
<td>15%</td>
<td>NR</td>
</tr>
<tr>
<td>Focal proliferative—WHO Class III</td>
<td>2%</td>
<td>NR</td>
<td>28%</td>
<td>NR</td>
</tr>
<tr>
<td>Diffuse proliferative—WHO Class IV</td>
<td>37%</td>
<td>NR</td>
<td>47%</td>
<td>NR</td>
</tr>
<tr>
<td>Membranous—WHO Class V</td>
<td>8%</td>
<td>NR</td>
<td>16%</td>
<td>NR</td>
</tr>
<tr>
<td>Neuropsychiatric disease</td>
<td>NR</td>
<td>21%</td>
<td>16%</td>
<td>NR</td>
</tr>
<tr>
<td>Serositis</td>
<td>NR</td>
<td>3%</td>
<td>12%</td>
<td>17%</td>
</tr>
<tr>
<td>Pericarditis</td>
<td>NR</td>
<td>3%</td>
<td>12%</td>
<td>17%</td>
</tr>
<tr>
<td>Pleuritis</td>
<td>NR</td>
<td>NR</td>
<td>1%</td>
<td>NR</td>
</tr>
<tr>
<td>Cardiac disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myocarditis</td>
<td>NR</td>
<td>NR</td>
<td>0%</td>
<td>NR</td>
</tr>
<tr>
<td>Endocarditis</td>
<td>NR</td>
<td>NR</td>
<td>0.4%</td>
<td>NR</td>
</tr>
<tr>
<td>Pneumonitis</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Gastrointestinal disease</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

NR, not reported; WHO, World Health Organization.

glomerulonephritis accounts for more than 50% of all pediatric lupus nephritis in most series. Acute renal failure can occur in a third of children presenting with proliferative lupus nephritis. Mesangial and membranous nephritis are less common and can occur in conjunction with proliferative lupus nephritis. Posterior reversible encephalopathy (PRES) is an increasingly recognized central nervous system (CNS) complication of lupus nephritis and hypertension.49 Neuropsychiatric disease affects about a quarter of children with pSLE. Headaches in conjunction with psychosis or cerebrovascular disease presenting as seizures or severe cognitive dysfunction are the most common clinical phenotypes. Isolated lupus headache is uncommon.7 Psychosis in pSLE is characterized by optic and acoustic hallucinations and visual distortions. Many patients have overlapping features of cognitive dysfunction, headaches, and mood disorder.7 Isolated mood disorders such as depression are uncommon in pSLE. CNS vasculitis in pSLE more commonly affects the small vessels. Angiography-positive disease and strokes are uncommon.7 Transverse myelitis is an uncommon, serious manifestation of pSLE.50 Peripheral neuropathies are uncommon in children with lupus.

The diagnosis of NPSLE in children is based on clinical assessment including comprehensive neurocognitive testing, inflammatory markers, and neuroimaging. Children with neuropsychiatric disease commonly have
antiphospholipid antibodies, in particular anti-β₂-glycoprotein I (anti-β₂GPI). A positive lupus anticoagulant is often detectable in children with cerebrovascular disease including sinus vein thrombosis (SVT) and chorea. Neuroimaging demonstrates only subtle abnormalities in half of patients including the majority of children presenting with psychosis.

Hematologic disease is common in pSLE and affects a quarter of children. Treatment-refractory idiopathic thrombocytopenic purpura (ITP) or severe autoimmune hemolytic anemia can be the presenting features of pSLE. A positive ANA and older age were found to be risk factors of pSLE in ITP. Serositis including pleuritis, pericarditis, and less commonly peritonitis affects about 20% of children with pSLE. Chest pain is the most common presenting feature. Inflammatory lung lesions including capillaritis and alveolar hemorrhage are uncommon. Although cardiomegaly due to pericarditis and arrhythmia/conduction anomaly occur frequently, myocarditis and coronary arteritis are serious yet uncommon lupus features in children. The most common gastrointestinal (GI) manifestation of pSLE is lupus hepatitis. Inflammatory bowel manifestations are rare. Overall atypical presentations can be found in up to a quarter of children ultimately diagnosed with pSLE and, when present, were found to correlate with poor outcome. Endocrinopathies of pSLE include hypothyroidism or hyperthyroidism and diabetes mellitus. Menstrual cycle disturbances and transient amenorrhea are common in girls with pSLE and may be associated with pituitary dysfunction or treatment with cyclophosphamide, leading to a decreased progesterone production.

**Diagnosis and Diagnostic Tests**

The diagnosis of pSLE is based on clinical findings, laboratory test results including inflammatory markers, complement levels, markers of organ involvement, and specific autoantibodies. Tissue biopsies and imaging studies can further support and/or classify pSLE subtypes. The presence of 4 of 11 ACR classification criteria was found to have a sensitivity of 96% and a specificity of 100%. A careful evaluation of potential organ involvement of pSLE is mandatory. Characteristic abnormal laboratory markers of active pSLE may include a raised erythrocyte sedimentation rate (ESR), anemia, which may be Coombs positive and hemolytic, a low white blood count with predominant lymphopenia, low platelets, and low C3 and/or C4 complement levels. Paradoxically, the C-reactive protein (CRP) is normal in the vast majority of children with active lupus, except for those presenting with serositis or concurrent infections. An abnormal urinalysis including microscopic evidence of casts indicated renal involvement. Renal function impairment is best evaluated by serum creatinine, albumin, and urine protein-to-creatinine ratio. Lupus anticoagulant and specific lupus autoantibody testing including ANA, dsDNA, ENA, and antiphospholipid antibodies is mandatory. ANA is found is almost every child with pSLE, while dsDNA is detected in more than 80%. Novel antibodies have been proposed and require prospective validation in pSLE. Children may have frank hypothyroidism or hyperthyroidism or solely raised titers of thyroid antibodies.

Lipid abnormalities are increasingly recognized in children with pSLE. Dyslipidemia may contribute to a risk of premature arteriosclerosis in pSLE patients. Cytokines derived from adipocytes including leptin, adiponectin, and ghrelin were recently shown to correlate with disease activity in pSLE. Markers of bone health and fracture risk have been extensively evaluated in children with pSLE. A comprehensive assessment may include these markers in conjunction with bone density measurement.

The differential diagnosis of pSLE is wide and includes infections such as cytomegalovirus (CMV), Epstein-Barr virus, and tuberculosis, malignancies such as lymphomas, endocrinopathies, primary inflammatory or idiopathic organ diseases such as membranoproliferative nephritis, or idiopathic psychosis and pediatric autoimmune diseases including JDM, systemic sclerosis, Sjögren’s syndrome, overlap syndromes, and polyarticular and systemic juvenile idiopathic arthritis. Children with a recent diagnosis of pSLE can present with concurrent infections such as CMV, which may confound the clinical presentation and response to therapy. Uncommon infections may be present. Macrophage activation syndrome (MAS, a secondary hemophagocytic lymphohistiocytosis [HLH] syndrome) is a newly recognized inflammatory emergency and can present in pSLE patients. MAS has to be considered in a child with pSLE and unexplained fever and cytopenia, when associated with marked hyperferritinemia.

**Treatment**

The care for children with lupus, similar to other multiorgan rheumatic diseases, mandates a multidisciplinary approach of different subspecialty physicians including pediatric rheumatologists, nephrologists, psychiatrists, and adolescent medicine, allied health care specialists, social workers, and teachers.

**Immunosuppression**

The choice and dosing of immunosuppressive therapy regimen must be tailored to the extent and severity of the child’s organ disease. A thorough diagnostic evaluation is mandatory before initiating therapy. Immunosuppressive treatment protocols are commonly adopted from adult trials and meta-analyses. Treatment response criteria for pSLE were recently developed and validated.

Corticosteroids are the mainstay of lupus therapy. The general approach for major organ disease includes initial high-dose treatment with 2 mg/kg prednisone equivalent in two to three divided doses followed by a slow taper. This includes treatment of proliferative lupus nephritis, neuropsychiatric lupus except for chorea and SVT, myocarditis, and lung disease. Pulse intravenous (IV) methylprednisolone is frequently used for emergent situations including acute psychosis, MAS, and myocarditis. Arthritis, serositis, nonproliferative lupus nephritis, and mucocutaneous disease may require smaller initial doses of corticosteroids. The efficacy of corticosteroids in pSLE is well established. However, the significant toxicity limits its long-term use at high doses. Short-term side effects include weight gain, sleep disturbances, emotional instability, increased hair growth, and impaired glucose metabolism. Long-term effects
include cataracts, growth arrest, vertebral fractures, and avascular necrosis. Combination immunosuppressive regimens are commonly used in children with major organ involvement. Cyclophosphamide, mycophenolate mofetil (MMF), and azathioprine have been studied in observational cohorts of pSLE patients. IV cyclophosphamide was considered the gold standard for severe organ disease such as proliferative lupus nephritis and neuropsychiatric disease. In accordance to adult SLE protocols, cyclophosphamide is commonly used for induction over 6 months, followed by either azathioprine or MMF. Induction therapy with MMF was found to be safe, well tolerated, and effective in a small cohort of renal and nonrenal pSLE patients. Complete renal remission is achieved in 40% to 50% of children at 6 months and 75% at 12 months. Commonly reported side effects that led to discontinuation of MMF therapy included severe diarrhea and abdominal pain. In children, optimal dosing may require pharmacokinetic evaluation on a stable dose. Efficacy and safety of MMF as maintenance drug is well established in pSLE. Induction therapy with azathioprine was found to be equally efficacious as cyclophosphamide in children with proliferative lupus nephritis and renal failure. Azathioprine has a good efficacy and safety profile. Routine monitoring of blood counts and liver function is required. Significant toxicity may occur in children with mutations in the gene encoding thiopurine methyltransferase or thiopurine S-methyltransferase (TPMT); however, the role of genotyping remains controversial. Azathioprine is also used as a maintenance drug following induction with cyclophosphamide.

Dialysis is required in children with end-stage renal disease. Plasmapheresis is indicated for specific disease manifestations such as thrombotic-thrombocytopenic purpura (TTP), transverse myelitis, and steroid-resistant nephritis. B cell depletion with the anti-CD20 antibody rituximab is the main biologic therapy currently used in pSLE. It was shown to be effective as a single agent in hematologic disease and in addition to standard therapy in refractory, difficult-to-treat pSLE. The safety profile remains to be systematically studied in pSLE. Autologous stem cell transplantation is rarely performed in pSLE.

MAS therapy in pSLE may include IV immunoglobulin, IV methylprednisolone pulse therapy, cyclosporine, or even chemotherapy according to HLH protocol.

**Antimalarials**

Antimalarials are strongly recommended for children and adults with SLE. On the basis of predominantly adult studies and meta-analysis, antimalarials are thought to decrease overall mortality and improve long-term outcome,

modify lipid profiles,

and control joint and skin disease, in particular discoid lupus lesions.

**Adjunctive Therapy**

Supportive medical therapies include angiotensin-converting inhibitors for renal protection and hypertension, contraception when applicable, organ dysfunction therapies including anticonvulsants and antipsychotic medication, and anticoagulation when applicable. High-factor sunscreen for sunburn protection is important. Educational efforts have to include disease and treatment, medication side effects, infection risk, and impact on social life and school.

Vitamin D is known to be a strong factor of bone protection in pSLE. Sufficient vitamin D doses in addition to calcium intake and physical activity are required to maintain good bone health. More recently a novel role of vitamin D in maintaining immune homeostasis was recognized. This is supported by studies demonstrating an inverse correlation between vitamin D levels and disease activity, in particular in overweight children with pSLE.

**Outcome**

Lupus in children and adults is a relapsing/remitting disease. The burden of pSLE is complex to determine because many factors such as access to health care, individual patient characteristics, disease activity, confounding diseases such as infections, and responsiveness to treatment all contribute to overall mortality and morbidity.

The overall mortality as captured by standard mortality rate (SMR) for all SLE in the United States between 1992 and 2001 was 3.06 deaths per million inhabitants per year, in Brazil between 1985 and 2004 it was 3.8 (2601 deaths, 90% female), and in Denmark it was 4.6. With improved therapies, mortality rates were shown to decrease in Canada (SMR, 10.1 in 1970-1977; 4.8 in 1978-1985, and 3.3 in 1986-1994). When comparing childhood- with adult-onset SLE, childhood-onset SLE was found to be independently associated with an increased mortality risk (hazard ratio [HR], 3.1), as was low socioeconomic status measured by education (HR, 1.9), and end-stage renal disease (HR, 2.1).

Young age at disease onset was repeatedly shown to be a predictor of adverse outcome. Children with pSLE in poor countries clearly have a higher mortality: In a small study from Nigeria the mortality was 30%. The Latin American LUMINA cohort had an 81% survival at last follow-up, the recent 5-year patient survival rate in Iran was 82.5%, and in Canada it was 100%. Infections continue to be the main cause of death in developing countries with limited access to health care.

Nephritis has been consistently identified as a predictor of poor outcome in pSLE. Histologic subtype of proliferative disease, evidence of disease relapse, certain ethnicities, and poor response to therapy were strong predictors of end-stage renal disease in pSLE. Gibson demonstrated that treatment resistance portended a high risk of end-stage kidney disease and disproportionately affected African-American children with lupus nephritis.

Children with pSLE accrue disease- and treatment-related damage as captured in the domains of the Systemic Lupus International Collaborative Clinics (SLICC) Damage Index (see Chapter 80), constantly adding to the overall disease burden. Osteoporosis, cataracts, and osteonecrosis/avascular necrosis (AVN) are the leading domains of damage accrual. Individual patient characteristics, disease activity, corticosteroid therapy, calcium/vitamin D deficiency, and immobility contribute to impaired bone health in pSLE. AVN occurs in 6% to 10% of pSLE patients overall and is associated with corticosteroid therapy.
Drug-Induced Lupus Erythematosus

Several medications can cause systemic and subacute or chronic cutaneous lupus phenotypes in children. The cutaneous manifestations of systemic drug-induced lupus (DIL) include malar rash, purpura, erythema nodosum, urticaria, and photosensitivity. Systemic symptoms include arthritis, oral ulcers, pleuritis, hematologic manifestations, and less commonly renal disease. Characteristic laboratory findings of DIL are positive ANA and antihistone antibodies. Drugs implicated are minocycline, anticonvulsive drugs, hydralazine, procainamide, and isoniazid.

Management of drug-induced lupus is based on the withdrawal of the offending drug. Topical and/or systemic corticosteroids and other immunosuppressive agents may be required in resistant cases.

Neonatal Lupus Erythematosus

Neonatal lupus erythematosus (NLE) is an acquired disease of the newborn caused by placental transfer of maternal anti-SSA/Ro and anti-SSB/La IgG antibodies. These can be present in mothers with SLE, Sjögren’s syndrome, and other autoimmune connective disorders, as well as clinically healthy women. Antibody transfer can lead to inflammation of the cardiac conducting system and subsequent fibrosis resulting in congenital heart block (CHB), which may be detected as early as 20 weeks of gestation. A prolonged PR interval is the first electrocardiographic sign of conduction system abnormality in NLE. The degree of heart block can vary, and rapid clinical progression from normal sinus rhythm to complete CHB over 2 weeks may be observed, causing life-threatening cardiomyopathy and fetal hydrops in the most severe cases. Isolated endocardial fibroelastosis can be found in some infants. Interestingly, infants with prenatal exposure to high-titer anti-SSB/La antibody levels are more likely to have noncardiac features of NLE, whereas cardiac disease tends to be associated with moderate or high maternal anti-SSA/Ro levels, independent of anti-SSB/La titers in CHB. The overall risk of CHB in anti-SSA/Ro-positive women is estimated to be 2% to 5%, but this risk may be increased by 10-fold in women with a previous child with CHB. A recent study suggests an overall recurrence rate of cardiac NLE of 17%, independent of maternal health, antenatal use of steroids, antibody status, severity of cardiac disease in the first affected child, or sex of the subsequent child.

In addition to heart block, newborns with NLE can present with a characteristic NLE rash, hepatic dysfunction, and hematologic abnormalities including significant thrombocytopenia. Typically the NLE rash is located around the eyes but may present elsewhere on the body. Hepatobiliary disease can have three distinct presentations: (1) transient conjugated hyperbilirubinemia with mildly raised liver function tests (LFTs) in the first weeks of life; (2) mild elevations of LFTs at 2 to 3 months of life; and (3) severe liver failure during gestation or in the neonatal period. NLE neurologic involvement can include magnetic resonance imaging (MRI) findings of nonspecific white matter changes and calcification of the basal ganglia. NLE “vasculopathy” is reported. Recently, an association of NLE and hydrocephalus has been recognized. Chondrodysplasia punctata, a stippling of the epiphyses, and pulmonary capillaritis are rare clinical presentations of NLE.

In a prospective multicenter study of 128 infants whose mothers had been referred for the presence of anti-SSA/Ro antibodies, regardless of their diagnosis, hematologic abnormalities and raised liver enzymes were seen in 27% and 26%, respectively. Cutaneous NLE manifestations were present in 16%. Only 2 of the 128 infants (1.6%) presented with complete CHB. In a recent Japanese review, 193 infants with NLE were described reporting CHB in 23%. Treatment of CHB in NLE remains controversial. Prevention of progression to complete CHB may be achieved by treating the mother with fluorinated steroids (dexamethasone or betamethasone), which are not metabolized by the placenta and are available to the fetus in an active form. IV immunoglobulin has been used to prevent the development of CHB in the index patient and in subsequent pregnancies. The current recommendation is to screen anti-Ro(SSA) antibody-positive mothers with serial echocardiograms and obstetric sonograms biweekly starting from week 16 of gestation. Early detection of cardiac manifestations of NLE including premature atrial contractions or moderate pericardial effusion preceding CHB may potentially be targeted with preventive therapy. First-degree heart blocks can be reversed by dexamethasone treatment of the mother. Once third-degree block is unequivocally identified, reversal is unlikely to be achieved. The majority of children with CHB require pacemakers.

The treatment approach to extracardiac manifestations of NLE is conservative. Skin disease may require topical corticosteroids and sun protection. Transient elevations of LFTs and cytopenias commonly do not require therapy. The morbidity of cardiac neonatal lupus is estimated to be 20%. Mortality is particularly high in patients with CHB and concurrent cardiomyopathy. Children with NLE can develop SLE later in life. Concerns of potential long-term neurocognitive deficits of NLE patients need further evaluation.

JUVENILE DERMATOMYOSITIS

Definition and Criteria

Juvenile dermatomyositis (JDM) is an inflammatory, immune-mediated vasculopathy with predominant involvement of muscle and skin that may involve other organs as well. JDM is by far the most common form of idiopathic inflammatory myopathy (IIM) in children and adolescents; therefore other types (such as juvenile polymyositis) are not addressed in this chapter.
Epidemiology

The incidence of JDM has been reported at approximately two to three cases per million children with a mean age of onset of JDM of 6 to 9 years of age. JDM may begin before 4 years of age in approximately 25% of cases. Females are affected more commonly than males in a ratio of approximately 2:1. Birth distributions for some subgroups of JDM patients have been noted and suggest that perinatal exposures may influence the onset of disease.

Genetics, Etiology, and Pathogenesis

It is thought that JDM is an autoimmune disorder in which environmental factors trigger an immune vasculopathy in genetically susceptible individuals. There is circumstantial evidence supporting the possible role of infection in the pathogenesis of JDM. Gene expression profiling in newly diagnosed JDM patients demonstrated interferon signature patterns in affected muscle tissue, suggesting preceding viral infection. Electron microscopic studies of affected muscle demonstrate tubuloreticular inclusions, which can also indicate a type I interferon response. Many different infections have been associated with JDM.

Both human leukocyte antigen (HLA) and non-HLA genetic relationships have been reported to be disease associated or protective. In Caucasians, the HLA allele HLA DRB1*0301 is the strongest HLA risk factor (odds ratio [OR], 3.9). The HLA associations do not seem to affect disease course or complications. Similarly, the presence of polymorphisms of cytokine genes may confer an increased risk on the development of JDM or may be protective. Polymorphisms at these and other alleles may also be associated with disease complications and course.

The central events of the immune angiopathy of JDM include an overexpression of major histocompatibility complex (MHC) type I molecules on the surface of myofibers and a type I interferon response. An immune complex–mediated vasculopathy with the presence of the C5-9 membrane attack complex, as well as immunoglobulins and C3 complement with complement activation, is evident. There is also a perivascular and perimysial infiltration of plasmacytoid dendritic cells, leading eventually to infiltration by CD4+ T cells, B cells, macrophages, proinflammatory cytokines, and chemokines. This leads to vascular damage, capillary dropout, and muscle ischemia. Upregulation of MHC class I molecules on myofibers is associated with the activation of nuclear factor κB (NFκB), which can lead to muscle damage. Maternal microchimerism has been noted in the majority of patients with JDM and in frequencies much greater than siblings or healthy controls, suggesting that mechanisms similar to graft-versus-host disease may also play a role in the pathogenesis of JDM.

Clinical Features

Patients usually present with an insidious onset of fatigue, decreased functional ability, stiffness, and weakness. Irritability may result from muscle pain and an inability to participate in routine activities, especially in young children. A more acute onset, with fever, may occur in about half of patients. Weakness with rash is the presenting problem in about 50% of cases, weakness alone in about 25%, and the remainder present with predominantly skin symptoms. The clinical features at presentation are shown in Table 108-2.

Symmetric proximal muscle weakness typically presents first in the lower extremities, manifesting as difficulty climbing stairs and running. Patients may demonstrate a Gower sign reflecting weakness of the lower limb and trunk muscles. Reaching for objects above the head and hair brushing are difficult because of weakness of the shoulder girdle. An increased lumbar lordosis results from weakness of the trunk muscles, which also makes it difficult to roll in bed and get out of bed. Neck flexor weakness makes it difficult to hold the head upright. The muscles may be painful and tender to touch due to edema. Weakness of the distal musculature is unusual but may be present late in disease. Weakness of the palatal musculature results in dysphonia, dysphagia, and nasal regurgitation. Rarely, patients may be so weak as to be bedbound.

Skin rashes are present in approximately 75% of patients at presentation. Rash may be the presenting feature and in fact the only symptom (clinically amyopathic dermatomyositis), although subtle abnormalities may be detected on muscle MRI. Some skin features are predictive of poor outcome such as calcinosis and severe nail-fold capillary abnormalities.

The cutaneous features of JDM have been well summarized in a comprehensive review. It is important to note that activity of the skin disease frequently does not correlate with activity of the muscle disease, and cutaneous abnormalities may have a significant impact on the patient’s quality of life.

Gottron papules occur over the extensor surfaces of the metacarpophalangeal (MCP) and interphalangeal (IP) joints, as well as the knees, elbows, and medial malleoli (Figure 108-3A). They are erythematous to violaceous papules and may have associated scaling, crusting, erosions, ulcerations, or pigmented change. This is differentiated from Gottron sign, which involves macular lesions occurring in the same distribution.

The pathognomonic heliotrope rash consists of violaceous to erythematous discrete or confluent macules confined to the upper eyelids (Figure 108-3B). This can extend periorbitally and often presents with generalized periorbital edema with discoloration.

Other cutaneous lesions include erythematous lesions in both sun-exposed and non–sun-exposed areas. Common
Calcinosis cutis can occur in up to 40% of patients with JDM; occasionally it may be present at the time of diagnosis. It is a dystrophic calcification and occurs more often in patients who have been “undertreated” or had a delay in the start of treatment. Calcinosis may take several forms: superficial plaques and nodules; tumoral; fascial plane deposition; and an exoskeleton (Figure 108-4A, D). Patients may have more than one pattern. The superficial lesions are often subject to minor trauma and can lead to skin breakdown. They may also extrude a chalklike material (see Figure 108-4A). Tumoral deposits can impair function and lead to skin breakdown, especially when they involve flexural areas. Occasionally the calcinosis in these areas leads to an intense inflammatory reaction resembling cellulitis. Fascial plane calcinosis may impair function when crossing joint lines (see Figure 108-4B). An exoskeleton may give a scleroderma-like picture. Contrary to old beliefs that calcinosis was purely a healing process indicating that JDM was inactive, these lesions are often associated with ongoing active disease requiring more aggressive systemic treatment.

Lipodystrophy is a late manifestation reported in up to 40% of patients, developing a median of 4.6 years after diagnosis. It may be localized, partial, or generalized and frequently occurs with metabolic syndrome, which includes hyperglycemia, hypertriglyceridemia, insulin resistance, hepatomegaly, transaminitis, and premature organ failure. Acanthosis nigricans may occur as well.
GI manifestations include dysphagia from both palatal muscle weakness and involvement of the distal esophagus. Intestinal vasculopathy can result in diffuse abdominal pain, lower GI bleeding, or bowel perforation with peritoneal free air.

Arthritis occurred in 61% of patients in one cohort, reported a median of 4.5 months after the JDM onset. Osteopenia may result from disuse and treatment, and pathologic fractures may develop. Rhabdomyolysis is a rare complication that may follow infection.

Multiple respiratory manifestations may occur and are often subclinical. In one large case series followed for a mean of 16.8 years, a low total lung capacity (TLC) was found in about 25% and a low diffusing capacity for carbon monoxide in 10%.

Clinically amyopathic JDM (CAJDM) is a unique form of JDM in which patients have a characteristic skin rash in the absence of any abnormalities of muscle including MRI changes for at least 6 months. Some patients have cutaneous changes and no weakness but abnormal muscle tests and may be categorized as "hypopathic JDM." In a systematic review, 26% of 68 patients with CAJDM developed classic JDM, with weakness developing a mean interval of 1.9 years after the onset of cutaneous symptoms; calcinosis developed in 4%. No laboratory or ancillary muscle testing was predictive of the development of muscle disease. Therefore immunosuppressive treatment is generally not indicated (other than required for the skin disease), but close follow-up is essential.

Figure 108-4  A, Elbow and forearm of a 14-year-old girl with chronic continuous juvenile dermatomyositis showing multiple sites of tumoral calcinosis. Several of these, particularly over pressure points, have ulcerated extruding a chalklike material. B, Solitary site of tumoral calcinosis posterior to the knee resulted in limited range of motion. C, Fascial and tumoral calcinosis affecting the lower extremity. D, Same patient as in A demonstrating marked calcification giving an exoskeleton-like appearance across the chest wall.
monoxide (DLco) in about 50% of patients. Just over one-third of patients had abnormalities on high-resolution computed tomography (CT) scanning such as interstitial lung disease, chest wall calcinosis, and airway disease. Spontaneous pneumothorax has been reported.

Cardiac manifestations are uncommon and can include pericarditis, myocarditis, and arrhythmias. However, subclinical left ventricular diastolic dysfunction, systolic hypertension, and electrocardiogram abnormalities occurred in 22% of one series of patients and were associated with cumulative organ damage. Hypertension is usually associated with high-dose glucocorticoid therapy.

The association with malignancy is limited to case reports such that a search for malignancy is not required in children with JDM.

**Disease Monitoring**

Muscle enzyme levels are frequently normal soon after treatment and are therefore unreliable indicators of disease activity. Several tools have been developed and validated to monitor the course and outcome of children with JDM. These include the 0- to 10-point Manual Muscle Test, the Childhood Health Assessment Questionnaire (CHAQ), the Childhood Myositis Assessment Score (CMAS), the Disease Activity Score (DAS), the Myositis Damage Index (MDI), and the Intention to Treat Index (MITAX). The Myositis Disease Activity Assessment Visual Analogue Scale (MYOACT) consists of a series of 10 visual analogue scales in different organ systems. In 2008 a prospective validation study of a core set for the evaluation of response to therapy in JDM was published under the auspices of the Paediatric Rheumatology International Trials Organization, the ACR, and the European League Against Rheumatism (EULAR) (Table 108–3). Provisional criteria for the evaluation of response to therapy in JDM include at least 20% improvement from baseline in three of six core set variables with no more than one of the remaining worsening by more than 30%, which cannot be muscle strength. The Cutaneous Assessment Tool has undergone preliminary validation in a series of 113 children with JDM. It measures both skin activity and skin damage.

**Diagnosis and Diagnostic Tests**

The diagnostic criteria published by Bohan and Peter require that patients have a characteristic skin rash plus three of the following four to meet the definition of “definite” JDM: symmetric proximal muscle weakness, elevated serum levels of muscle-derived enzymes, myopathic electromyogram (EMG), and histologic evidence of myositis. Patients with rash and two of these four criteria may be diagnosed with “probable” JDM. Because both EMG and biopsy are invasive procedures, many practitioners now rely on MRI studies of muscle to support the diagnosis (see later).

Laboratory evaluation in children with JDM helps to support the diagnosis and exclude other causes of muscle weakness. Systemic markers of inflammation such as the ESR and CRP generally reflect the degree of disease activity. Serum levels of neopterin and elevated levels of CD19* B lymphocytes have been suggested as good markers of disease activity, as has von Willebrand factor.

Elevated serum levels of muscle enzymes form one of the diagnostic criteria of JDM. Measurements of creatine kinase, lactate dehydrogenase, aspartate transaminase, and aldolase should all be obtained. However, their degree of elevation does not necessarily correlate with active disease and they can occasionally be normal, even at presentation, particularly with long-standing disease. Serum levels of muscle enzymes drop dramatically with treatment, often before a clinical improvement is seen.

Electromyography has generally fallen out of favor despite still being part of the diagnostic criteria, especially if MRI is available. Characteristic EMG changes include spontaneous fibrillations, increased insertional activity, decreased amplitude, and duration of action potentials.

Although abnormal muscle biopsy is one of the criteria proposed by Bohan and Peter in the diagnostic criteria for juvenile dermatomyositis, many clinicians elect not to do muscle biopsies when patients present with classic clinical features of JDM. This is because the procedure is invasive, painful, may not add to the diagnostic accuracy in individual patients, and may be normal in up to 20% of patients. Normal results may occur because of sampling error or patchy muscle involvement. Better yield might be afforded through the use of MRI to determine best sites for biopsy. Care should be taken not to biopsy a site that has previously undergone electromyography.

The characteristic light microscopic features are suggestive of an inflammatory vasculopathy. This includes endothelial cell swelling, capillary dropout with a reduced capillary-to-muscle fiber ratio, microthrombosis, and infarction. There is a relatively sparse inflammatory infiltrate consisting mainly of T cells, but there may be myeloid cells present as well. Muscle fibers demonstrate perifascicular atrophy, overexpression of class I major histocompatibility complex, deposition of immunoglobulin, and the membrane attack complex C5-9. There are also areas of degenerating and regenerating muscle fibers. Chronic

---

**Table 108–3** Domains and Suggested Variables Included in the Final Core Set for the Evaluation of Response to Therapy in Juvenile Dermatomyositis (JDM)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Suggested Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physician’s global assessment of patient’s overall disease activity</td>
<td>10-cm VAS</td>
</tr>
<tr>
<td>Muscle strength</td>
<td>CMAS (or MMT)</td>
</tr>
<tr>
<td>Global JDM disease activity tool</td>
<td>DAS (or MYOACT or MITAX)</td>
</tr>
<tr>
<td>Parent’s global assessment of patient’s overall well-being</td>
<td>10-cm VAS</td>
</tr>
<tr>
<td>Functional ability assessment</td>
<td>CHAQ</td>
</tr>
<tr>
<td>Health-related quality of life assessment</td>
<td>CHQ physical summary score</td>
</tr>
</tbody>
</table>

CHAQ, Childhood Health Assessment Questionnaire; CHQ, Child Health Questionnaire; CMAS, Childhood Myositis Assessment Scale; MITAX, Myositis Intention to Treat Index; MMT, Manual Muscle Test; MYOACT, Myositis Disease Activity Assessment; VAS, Visual Analogue Scale.

changes include an increase in the perimysial and endo-
myrial connective tissue, thought to reflect muscle fiber
damage and loss. Histopathologic changes have been
shown to correlate with both ulcerative disease and poor
prognosis.\textsuperscript{172}

Recently, an international group has developed a scoring
system that can be used in routine laboratories. The scoring
tool uses four domains to reflect the degree of pathology:
inflammatory, vascular, muscle fiber, and connective tissue.
It also includes a visual analogue scale from 0 to 10 to reflect
overall damage.\textsuperscript{173}

Currently the diagnosis of JDM is made on the basis
of Bohan and Peter criteria. However, many pediatric rheuma-
tologists now have turned to MRI to assist with the diagno-
sis and avoid more invasive tests such as EMG and muscle
biopsy.\textsuperscript{174} Inflammation characteristic of JDM is seen as
high-signal intensity on fat-suppressed weighted and short
tau inversion recovery (STIR) images.\textsuperscript{175} STIR sequences
can also reveal fasciitis and panniculitis. Muscle atrophy is
best appreciated on T1-weighted sequences as increased
signal between muscle planes. An increase in mean T2
relaxation time correlates with increased muscle disease
activity and muscle strength.\textsuperscript{176}

ANA positivity is seen in 10% to 85% of patients with
JDM. Myositis-associated (MAA) and myositis-specific
antibodies (MSA) are uncommon unless the JDM is part of
an overlap syndrome.\textsuperscript{177,178} In those patients who are posi-
tive for MAAs and MSAs (anti–signal-recognition particle,
anti-synthetase, and anti-Mi2), the clinical associations are
the same as in adult disease. The newly described autoanti-
body anti-p155/140 has been identified in up to 29% of one
series of JDM patients,\textsuperscript{179} and anti-p140 has been identified
in 23% of JDM patients, associated with calcinosis in one
series.\textsuperscript{180}

**Differential Diagnosis**

The most important differential diagnosis for patients pre-
senting with rash and muscle weakness is SLE, particularly
for those patients presenting with significant arthritis and a
malar rash. Characteristic autoantibodies of SLE, cytope-
nias, renal disease, and hypocomplementemia help to dif-
ferentiate these disorders. Patients with systemic sclerosis
and prominent myositis may be difficult to differentiate
from patients with JDM. Patients with mixed connective
tissue disease and other overlap syndromes may have fea-
tures of JDM in addition to those of other autoimmune
connective tissue diseases.

Other idiopathic inflammatory myopathies are extremely
rare in children. These include inclusion body myositis,
granulomatous myositis, and macrophagic myositis.

Patients who present with either no rash or mild rash
and predominant muscle weakness may need to be differen-
tiated from patients with primary myopathies. Patients with
muscular dystrophies often have positive family histories
and an insidious onset of disease with characteristic muscle
groups involved. Congenital myopathies usually present
in infancy with hypotonia. Metabolic myopathies may be asso-
ciated with developmental delay. Cramping and weakness
after exertion may also be signs of metabolic myopathy.

Various infections may lead to an acute myositis. Perhaps
the best recognized is influenza B, presenting with acute calf
pain, weakness, and raised muscle enzyme levels. *Trichinella*
infection may be associated with periorbital edema and
significant peripheral eosinophilia. Many other bacteria,
viruses, and parasites may cause myositis; they should be
suspected in the appropriate clinical circumstance.

**Treatment**

The management of patients with JDM requires a multidis-
ciplinary team approach with medical specialists (rheuma-
tologists, dermatologists, neurologists); nurses; rehabilitation
specialists; social workers; and nutrition specialists.

Early aggressive treatment has been shown to result in
better long-term outcomes and prevent disease-related
complications.\textsuperscript{139,141,181,182} The cornerstone of treatment
is high-dose, daily corticosteroid, usually combined with a
second-line agent, typically methotrexate (MTX). Some
practitioners advise the early use of high-dose, IV pulse
methylprednisolone to ensure appropriate absorption
when there may be a concern of intestinal vasculitis,\textsuperscript{183}
when there is a flare of disease, or when a patient seems
unresponsive to standard steroid therapy. The usual course
has been to start at 2 mg/kg/day in one to three divided
doses and to begin to taper when muscle enzymes have
normalized and strength has improved, with a subsequent
slow taper over 18 to 24 months in uncomplicated cases.

With steroid treatment alone, a significant number of
patients do not respond fully and have complications,
and some patients may be overtreated and have steroid-
related complications. As a result, other agents are com-
monly prescribed.

MTX has been used for decades in children with
steroid-resistant JDM and has recently been incorporated
into many treatment protocols. In addition to its anti-
inflammatory effect, it allows for a lower cumulative dose of
corticosteroids.\textsuperscript{184} A recent survey of North American pedi-
atriic rheumatologists documented that the most common
treatment approach to patients with JDM is a combination
of prednisone and MTX.\textsuperscript{185}

In patients who do not respond adequately, there
are several options. IV immunoglobulin (IVIG) has been
shown in a randomized controlled trial in adults with der-
matomyositis to be effective,\textsuperscript{186} as well as in case series in
childhood JDM.\textsuperscript{187} Several different protocols have been
described. Generally, if there is no improvement within
2 months, it is unlikely that additional IVIG will be
effective.

Mycophenolate mofetil at an initial dose of 20 mg/kg in
two divided doses was studied in 30 patients who had not
responded to prednisone and MTX.\textsuperscript{188} A significant improve-
ment in muscle and skin DAS was noted at 12 months, with
a significant reduction on steroid dose and no serious adverse
events. There was an increase in mean height and weight
as well.

Cyclosporine A is used frequently in Europe as a second-
line agent with good results.\textsuperscript{189} A trial is currently under way
in Europe for newly diagnosed patients with JDM compar-
ing prednisone alone with prednisone plus MTX with pred-
nisone plus cyclosporine A.

Cyclophosphamide has generally been reserved for
patients with treatment-resistant disease, severe ulcerative
disease, or lung involvement. Major clinical benefit was
noted in a small cohort of patients without serious toxicity.\textsuperscript{190}

The results with anti–tumor necrosis factor (TNF) treatment with both etanercept and infliximab have been reported in several small case series, with both positive and negative outcomes.\textsuperscript{191} There are reports of several cases of myositis developing while on etanercept treatment. The use of rituximab has been reported in only a small number of patients with good results.\textsuperscript{192} The results of the rituximab in myositis trial are awaited. Nevertheless, a trial may be warranted in patients with severe unresponsive disease. It is possible that patients who have myositis-specific autoantibodies would respond better. A few patients have undergone successful autologous stem cell transplantation.\textsuperscript{193}

One approach to the treatment of patients with JDM has been to use a step-wise addition of medications if patients fail to improve according to a predetermined outcome (similar to “treat-to-target” approach in rheumatoid arthritis). Using this approach, Kim and colleagues\textsuperscript{192} reported excellent outcomes in a series of patients who were treated, progressively, with prednisone (98%), MTX (78%), IV methylprednisolone (84%), cyclosporine A (27%), IVIG (20%), plasma exchange (8%), and cyclophosphamide (4%).

Recently, the Childhood Arthritis and Rheumatology Research Alliance (CARRA), using consensus building techniques, proposed three protocols for the treatment of moderately severe JDM. All include corticosteroids in one to two doses per day (maximum 60 mg) plus MTX, preferably given by the subcutaneous route at the lower dose. Clinicians may also add either pulse steroids or IVIG of 15 mg/m\textsuperscript{2} or 1 mg/kg, maximum 40 mg/wk.\textsuperscript{194} One approach to the pharmacologic management of JDM is listed in Table 108-4.

Cutaneous disease occasionally requires specific treatment irrespective of the treatment of the muscle disease. Sun protection with broad-spectrum sunscreens with an SPF of 30 or higher should be used daily. Topical emollients are helpful for dry and pruritic skin. Antihistamines may be helpful in reducing pruritus. Topical corticosteroids and tacrolimus may be indicated for the very inflamed lesions in patients who either do not respond to, or do not require, systemic therapy. Hydroxychloroquine was associated with a significant improvement of skin disease in a small cohort of patients who had an incomplete response to systemic corticosteroids.\textsuperscript{195}

Physiotherapy and occupational therapy are important to all children with JDM. Early in the disease, attention must be paid to stretching joints to prevent muscle contractions. Occasionally, contracted joints should be splinted. As the muscles heal, a more vigorous active exercise program can be initiated to strengthen muscles that may have atrophied from the combined effects of the inflammatory myopathy, disuse, and high-dose corticosteroid treatment. Active exercise may in fact exert anti-inflammatory effects in exercised skeletal muscle.\textsuperscript{196}

Treatment for calcinosis has been disappointing. Case reports with positive results have included bisphosphonates and calcium channel inhibitors, but most patients do not respond. Surgical excision may be necessary for particularly troublesome individual lesions and can have excellent results.

### Table 108-4  An Approach to the Management of a Patient with Juvenile Dermatomyositis

<table>
<thead>
<tr>
<th>For Muscle Weakness</th>
<th>Initial Treatment</th>
<th>If Failure to Respond to Initial Treatment, Consider Adding:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prednisone</td>
<td>IV pulse methylprednisolone ± IVIG</td>
</tr>
<tr>
<td></td>
<td>Methotrexate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>± MTX</td>
<td></td>
</tr>
<tr>
<td></td>
<td>± 1 mg/kg, max.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40 mg/wk</td>
<td></td>
</tr>
<tr>
<td>For Skin</td>
<td>Hydration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sun protection</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| For Failure to Respond to Second-Line Treatment, Consider Adding: |
| Cyclophosphamide    |
| Tacrolimus          |
| Infliximab          |
| Rituximab           |

*IV, intravenous; IVIG, intravenous immunoglobulin.*

### Outcome

In the precorticosteroid era, approximately one-third of patients with JDM went into a complete clinical remission, one-third had a chronic course, and one-third died.\textsuperscript{197} Advances in medical therapy, earlier diagnosis, and more aggressive treatment protocols have led to significantly improved outcomes; however, there is still significant morbidity and a small mortality associated. Current mortality is less than 5%,\textsuperscript{198,199}.

Patients may follow one of three disease courses: monophasic, polycyclic (flares of disease while off treatment), or chronic continuous courses. Delayed recognition of disease and initial undertreatment may result in more prolonged disease course. In the “modern era,” two large series showed that approximately 40% of patients pursued a monophasic course and the remainder had either a polycyclic or a chronic continuous course.\textsuperscript{141,159} Using an aggressive early stepwise treatment approach, only 4% had a chronic disease course, suggesting that early control of muscle inflammation prevents long-term morbidity.\textsuperscript{192}

The largest follow-up study included 490 patients from Europe and Latin America followed for a mean of 7.7 years seen between January 1980 and December 2004.\textsuperscript{194} Reduced muscle strength and/or endurance were documented in 40% to 50%, although it was severe in less than 10% of patients. Persistent disease activity was noted in 40% to 60%. Cumulative damage occurred in 70%, primarily cutaneous. Decreased functional ability was reported in 40% and major impairment in 7%. A chronic course was the strongest predictor of poor prognosis.

There are a number of factors considered to predict a poor outcome including a delay in treatment or...
undertreatment,181 chronic continuous course,199 presence of rash and capillary nail-fold dropout at 6 months of treat-
ment,14 ulcerative skin disease, and abnormalities on initial
muscle biopsy.172 In addition, the presence of myositis-
specific antibodies, unique HLA types, and the TNF poly-
morphism 308 may also determine a poorer outcome.

**SCLERODERMA**

The scleroderma disorders in children can be classified into
systemic, localized, and others (Table 108-5). The systemic
scleroderma are rare in the pediatric age group and are outnumber-
ered by the localized forms by approximately 10:1.

**Systemic Sclerosis**

**Epidemiology**

Juvenile systemic sclerosis (JSSc) makes up approximately
10% of all cases of systemic sclerosis120; the incidence rate
in a recent U.K. study was reported as 0.27 per million
children. Females outnumber males anywhere from 4:1 to
10:1.201 The incidence seems to increase with increasing age,202,203 although in one large multicenter review the mean
age of onset was 8.1 years.203 Diffuse disease is much more
common than limited disease. An increased family history
of autoimmune disease including scleroderma has been
noted in some series.

Little work on etiology and pathogenesis has been done
specifically in juvenile systemic sclerosis, and it is assumed
to be identical to adult disease. The interested reader is
referred to Chapter 83.

**Clinical Features**

The onset of diffuse systemic sclerosis is often insidious,
and delay in diagnosis ranged from a median of 1 to 2.8 years
in three large series.200,202 The most common presenting
features are Raynaud’s phenomenon, skin edema, and
sclerodactyly (Table 108-6). The diagnosis should be con-
sidered suspect in the absence of Raynaud’s phenomenon.

Although skin edema is the earliest cutaneous abnormal-
ity, it is followed fairly quickly by induration, sclerodactyly,
and loss of facial creases.

**Table 108-6**  Clinical Features at Diagnosis and
during Course in 153 Patients with Juvenile
Systemic Sclerosis from 55 Centers

<table>
<thead>
<tr>
<th>Feature</th>
<th>% At Diagnosis</th>
<th>% During Course</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Skin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edema</td>
<td>35</td>
<td>46</td>
</tr>
<tr>
<td>Sclerodactyly</td>
<td>46</td>
<td>66</td>
</tr>
<tr>
<td>Skin induration</td>
<td>74</td>
<td>76</td>
</tr>
<tr>
<td>Calcinosis</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td><strong>Peripheral Vascular</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raynaud’s phenomenon</td>
<td>75</td>
<td>84</td>
</tr>
<tr>
<td>Digital infarcts</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td>Digital pitting</td>
<td>28</td>
<td>38</td>
</tr>
<tr>
<td>Abnormal capillaroscopy</td>
<td>25</td>
<td>52</td>
</tr>
<tr>
<td><strong>Respiratory</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyspnea</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>Abnormal chest radiograph</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>Abnormal chest CT scan</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>Reduced DLco</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>Reduced FVC</td>
<td>11</td>
<td>42</td>
</tr>
<tr>
<td><strong>Cardiac</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pericarditis/arrhythmia</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Heart failure</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Pulmonary hypertension</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td><strong>Musculoskeletal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle weakness</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Arthritis</td>
<td>17</td>
<td>27</td>
</tr>
<tr>
<td>Arthralgia</td>
<td>26</td>
<td>36</td>
</tr>
<tr>
<td>Tendon friction rubs</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td><strong>Gastrointestinal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dysphagia</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>Reflux</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Weight loss</td>
<td>18</td>
<td>27</td>
</tr>
<tr>
<td><strong>Renal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raised creatinine/proteinuria</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Renal crisis</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hypertension</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><strong>Nervous System</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seizures</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Peripheral neuropathy</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Abnormal brain MRI</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

- CT, computed tomography; DLco, diffusing capacity for carbon monoxide; FVC, forced vital capacity.

Raynaud’s phenomenon involves the distal extremities
and rarely other acral areas such as the earlobes and tip of
the nose. Occasionally it may lead to digital infarcts result-
ing in pitting or more significant gangrenous change. Capil-
lar nail-fold abnormalities, visible either to the naked eye
or by capillary microscopy, have been reported in at least
50% of patients and include areas of dropout and abnormal
capillaries with dilatation and tortuosity.

Abnormalities of the musculoskeletal system include
mild inflammatory synovitis, which is nonerosive; joint
contractures most commonly resulting from skin and sub-
cutaneous tightness around the joints; and myositis. In
patients with overlap syndromes the muscle involvement is

---

**Table 108-5**  Classification of Scleroderma in
Childhood

<table>
<thead>
<tr>
<th>Systemic Sclerosis</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffuse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Overlap Syndromes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed connective tissue disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overlap with features of SLE, JDM, and JIA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Localized Scleroderma</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circumscribed morphea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear morphea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalized morphea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pansclerotic morphea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed morphea</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Proposed Pavia criteria.211
JDM, juvenile dermatomyositis; JIA, juvenile idiopathic arthritis; SLE, systemic lupus erythematosus.
more significant. Otherwise scleroderma myositis is generally quite mild. Tendon friction rubs may be felt or heard in a minority of patients.

Involvement of the respiratory system has become the most significant cause of morbidity and mortality in patients with systemic sclerosis. Early changes may be documented by high-resolution computed tomography, although this may not offer much greater benefit than well-performed pulmonary function tests (PFTs). It is important to document the pulmonary status early in the disease process because it is only then, before irreversible fibrosis occurs, that reversal of abnormalities may be possible. Pulmonary function abnormalities include a reduced forced vital capacity, a reduced FEV₁-to-forced vital capacity (FVC), and a reduced DLco. Severe chest wall involvement with restriction of movement may also lead to abnormal PFTs. Early CT changes include a ground-glass appearance suggestive of alveolitis. Rarely, pleural effusions may occur. Cardiac involvement includes pericarditis (which may be asymptomatic), arrhythmia, and congestive heart failure (CHF). Pulmonary arterial hypertension may develop in the face of severe lung disease. Cardiac disease (resulting primarily from CHF) was the most common cause of mortality in one series of 135 patients with juvenile systemic sclerosis.

Involvement of the GI system is common with symptoms of dysphagia and gastroesophageal reflux. More diffuse involvement can lead to reduced GI motility with bacterial overgrowth resulting in malabsorption. Severe constipation may occasionally occur. Renal involvement is much less common in juvenile than adult scleroderma and can include proteinuria, hypertension, and the eventual development of renal crisis. Neurologic involvement is unusual but can include seizure, stroke, and peripheral neuropathy.

**Diagnosis and Diagnostic Tests**

The diagnosis of systemic sclerosis rests on the presenting signs and symptoms and the investigation of organs that may be affected by the process. Seventy-five to 97% of patients were antinuclear antibody positive in three series. Specific autoantibodies include antitopoisomerase in approximately 33%, anticientromere antibody in less than 10%, and autoantibodies associated with overlap syndromes (anti-PM-Scl, anti-U1-RNP, anti-Ro) in a smaller number depending on the series. Rheumatoid factor may be present in up to 20% of patients.

Skin biopsy is rarely performed. Pathologic findings include dense collagenization and loss of adnexal structures. An inflammatory infiltrate composed primarily of mononuclear and mast cells is seen early in the course.

It is important to investigate the various organ systems that may be involved by systemic sclerosis in order to help prognosticate and develop an appropriate management plan. Investigations should include, at a minimum, PFTs with a DLco, an electrocardiogram, an echocardiogram, and a chest radiograph. For patients unable to perform PFTs, high-resolution CT scan is indicated. This is also indicated in patients with abnormal PFTs to detect early alveolitis. Serum KL-6 is a mucin-like glycoprotein strongly expressed in type II pneumocytes and may be a useful noninvasive marker of pulmonary fibrosis in children with JSSc.

It can be assumed that most patients have GI involvement and that an upper GI series will be abnormal. This test is indicated for patients with severe pain unresponsive to standard agents and severe dysphagia. Hydrogen breath test and measuring fat-soluble vitamins may be helpful in patients with suspected malabsorption. Radiographs of the hands may be helpful in showing distal acro-osteolysis in patients with severe Raynaud’s phenomenon and occasionally may show calcinosis.

When patients present with Raynaud’s phenomenon, capillary nail-fold abnormalities, and edematous or indurated skin, the diagnosis is clear and the differential diagnosis is limited. However, patients are often seen early in the course with just Raynaud’s phenomenon and a positive ANA. In those situations the most important differential diagnoses to consider are systemic lupus erythematosus, overlap syndrome/mixed connective tissue disease, and juvenile dermatomyositis. Although many other diseases are associated with skin fibrosis, they are not associated with the characteristic multisystem involvement as systemic sclerosis and should not pose a diagnostic challenge.

Provisional classification criteria have been developed and accepted by the Pediatric Rheumatology European Society (PRES), ACR, and EULAR on the basis of a Delphi survey and nominal group techniques. These were proposed because current criteria had not been studied in children and are not sensitive enough to detect early disease. Furthermore, they do not include key features of early disease such as Raynaud’s phenomenon, ANA positivity, and capillary nail-fold changes. The proposed criteria are listed in Table 108-7. The presence of the one major and at least two minor criteria in a patient younger than 16 years classifies the patient as having juvenile systemic sclerosis, with a sensitivity of 90% and a specificity of 96%.

**Treatment**

The approach to the management of patients with systemic sclerosis should include treating the basic disease process itself, as well as the various organ manifestations. No controlled studies exist in children, and data must be extrapolated from the adult literature. The EULAR Scleroderma Trials and Research Group has made recommendations regarding management of patients with systemic sclerosis, and they should be considered for all patients. Because no specific treatment studies have been reported in children and adolescents, the reader is referred to Chapter 84 for a discussion of treatment where the same principles hold. It should be noted that for patients with rapidly progressive disease and progressive lung disease, autologous stem cell transplantation may provide the only opportunity for survival and has been successful in a number of patients with juvenile systemic sclerosis.

**Outcome**

Very little long-term outcome data are available for JSSc. Morbidity is substantial from the multisystem involvement with marked impact on the quality of life. The survival rates in juvenile systemic sclerosis are better than adult disease. Mortality in three large series varied from 12% to 30%,. The most common causes of death include...
Localized Scleroderma

Epidemiology

Localized scleroderma (LSc), also known as morphea, had a reported incidence of 2.7 per 100,000 of the general population in the Mayo Clinic series, and 3.4 per million children in the United Kingdom, making it much more common than systemic sclerosis in the pediatric age group. The female-to-male ratio is approximately 2:1, and the average age of onset is approximately 7 years. Mild lesions may never get to medical attention, so the incidence may be even higher. Congenital morphea has been reported. Several classification schemes have been developed; the proposed Pavia criteria are listed in Table 108-5. Linear lesions are more common in the pediatric age group.

Etiology and Pathogenesis

Like systemic sclerosis, it is likely that environmental factors trigger immune activation leading ultimately to fibrosis. There have been no genetic studies to date to suggest that the HLA or non-HLA systems play a role, although those studies are under way. Familial occurrence is uncommon, but there is a strong family history of other autoimmune disorders. Trauma has been considered as a possible inciting feature. Borrelia infections have been implicated in Europe but not in North America. The presence of elevated levels of serum cytokines (such as TNF and interleukin (IL)-1) that may influence fibroblast proliferation provide strong evidence for the role of immune activation in disease pathogenesis. Microchimerism has been reported, and as with other connective tissue diseases it suggests that mechanisms similar to graft-versus-host disease also may play a role. Several drugs and toxins may lead to cutaneous fibrosis, although none has been consistently identified in patients with LSc.

Clinical Features

There is usually a delay in presentation of several months because the lesions themselves are typically not symptomatic. Active circumscribed (Figure 108-5A) and linear lesions (Figure 108-5B) usually have a shiny, waxy appearance surrounded by a violaceous, erythematous border. They may be warmer than the surrounding and contralateral skin. Rarely there may be local itching or tingling. The lesions are indurated, and they may be either superficial or can extend to muscle and bone. Occasionally with linear lesions, extensive fascial involvement may occur. Lesions heal with hyperpigmentation or hypopigmentation and generally soften with time. Atrophy of the subcutaneous tissues is common. Linear lesions, if untreated, may result in growth deformity, joint contracture, loss of muscle bulk, and marked extremity weakness.

Lesions on the face and head can take the form of either a “saber-cut”—like lesion (en coup de sabre) or progressive hemifacial atrophy (also known as Parry-Romberg syndrome), where the epidermal changes are minimal but there is marked dermal and subcutaneous atrophy (Figure 108-5C). These may coexist in the same patient. With time, as the unaffected side of the face grows normally, there is progression of the facial asymmetry even though the disease may be inactive. Facial lesions may be associated with hemiatrophy of the tongue (Figure 108-5D), dental abnormalities, and ocular abnormalities. A small number of children develop seizures.

Pansclerotic morphea is rare but can be life threatening. There is marked thickening of the skin and deeper tissues involving the extremities and trunk, sparing the distal extremities. Although it is similar to systemic sclerosis in the extent of the fibrosis, internal organ involvement does not occur and Raynaud’s phenomenon is not common. Extracutaneous signs and symptoms have been reported in up to 20% of patients. They are more common in patients with linear lesions. The most common is arthritis,

Table 108-7  Provisional Criteria for the Classification of Juvenile Systemic Sclerosis (SSc)

<table>
<thead>
<tr>
<th>Major Criterion (Required)</th>
<th>Minor Criteria (At Least 2 Required)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximal skin sclerosis/induration of the skin</td>
<td>Cutaneous&lt;br&gt;Sclerodactyly&lt;br&gt;Peripheral Vascular&lt;br&gt;Raynaud’s phenomenon&lt;br&gt;Nail-fold capillary abnormalities&lt;br&gt;Digital tip ulcers&lt;br&gt;Gastrointestinal&lt;br&gt;Dysphagia&lt;br&gt;Gastroesophageal reflux&lt;br&gt;Cardiac&lt;br&gt;Arrhythmias&lt;br&gt;Heart failure&lt;br&gt;Renal&lt;br&gt;Renal crisis&lt;br&gt;New-onset arterial hypertension&lt;br&gt;Respiratory&lt;br&gt;Pulmonary fibrosis (HRCT/radiography)&lt;br&gt;Decreased DLco&lt;br&gt;Pulmonary arterial hypertension&lt;br&gt;Neurologic&lt;br&gt;Neuropathy&lt;br&gt;Carpal tunnel syndrome&lt;br&gt;Musculoskeletal&lt;br&gt;Tendon friction rubs&lt;br&gt;Arthritis&lt;br&gt;Myositis&lt;br&gt;Serologic&lt;br&gt;Antinuclear antibodies&lt;br&gt;SSc-selective autoantibodies (anticentromere, anti-topoisomerase 1 [ScI70], antifibrillarin, anti-PM-Scl, antifibrillin or anti-RNA polymerase I or III)</td>
</tr>
</tbody>
</table>

DLco, diffusing capacity for carbon monoxide; HRCT, high-resolution computed tomography.

hypergammaglobulinemia have been reported to correlate with active lesions, but this is not always the case. Rheumatoid factor is present in 10% to 25% and ANA positivity in approximately 50% of cases. Multiple specific autoantibodies have been reported, but antibodies to topoisomerase-I and centromere are distinctly unusual.

Treatment

The treatment of localized scleroderma depends on the stage of the lesion, as well as the extent of involvement. Few controlled studies have been done; therefore treatment recommendations have relied on general experience. Plaque lesions can generally be treated topically with either corticosteroids and/or calcipotriene. Markedly indurated lesions may respond better to imiquimod. Topical tacrolimus may also be used. Systemic treatment is usually indicated for rapidly progressive lesions, for lesions crossing joint lines, and for lesions that are potentially cosmetically deforming. A combination of corticosteroids and MTX is generally recommended. Corticosteroids may be administered as monthly pulses or orally in a dose of 1 to 2 mg/kg with a taper over 3 to 6 months. MTX is

Diagnosis and Diagnostic Tests

The diagnosis of LSc is usually made on the basis of characteristic cutaneous features. A skin biopsy can be of assistance when the diagnosis is not clear. Abnormalities consist of edema, an early infiltration by mononuclear cells, and excessive deposition of collagen. With time there is loss of skin appendages and rete pegs. Other skin diseases that may have a similar clinical presentation include lichen sclerosis et atrophicus, connective tissue nevus, collagenoma, and localized fibrotic disorders. The absence of significant internal organ involvement and Raynaud's phenomenon help differentiate LSc from systemic sclerosis and other autoimmune conditions.

Laboratory investigations are nonspecific and show mild or no systemic inflammation (ESR, CRP). Eosinophilia and hypergammaglobulinemia have been reported to correlate with active lesions, but this is not always the case. Rheumatoid factor is present in 10% to 25% and ANA positivity in approximately 50% of cases. Multiple specific autoantibodies have been reported, but antibodies to topoisomerase-I and centromere are distinctly unusual.

not necessarily associated in the area of skin involvement. It is seen more commonly in patients with a positive rheumatoid factor. Neurologic manifestations of seizure and headache occur almost exclusively in patients with facial lesions. Ocular abnormalities including asymptomatic anterior uveitis were reported in 3% of one large series.

Figure 108-5  A, Circumscribed/plaque morphea on the abdomen of a 5-year-old girl with localized scleroderma. Note the central ivory-colored area of induration surrounded by intense erythema indicative of an active lesion. B, Linear localized scleroderma involving the inner aspect of the left leg of an 11-year-old girl. Note areas of porcelain-white lesions surrounded by erythema. C, A 10-year-old girl with a history of linear scleroderma of the face with both en coup de sabre and Parry-Romberg lesions of 7 years’ duration. Note frontal alopecia and two bands of linear scleroderma on the right side of the forehead and face. There is hyperpigmentation and marked subcutaneous atrophy. Note also the loss of eyebrow and eyelashes on the right, small right nares, and thin lips on the right. D, Same patient as in C showing atrophy of the tongue.
administered at a dose of up to 1 mg/kg or 15 mg/m² weekly. At higher doses, the subcutaneous route is probably more effective. Treatment should be administered for at least 2 years, and 1 year after all activity has disappeared because there is about a 30% chance of recurrence if treatment is stopped too early. MMF may be used for patients who have not responded to this combination. Imatinib, cyclosporine A, and tacrolimus have also been effective in a small number of cases. Ultraviolet A therapy is used more frequently in Europe with good success. Autologous stem cell transplantation may be required for patients with pansclerotic morphea.

Patients with facial lesions have undergone cosmetic repair with generally good outcomes. Surgery may also be required to lengthen Achilles tendons.

In addition to medical and surgical treatment, a combined team approach is often required for patients with more extensive disease. Physical and occupational therapy are essential in improving and maintaining muscle strength, range of motion, and function. Psychosocial support is especially helpful for patients with facial lesions. Other medical personnel whose involvement may be required are neurologists, ophthalmologists, craniofacial surgeons, orthopedic surgeons, dentists, and orthodontists.

**Disease Monitoring**

To date, it has been difficult to monitor the course of the disease as clinicians have relied on insensitive measures such as warmth, color change, and change in size over time. Recently, some more objective measures have been studied including ultrasound, computerized skin score, and laser Doppler flow. A disease activity score has recently been developed and initial validation of a disease damage score has been undertaken. CARRA is currently establishing scores for activity and damage.

**Outcome**

Lesions tend to soften spontaneously over several years and to heal with pigmented change (usually hyperpigmentation) and subcutaneous atrophy. Linear limb lesions may lead to marked atrophy and joint contracture. Lesions may recur after many years of apparent inactivity. Neither self-esteem nor health-related quality of life appear to be diminished compared with controls, although not many patients with more disfiguring lesions were studied. Rarely patients have developed other autoimmune connective tissue disorders including systemic sclerosis and SLE. Patients with pediatric onset of LSc have a higher incidence of autoimmune disorders as adults.

**Eosinophilic Fasciitis**

Eosinophilic fasciitis (EF) is included by some within the classification of localized scleroderma, and several authors have reported that children with EF have a disease evolution to morphea. EF is extremely rare in the pediatric population. Affected children present with marked induration of cutaneous and subcutaneous tissues of the upper or lower extremities and occasionally the trunk or face. Its onset may be preceded by intense exercise. Raynaud’s phenomenon, internal organ involvement, and nail-fold capillary abnormalities are rare but may occur.

**Mixed Connective Tissue Disease**

Mixed connective tissue disease (MCTD) was initially reported as a disorder associated with a favorable prognosis and an excellent initial response to low-dose glucocorticoid therapy. It had a frequency of 0.3% in the U.S. Pediatric Rheumatology Database. Children present with arthritis, myositis, and cutaneous disease characteristic of scleroderma, SLE, or JDM. A decrease in aerobic capacity may occur from reduced muscle strength. Progression to a more scleroderma-like disease has occurred, with sclerodactyly and GI involvement, or an SLE-like disease may evolve. Nephritis may be more frequent and more severe in children than in adults. Children often have less pulmonary disease (hypertension) and more hematologic complications (thrombocytopenia) than adults. ANAs are present in high titers, often in a speckled pattern, to an extractable nuclear antigen and ribonuclear protein (RNP).

**VASCULITIS**

Vasculitis is a common clinical phenomenon in children. Vasculitis can occur in association with infections, medications, hypersensitivity reactions, and in the context of childhood systemic rheumatic diseases such as lupus. The most common primary or idiopathic vasculitis types are Henoch-Schönlein purpura (HSP) and Kawasaki disease (KD). Incidences of vasculitis subtype vary widely depending on characteristics of populations such as ethnicity and the method of ascertainment.

Similar to adults, childhood vasculitis is categorized by predominantly affected vessel size as small, medium, or large vessel vasculitis. The histopathologic characteristics vary between diseases and include karyorhexis, neutrophil infiltration and necrosis, giant cell formation, and lymphocyte infiltrates. In 2005 a classification system for childhood vasculitis was proposed. In 2008 the so-called “EULAR/PReS endorsed consensus criteria for childhood vasculitis” were validated.

**Small Vessel Vasculitis**

Inflammation of the small vessels is the most common vasculitis subtype in children. Typically exposure to infectious agents, medications, hypersensitivity such as serum sickness, or systemic illness can cause migration of neutrophils through the vessel wall, leukocytoclasis, and fibrinoid necroses. The resulting histologic diagnosis of leukocytoclastic vasculitis is a common result found on superficial punch biopsies done for suspected vasculitis. Additional immunofluorescence studies may reveal immunoglobulin (Ig) deposits along the vessel wall, evidence of immune complexes and/or complement activation. Deposition of IgA is the hallmark of HSP.

**Henoch-Schönlein Purpura**

HSP (or anaphylactoid purpura) is an IgA-mediated small vessel vasculitis predominantly affecting the skin and
causing a palpable purpura. Histologically, a leukocytoclastic vasculitis with extravasation of leukocytes and red cells, vessel wall damage, fibrinoid necrosis, and IgA deposition at the vessel wall and in the mesangium of the kidney can be found. Preceding upper respiratory tract infections are reported in more than 50% of children. A variety of bacterial and viral triggers, environmental stimuli, and host susceptibility factors such as autoinflammatory disease genes have been reported. HSP can occur before or during the course of systemic diseases such as antineutrophil cytoplasm antibody (ANCA) vasculitis or Crohn’s disease.

IgA appears to play a pivotal role in the pathogenesis of HSP. Abnormal glycosylation of the hinge region O-linked glycan of IgA1 has been implicated in the etiopathology of HSP. Abnormal IgA1 molecules were found to have a higher tendency to aggregate, interact with IgG, and form IgA-IgG complexes and deposits in the kidney. Similarly, serum levels of galactose-deficient IgA1 are elevated in Caucasian and Asian patients with IgA nephropathy. Schmitt demonstrated deposits of IgA-binding streptococcal M protein in the skin and kidney of HSP patients directly linking infection and vasculitis. Wu studied the role of leukotrienes and suggested that abnormally high levels of leukotriene B4 and lipoxin A4, potent activators of neutrophils, found in HSP patients with nephritis differentiate them from those without nephritis.

**Definition and Classification.** HSP is the most common defined childhood small vessel vasculitis characterized by the classic triad of palpable purpura, arthritis, and abdominal pain. In 2006 EULAR/PReS-endorsed consensus criteria for HSP were proposed (Table 108-8). The sensitivity of these criteria was found to be 100%, the specificity 87%.

**Epidemiology.** Overall the incidence of HSP is estimated at 10 to 20 per 100,000 children younger than 17 years old. Boys are more commonly affected than girls; the male-to-female ratio was recently reported at 1.8 : 1. Gardner-Medwin reported an annual incidence of 70.3 per 100,000 between the ages of 4 and 6 years in a population-based study. In contrast, HSP is a rare disease in adults but may have a severe course and poor renal outcome.

**Clinical Presentation.** Purpura is the leading symptom of HSP. In a recent Italian study purpura was present in all patients. Histologically, a leukocytoclastic vasculitis with extravasation of leukocytes and red cells, vessel wall damage, fibrinoid necrosis, and IgA deposition at the vessel wall and in the mesangium of the kidney can be found. Preceding upper respiratory tract infections are reported in more than 50% of children. A variety of bacterial and viral triggers, environmental stimuli, and host susceptibility factors such as autoinflammatory disease genes have been reported. HSP can occur before or during the course of systemic diseases such as antineutrophil cytoplasm antibody (ANCA) vasculitis or Crohn’s disease.

IgA appears to play a pivotal role in the pathogenesis of HSP. Abnormal glycosylation of the hinge region O-linked glycan of IgA1 has been implicated in the etiopathology of HSP. Abnormal IgA1 molecules were found to have a higher tendency to aggregate, interact with IgG, and form IgA-IgG complexes and deposits in the kidney. Similarly, serum levels of galactose-deficient IgA1 are elevated in Caucasian and Asian patients with IgA nephropathy. Schmitt demonstrated deposits of IgA-binding streptococcal M protein in the skin and kidney of HSP patients directly linking infection and vasculitis. Wu studied the role of leukotrienes and suggested that abnormally high levels of leukotriene B4 and lipoxin A4, potent activators of neutrophils, found in HSP patients with nephritis differentiate them from those without nephritis.

**Table 108-8** European League Against Rheumatism/Pediatric Rheumatology European Society Endorsed Consensus Criteria for the Classification of Childhood Vasculitides

<table>
<thead>
<tr>
<th>Vasculitis Type</th>
<th>Classification Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predominant Small Vessel Vasculitis</td>
<td></td>
</tr>
<tr>
<td>Henoch-Schönlein purpura</td>
<td>Palpable purpura (mandatory criterion) plus at least one of:</td>
</tr>
<tr>
<td></td>
<td>- Diffuse abdominal pain</td>
</tr>
<tr>
<td></td>
<td>- Any biopsy showing predominant IgA deposition</td>
</tr>
<tr>
<td></td>
<td>- Arthritis or arthralgia (acute, any joint)</td>
</tr>
<tr>
<td></td>
<td>- Renal involvement (any hematuria and/or proteinuria)</td>
</tr>
<tr>
<td>Childhood granulomatosis with polyangiitis</td>
<td>At least 3 of the following 6 criteria must be present:</td>
</tr>
<tr>
<td></td>
<td>- Abnormal urinalysis (hematuria and/or significant proteinuria)</td>
</tr>
<tr>
<td></td>
<td>- Granulomatous inflammation on biopsy (if a kidney biopsy is done it</td>
</tr>
<tr>
<td></td>
<td>characteristically shows pauci-immune necrotizing glomerulonephritis)</td>
</tr>
<tr>
<td></td>
<td>- Nasal sinus inflammation</td>
</tr>
<tr>
<td></td>
<td>- Subglottic, tracheal, or endobronchial stenosis</td>
</tr>
<tr>
<td></td>
<td>- Abnormal chest radiograph or computed tomography</td>
</tr>
<tr>
<td></td>
<td>- Proteinase 3 ANCA or c-ANCA staining (sensitivity specificity calculated for any</td>
</tr>
<tr>
<td></td>
<td>positive ANCA)</td>
</tr>
<tr>
<td>Childhood microscopic polyangiitis</td>
<td>No proposed criteria; for description see text</td>
</tr>
<tr>
<td>Childhood Churg-Strauss syndrome</td>
<td>No proposed criteria; for description see text</td>
</tr>
<tr>
<td>Predominant Medium Vessel Vasculitis</td>
<td></td>
</tr>
<tr>
<td>Childhood (systemic) polyarteritis nodosa</td>
<td>Biopsy evidence of necrotizing vasculitis or angiographic abnormalities (mandatory</td>
</tr>
<tr>
<td></td>
<td>criterion) plus at least one of:</td>
</tr>
<tr>
<td></td>
<td>- Skin involvement (livedo reticularis, tender subcutaneous nodules, other vasculitic</td>
</tr>
<tr>
<td></td>
<td>lesions, superficial or deep infarctions)</td>
</tr>
<tr>
<td></td>
<td>- Myalgia or muscle tenderness</td>
</tr>
<tr>
<td></td>
<td>- Systolic/diastolic hypertension (&gt;95th percentile)</td>
</tr>
<tr>
<td></td>
<td>- Mononeuropathy or polyneuropathy</td>
</tr>
<tr>
<td></td>
<td>- Renal involvement (proteinuria, hematuria, impaired renal function)</td>
</tr>
<tr>
<td>Childhood cutaneous polyarteritis nodosa</td>
<td>No proposed criteria; for description see text</td>
</tr>
<tr>
<td>Predominant Large Vessel Vasculitis</td>
<td></td>
</tr>
<tr>
<td>Childhood Takayasu’s arteritis</td>
<td>Angiography of the aorta or its main branches and pulmonary arteries showing</td>
</tr>
<tr>
<td></td>
<td>aneurysms/dilatation, occlusion or thickened arterial wall not due to</td>
</tr>
<tr>
<td></td>
<td>fibromuscular dysplasia (mandatory criterion) plus at least 1 of:</td>
</tr>
<tr>
<td></td>
<td>- Pulse deficit or claudication</td>
</tr>
<tr>
<td></td>
<td>- Blood pressure discrepancy (&gt;10 mm Hg)</td>
</tr>
<tr>
<td></td>
<td>- Bruits</td>
</tr>
<tr>
<td></td>
<td>- Hypertension</td>
</tr>
<tr>
<td></td>
<td>- Acute-phase reactant (erythrocyte sedimentation rate &gt;20 mm/hr, C-reactive protein</td>
</tr>
<tr>
<td></td>
<td>abnormal)</td>
</tr>
</tbody>
</table>

ANCA, antineutrophil cytoplasm antibody; c-ANCA, cytoplasmic antineutrophil cytoplasm antibody.
cases, arthritis/arthralgias in 74%, abdominal symptoms in 51% including intussusception in 0.6%, renal disease in 54% including severe nephropathy in 7%, and acute renal failure in 2%.\textsuperscript{272} Scrotal edema was reported in 13%. Correspondingly, Dolezalova and colleagues\textsuperscript{270} described purpura present in 100% of Czech HSP patients; arthritis/arthralgia in 52%; abdominal pain and/or GI bleeding in 40%; hematuria/proteinuria in 15%; and genital involvement in 2.8%. Eye findings including anterior uveitis can be seen in HSP patients.\textsuperscript{274}

Skin disease in HSP has a characteristic appearance: Petechial or palpable purpuric lesions are located on dependent areas including lower legs and feet, buttocks, and arms. Lesions can have different sizes and stages ranging from fresh petechial rash to confluent bruises. An associated edema is commonly found, hands and feet appear puffy, and scrotal edema may be present in boys. Children younger than 2 years have been reported to have more significant edema. Lesions occur in waves, and skin disease in HSP is reported to last from 4 to 8 weeks.\textsuperscript{272}

HSP arthritis is often painful, nonerosive, and nonmigratory. Ankles, knees, hands, and wrists are most commonly inflamed. Arthralgias are found in a similar distribution. GI symptoms are common in HSP patients including intermittent abdominal discomfort, pain, and vomiting. Abdominal complications including intussusception are rare; however, they always have to be considered when a child presents with HSP features and complains of abdominal pain and possibly associated bloody stools. Oftentimes bowel wall thickening on ultrasound is detected.

Overall renal disease in HSP occurs in 40% to 50% of children and manifests itself as microscopic hematuria or low-grade proteinuria, which completely resolves in the vast majority.\textsuperscript{275} Older children may be at higher risk for nephritis.\textsuperscript{276} Overall progression to end-stage renal disease occurs in 1% to 3% of children.\textsuperscript{277} Renal biopsies are done in children with renal compromise and histologically demonstrate IgA nephropathy.\textsuperscript{278} The degree of damage on renal biopsy and the degree of proteinuria predicts poor outcome.\textsuperscript{279} Reported 10-year renal survival rates for children undergoing renal biopsies for HSP ranged from 73% to 90%.\textsuperscript{279,280} Though uncommon in HSP, overall IgA nephropathy and HSP nephritis represent the most common chronic glomerulonephritis in childhood.\textsuperscript{277}

**Diagnosis and Diagnostic Tests.** Children with HSP may have a raised erythrocyte sedimentation rate (ESR) (57%), elevated serum IgA (37%), and proteinuria (42%).\textsuperscript{272} All children require serial urinalyses. Jauhola and colleagues\textsuperscript{276} demonstrated that HSP nephritis occurred on average 14 days after HSP diagnosis, and within 1 month in the majority of cases. The risk of developing HSP nephritis after 2 months was 2%. Laboratory tests or blood pressure measurement at onset did not predict the occurrence of nephritis. Overall specific diagnostic markers of HSP are not readily available. Alternative complement pathway markers including activated C3 and C4 were reported in children and adults and may be associated with disease progression.\textsuperscript{275} Urinary proteomic patterns and serum levels of galactose-deficient IgA1 are promising and are currently being studied.\textsuperscript{281}

Skin biopsies are done to confirm the diagnosis of HSP and to exclude differential diagnoses. Overlapping clinical features may be found in infection, inflammation or medication associated leukocytoclastic vasculitis, rheumatic fever, poststreptococcal glomerulonephritis, lupus, and systemic vasculitis.\textsuperscript{282,283} Renal biopsies are performed in a select group of HSP patients.\textsuperscript{275}

**Treatment.** In most children HSP is a benign disease and does not require specific therapy. There is significant variation demonstrated for inpatient therapy and evaluation of children with HSP, which may contribute to varying clinical outcomes.\textsuperscript{284} Immunosuppressive therapy of HSP targets severe disease presentations including nephritis and gastrointestinal vasculitis. In 2006, Ronkainen and co-workers published a randomized placebo-controlled trial demonstrating that prednisone reduces the severity of joint and abdominal pain, while having no effect on purpura, prevention of nephritis, or recurrence of HSP.\textsuperscript{285} A recent retrospective cohort study of 1895 children discharged with HSP between 2000 and 2007 from 36 tertiary care children's hospitals in the United States determined that early corticosteroid treatment was associated with significantly less abdominal surgery, endoscopy, and abdominal imaging during hospitalization suggesting a protective effect of corticosteroid therapy for abdominal complications of HSP.\textsuperscript{286} In contrast, a prospective study from Finland reported that corticosteroids although alleviating clinical symptoms, did not alter the clinical course of HSP during 6 months of follow-up. Prednisone prophylaxis did not affect the timing of the appearance of nephritis.\textsuperscript{287} The addition of cyclophosphamide did not show a benefit in adults with HSP nephritis in a small trial.\textsuperscript{289} A stepwise approach is often used, including the use of nonsteroidal anti-inflammatory medication for mild HSP, corticosteroids for moderate to severe HSP, and addition of angiotensin-converting enzyme (ACE) inhibitors for nephritis. Evidence is emerging that treatment with high-dose IV pulse methylprednisolone coupled with azathioprine or cyclophosphamide may be beneficial in patients with severe nephritis.\textsuperscript{288} Cyclosporine also has been successfully used for severe nephritis.\textsuperscript{289}

**Outcome.** In the majority of children HSP is a self-limiting disease, which resolves within 4 to 6 weeks.\textsuperscript{258} Patients with microscopic hematuria and trivial proteinuria have an excellent prognosis.\textsuperscript{280} In contrast, 30% of pediatric HSP patient with nephritis will have renal impairment and 5% will develop end-stage renal disease.\textsuperscript{289} Recurrence of HSP is seen in a third of patients; symptoms resolve within 4 to 6 weeks in the majority of patients. Children older than 8 years of age and those with nephritis are significantly more likely to experience recurrences.\textsuperscript{289}

**Antineutrophil Cytoplasm Antibody Vasculitis**

The group of childhood systemic vasculitides associated with ANCA include granulomatosis with polyangiitis (GPA; formerly known as Wegener's granulomatosis), microscopic polyangiitis (mPA), and Churg-Strauss syndrome (CSS). The conceptual framework provided by the Chapel Hill Consensus Conference on the Nomenclature of Systemic Vasculitis is commonly accepted for childhood ANCA vasculitis. Necrotizing small vessel vasculitis of venules, capillaries, arterioles, and small arteries is the hallmark of ANCA vasculitis. Pathogenetic studies were not primarily done in children and are therefore not
covered here. Pauci-immune necrotizing and crescentic glomerulonephritis, as well as hemorrhagic pulmonary capillaritis, frequently present as pulmonary-renal syndrome. Limited phenotypes including pauci-immune glomerulonephritis with ANCA and nonrenal, upper respiratory tract, or limited GPA are commonly included in the group of ANCA vasculitides of children and adults. These diseases are rare in childhood. However, recent collaborative efforts such as “A Registry for Childhood Vasculitis (ARCHiVe)” have substantially increased the knowledge and understanding of childhood ANCA vasculitis.

**Definition and Classification**

**Granulomatosis with Polyangiitis.** The proposed EULAR/PReS-endorsed consensus criteria for GPA are shown in Table 108-8. The sensitivity as determined in the validation process described earlier was found to be 93.3%, and the specificity was 99.2%, when including evidence of any ANCA. Cabral and colleagues have recently validated the proposed criteria in a large multicenter cohort of children with systemic vasculitis (no HSP) and determined a diagnostic sensitivity of 73.6% and specificity of 73.2%.

**Microscopic Polyangiitis.** In children, MPA is a rare diagnosis. No formal classification criteria were proposed for MPA in children by the “EULAR conference expert group.” Instead, the existing Chapel Hill description was modified by formally adding ANCA.

**Churg-Strauss Syndrome.** CSS (allergic granulomatosis) is even less common than MPA in children. No pediatric classification criteria exist, and therefore adult ACR criteria are commonly applied. The pediatric literature is limited to case reports and small series. A recent literature review identified a total of 33 children with systemic vasculitis (no HSP) and determined a diagnostic sensitivity of 73.6% and specificity of 73.2%.

**Clinical Presentation**

The clinical presentation of ANCA vasculitis is widely overlapping. In children, GPA and MPA patients are commonly reported in combined series. In the largest single-center cohort of 25 children, the median duration of symptoms before establishing the diagnosis was 2 months. Children presented most frequently with constitutional symptoms (96%) and glomerulonephritis (88%) with renal failure in half. Recovery from renal failure was uncommon (1/11). Upper airway disease was present in 84% including one child with subglottic stenosis. Overall, 80% had pulmonary involvement at diagnosis, most commonly nodules (44%) and pulmonary hemorrhage (44%). Five children with pulmonary hemorrhage required ventilation, and four children had venous thrombotic events. The ARChiVe group reported 117 pediatric patients including GPA (n = 76), microscopic polyangiitis (n = 17), ANCA-positive pauci-immune glomerulonephritis (n = 5), CSS (n = 2), and unclassified vasculitis (n = 17). In the 65 of 76 who met ACR criteria for GPA, the median interval from symptom onset to diagnosis was 2.7 months (range, 0 to 49 months). The most frequently presenting features by organ system were constitutional (89%); pulmonary (80%); ear, nose, and throat (80%); and renal (75%). Zwerina and co-workers suggested that children compared with adults with CSS had a predominance of cardiopulmonary disease manifestations, less peripheral nerve involvement, and higher mortality.

**Diagnosis and Diagnostic Tests**

Raised inflammatory markers including ESR and CRP, leukocytosis, and positive ANCAAs are commonly found at diagnosis of ANCA vasculitis. Aikikusa and colleagues documented ANCA positivity in 22 of 25 children. Cabral and co-workers identified cytoplasmic ANCA (c-ANCA) positivity in 66% of pediatric GPA patients, 93% of whom were anti-PR3 positive on enzyme-linked immunosorbent assay. Accordingly, 22% were perinuclear ANCA (p-ANCA) positive, of which 21% had anti-PR3 and 57% anti-MPO specificity.

Endothelial cell markers including von Willebrand factor antigen, antiendothelial cell antibodies, and circulating endothelial cells are potential biomarkers. Disease activity and damage tools are currently being developed based on modifications of the available adult vasculitis measures.

**Epidemiology**

Childhood ANCA vasculitis is a rare group of diseases. In Southern Alberta, Canada, the average annual incidence of GPA in children was estimated at 2.75 cases per million per year, with a steep increase over the past 5 years to 6.39 cases per million per year. A recent study from Japan reported the adult MPA incidence at 14.8 per million, whereas GPA was present in 2 per million adult patients. Within the group of ANCA vasculitis, MPA may be more common than GPA in adults, while it is definitely less common in children. However, Reinhold-Keller and associates suggested that the incidence of GPA of all age groups was two to three times greater than those of MPA and CSS. There was no regional difference in incidence rates found. Girls are consistently more commonly affected than boys; the male-to-female ratio is reported to be 1:3 to 4.
primarily treated with MTX and corticosteroids. In contrast, the initial treatment of CSS in children was corticosteroid monotherapy in 76% (n = 33), while only 24% received combination immunosuppressive therapy. In a retrospective series, Wright and Dillon reported that therapeutic plasma exchange may be of benefit during the acute phase of childhood systemic vasculitis. New drugs including rituximab and other biologic therapies are increasingly used in children. Supportive therapy regimens such as ACE inhibitors and antibiotic prophylaxis vary and have not been systematically studied in children with ANCA vasculitis.

Outcome

ANCA vasculitis is a relapsing/remitting chronic disease in children and adults. Phillip and Luqmani calculated overall 5-year survival rates for GPA of all ages at 75%, MPA at 45% to 75%, and CSS at 68% to 100% at 5 years. The pediatric CSS literature review captured six deaths (18%), all related to the underlying disease, occurring after a mean disease duration of 14 months. Mortality rates may be falling as a result of more effective intervention, but they remain elevated substantially in severe disease. Early deaths were attributed to active disease (multiorgan failure, infection). Late deaths reflected the cumulative burden of disease and treatment in addition to comorbidities.

Disease activity has consistently been reported to be high at diagnosis of ANCA vasculitis in children. Risk factors associated with mortality or morbidity remain to be identified in children.

Medium-Sized Vessel Vasculitis

KD, polyarteritis nodosa (PAN), and cutaneous PAN are characterized by medium-sized muscular artery inflammation. KD is the most common cause of acquired heart disease in Western countries. Vessel wall histology on autopsies was recently reported to demonstrate a proliferative, granulomatous process with accumulation of monocytes/macrophages and rare occurrence of fibrinoid necrosis. KD is a monophasic inflammatory disease that leads to transient coronary artery lesions in 20% to 25% of children, while coronary artery damage (e.g., aneurysms, stenoses) is found in 5 to 8% of treated patients. In contrast, PAN is a rare disease, with difficult-to-control chronic disease activity, and high morbidity and mortality rates. Biopsy specimens demonstrate systemic transmural necrotizing vasculitis. Cutaneous PAN has a similar histology; however, in general it has a less severe disease course.

Kawasaki Disease

Kawasaki disease (KD) is better classified as an inflammatory “syndrome.” Infectious triggers and possibly other environmental factors lead to a stereotypical inflammatory process in a susceptible host. Host susceptibility clearly differs between ethnic groups. This may be related to genetic polymorphisms of proinflammatory genes such as the signal transduction caspase-3 gene CASP3 or the ITPKC gene, which encodes a regulator of T cell activation. Genes involved in vascular remodeling such as matrix metalloproteinases may confer an increased risk of vessel damage and aneurysm formation.

Definition and Classification. The diagnosis of KD remains grounded in recognition of a clinical pattern: Children, in whom the diagnosis of typical KD is made, present with a minimum of 5 days of fever plus at least four of five criteria including oral changes of cracked lips/strawberry tongue (as seen in 94%), bilateral nonpurulent conjunctivitis (92%), rash (90%), erythema and/or swelling of hands and/or feet (77%), and cervical lymphadenopathy (64%).

The diagnosis of KD is more challenging in children with incomplete clinical features. In 2004 the American Heart Association (AHA) proposed an algorithm for diagnosing and treating suspected incomplete KD. Yellen and coworkers tested the performance in a retrospective multicenter series of 195 patients with KD and coronary artery aneurysms. The authors demonstrated that applying the AHA algorithm would have significantly increased the rate of IVIG treatment from 70% (classic KD criteria) to 97% and possibly prevented aneurysms. Similarly, Heucin and colleagues recently demonstrated a significantly increased detection rate of KD—in particular incomplete KD with coronary lesion—when applying the AHA algorithm.

Epidemiology. KD primarily affects young children; 80% of cases occur in children younger than 5 years of age. Boys are more commonly affected than girls; the male-to-female ratio is reported to be 1.4 to 1.9:1. Recurrence rates of KD are estimated at 3%. Atypical KD is more common in children younger than 1 year or older than 9 years of age, accounting for one-third of KD diagnoses in these age groups.

Incidence rates clearly vary between ethnic groups. Asian children are at highest risk: In 2010, Park and colleagues recently reported an average annual incidence rate of KD in Korea of 113.1 per 100,000 in children younger than 5 years. In Japan, the annual incidence rate was even higher at 218.6 per 100,000 children younger than 5 years of age. Around the same time, in the rest of the world annual KD incidence rates were reported between 5 and 13 per 100,000 children younger than 5 years.

Clinical Presentation. Fever in KD patients is typically continuous. It is reported to be either absent or less prominent or consistent in children younger than 1 year of age and older than 9 years of age. Eye findings include bilateral nonpurulent conjunctivitis, which is particularly prominent with fever. Other inflammatory eye findings including asymptomatic uveitis have been reported. Rash of all types can be associated with KD. The rash is also more prominent with fever. Frequently it is confluent in the diaper area and axilla in the acute phase. Skin peeling classically starts on the fingertips in the subacute phase. Blisters are uncommon. Oral changes include dry, red, and cracked lips; prominent follicles of the tongue (strawberry tongue); and an oral anaphema. Aphthous ulcers can be present, primarily when associated with a triggering herpes-group virus infection. Cervical lymphadenopathy is frequently asymmetric. The criteria state they should be 1.5 cm or greater. Nodes can be tender, and a secondary lymphadenitis can occur, which may require additional therapy. Hands and feet frequently appear puffy and erythematous.
Neurologic symptoms are common in children with KD: The majority of toddlers are extremely irritable. Often children are withdrawn, lethargic, and clingy or complain of headaches, in particular with fever. Transient hearing loss can be present in a significant number of patients, most commonly sensorineural hearing loss (20 to 35 dB). It may be related to salicylate toxicity in some children. Persistent hearing loss is rare.339

Other organ manifestations include acute arthritis,430 hepatitis,341 gallbladder hydrops, intussusception, or pseudo-obstruction presenting as acute abdomen,342 dysuria with sterile pyuria, genitai swelling, and muscle pain and weakness.

**Diagnosis and Diagnostic Tests.** Inflammatory markers are commonly raised in children with KD. ESR and CRP, leukocytosis, anemia, mildly raised liver function tests, and low albumin levels are expected in children with acute KD.343,344 Laboratory markers are included in the AHA algorithm329 and have been used to predict adverse outcome.344

Because infections are commonly found in children with KD, an infectious workup is mandatory to detect concurrent ongoing infections that may require additional therapy. Cardiac evaluation in KD patients includes a chest radiograph, electrocardiogram, and echocardiography. DeZorzi and associates defined body-surface area–adjusted standards (z scores) for coronary artery abnormalities on echocardiography.345 These scores have subsequently been used to establish a classification system for the entire spectrum of coronary artery abnormalities including aneurysms.346 Follow-up echocardiography is required at 2 weeks in children with evidence of coronary damage and at 6 weeks in all children because vascular disease commonly peaks at 2 to 4 weeks.347

A severe complication of KD is increasingly recognized: macrophage activation syndrome (MAS).348 Latino and co-workers reported 12 of 638 KD patients who developed clinical and laboratory features of MAS including hepatosplenomegaly, cytopenia in two or more cell lines, hyperferritinemia and elevated hepatic enzymes, hypertriglyceridemia and/or hypofibrinogenemia, increased p-dimers, and evidence of hemophagocytosis on biopsy (4/12). Early recognition of MAS and increased immunosuppression are crucial.

**Treatment.** The first-line treatment for children with KD is IVIG at a dose of 2 g/kg.349 In addition, high-dose acetylsalicylic acid (ASA) at 30 to 50 mg/kg or 80 to 100 mg/kg is frequently used as an antipyretic and anti-inflammatory drug while the child is febrile. However, a metaanalysis did not find sufficient evidence for ASA treatment in KD.350 Conceptually, there is strong support for the use of ASA in KD because IVIG and ASA were found to differentially modulate the expression of TNF and its downstream effects in the KD animal model.351 Importantly, low-dose ASA has an antithrombotic effect by inhibiting the production of thromboxane A2 and prostacyclin in platelets.

Recurrence of fever after a dose of IVIG is commonly treated with a repeat dose of IVIG.352 Children with refractory KD, defined as failure to respond to IVIG retreatment commonly receive corticosteroid therapy. A trial exploring the efficacy of early corticosteroids for primary KD therapy in addition to IVIG did not demonstrate a significant benefit.349 In nonresponders TNF inhibitors have been used.353,354 Son and co-workers reported that patients treated with infliximab had a faster resolution of fever and fewer days of hospitalization.355 Abciximab is a monoclonal antibody against glycoprotein (GP)IIb-IIIa on platelets. A small study of 18 children with KD and large coronary artery aneurysms suggested that abciximab treatment may be associated with improved vascular remodeling.356

Children with KD require cardiology follow-up at 6 weeks including clinical assessment and echocardiography. Commonly, low-dose ASA treatment is discontinued in all children in whom the coronary artery lesions have resolved. Children with evidence of coronary disease at 6 weeks require long-term care. In many centers asymptomatic KD patients are reassessed at 12 months and then discharged.

**Outcome.** The 5-year survival of children with KD is excellent at greater than 99%.318 However, one in 20 children with KD will develop permanent damage to their coronary arteries.357,358 Children may develop vascular aneurysms and stenoses beyond the coronary arteries (Figure 108-6). Early interventions including stenting or coronary bypass operations may be required.359 Even asymptomatic children with KD and aneurysms are at high risk of myocardial lesions.360 The long-term impact of “transient” coronary artery dilatations/ectasia remains to be determined.361 KD may lead to endothelial dysfunction and premature atherosclerosis.362 The psychosocial impact of KD was recently explored: Parents of KD patients report significant distress and anxiety even years after the acute illness.363

**Polyarteritis Nodosa**

Polyarteritis nodosa (PAN) is a rare necrotizing vasculitis of medium-sized vessels.321,364 The focal/segmental, transmural necrosis can lead to aneurysm formation. Classically lesions heal and scar with a palpable fibrotic nodule (nodosa). The disease. Gadolinium-enhanced, reconstructed magnetic resonance imaging angiography demonstrates multiple irregular aneurysms of the coronary arteries, the brachial and subclavian arteries, and both common iliac arteries and internal iliac arteries.352

**Figure 108-6** Giant aneurysms in a 6-month-old girl with Kawasaki disease. Gadolinium-enhanced, reconstructed magnetic resonance imaging angiography demonstrates multiple irregular aneurysms of the coronary arteries, the brachial and subclavian arteries, and both common iliac arteries and internal iliac arteries.
classic or systemic form can affect medium-sized arteries in multiple organs and typically presents with clinical and laboratory features of severe systemic inflammation. In contrast, cutaneous PAN is limited to the medium-sized arteries of the skin. It is more common and is characterized by periodic exacerbations often associated with Streptococcus infection.

**Definition and Classification.** The proposed EULAR/PReS-endorsed consensus criteria for (systemic) childhood PAN are shown in Table 108-8. The group modified the adult ACR PAN criteria by making biopsy evidence of necrotizing vasculitis or angiographic abnormalities a mandatory criterion.

The sensitivity as determined in the validation process described earlier was 89.6%, and the specificity was 99.6%. The mandatory criterion of biopsy evidence had the highest sensitivity and specificity (>80%). Peripheral neuropathy was the least sensitive (26%), and renal involvement was the least specific (37%) criterion. The consensus conference did not propose criteria for cutaneous PAN but recognized the need for a separate category. The disease characteristics were described as presence of subcutaneous nodular, painful, nonpruritic lesions with or without livedo reticularis with no systemic involvement except for myalgia, arthralgia, and nonerosive arthritis; biopsy evidence of necrotizing, nongranulomatous vasculitis; negative tests for ANCA; and an association with streptococcal infections.

**Epidemiology.** Systemic PAN is a rare disease. The overall incidence is estimated at 2 to 9 per million and varies among ethnicities. In a recent 5-year survey of all childhood vasculitis at 15 Turkish centers, PAN accounted for only 6% of cases. An association of systemic PAN with familial Mediterranean fever (FMF) was suggested. Associations with hepatitis B and other viruses with childhood PAN have been reported.

In 2004 an international PAN survey of 22 pediatric centers identified 110 children, of whom 63 children (57%) had systemic PAN and only 33 (30%) had cutaneous PAN. However, in many centers the number of children with cutaneous PAN is significantly higher than with systemic PAN. The association of cutaneous PAN with medications and systemic inflammatory/autoimmune diseases has been reported.

**Clinical Presentation.** Systemic inflammation often presents as fever, fatigue, and weight loss in children with systemic PAN. Decreased perfusion through medium-sized vessels can cause focal organ ischemia including severe abdominal pain, cardiac ischemia, muscle pain, skin infarction with gangrene, livedo reticularis, and renal hypertension. The severe focal vessel inflammation can present as cutaneous painful nodules often located on the calves or feet, focal muscle pain, peripheral or cranial neuropathy, and inflammatory CNS lesions among others. Necrotizing vascular inflammation in systemic PAN can lead to fragility of the medium-sized arterial vessel wall and significant hemorrhage. Renal disease in systemic PAN is classically renal hypertension due to segmental artery disease; however, small vessel involvement presenting as isolated proteinuria, nephritic or nephrotic syndrome, and renal failure are reported in a series of 26 children from Turkey.

Cutaneous PAN is characterized by the presence of deep skin nodules predominantly on the lower legs, which are frequently found at different stages of development. Most commonly a violaceous color or pigmentation with retiform appearance persists for months. Ulceration can be a complication. Pain, arthralgias/arthritis, malaise, and moderate fever are associated symptoms in children with cutaneous PAN.

**Diagnosis and Diagnostic Tests.** Inflammatory markers including ESR and CRP are commonly raised in children with active systemic and cutaneous PAN. Organ function parameter may be abnormal. Specific diagnostic markers for PAN are not available. Characteristic PAN skin biopsy features are identical to adult PAN (see Chapter 90). Characteristic angiography findings include aneurysms and stenoses of the medium-sized arteries.

**Treatment.** The treatment of childhood systemic PAN is based on adult studies and recommendations and is summarized in Chapter 90. Pediatric case reports and series supported the efficacy of corticosteroids, combination immunosuppression, and biologic therapies including TNF inhibitors and B cell depletion using rituximab. Addition of antiplatelet agents may be required. Children with cutaneous PAN are commonly treated with nonsteroidal anti-inflammatory medication or corticosteroids. Refractory patients require immunosuppressive combination therapy. Prophylactic antibiotics are considered in children with evidence of Streptococcus infections.

**Outcome.** The 1-year and 5-year survival rates of 26 Turkish children with systemic PAN was only 72.5% and 60%, respectively.

**Large Vessel Vasculitis**

Takayasu’s arteritis (TA) is the most common childhood large vessel vasculitis. Histologically, TA is indistinguishable from giant cell arteritis (GCA) in adults. TA and GCA may represent a disease spectrum rather than different entities. Both diseases are characterized by giant cells, which represent multinuclear cells formed by fusion of monocytes/macrophages and in TA can be found in the wall of large vessels including the aorta and its major branches.

**Definition and Classification**

The proposed EULAR/PReS-endorsed consensus criteria for TA are shown in Table 108-8. Angiographic abnormalities are a mandatory criterion. In addition, at least one of five criteria including pulse deficit or claudication, blood pressure discrepancy of greater than 10 mm Hg, bruits, hypertension, or elevated acute-phase reactant has to be present. The sensitivity and specificity as determined in the
The validation process described earlier were 100% and 99.9%, respectively.\footnote{257}

**Epidemiology**

The incidence of adult TA was recently reported at 0.8 per million in the United Kingdom.\footnote{379} It may vary between ethnicities with higher incidences in Asia,\footnote{380,381} Africa,\footnote{382} and Latin America.\footnote{383} No population-based pediatric data are available; however, children may account for up to 30% of patients in some studies.\footnote{384-386} Most commonly, childhood TA is diagnosed in adolescence. All series have a female predominance ranging from 14:1 to 1.4:1. Recent case series reported a total of 99 children with TA.\footnote{382,387-389} Associations of *Mycobacterium tuberculosis* and TA have consistently been suggested.\footnote{390}

**Clinical Presentation**

Children with TA often present with clinical signs of organ ischemia and systemic inflammation. Most commonly, headache or associated neurologic deficits such as strokes, seizures or syncope, chest or abdominal pain, claudication of extremities, fever, and weight loss are the presenting symptoms.\footnote{387,388} Examination frequently reveals hypertension, absent pulses, and bruits.

**Diagnosis and Diagnostic Tests**

The diagnosis of TA is based on clinical pattern recognition and confirmation by angiography. Inflammatory markers including CRP and ESR have limited sensitivity for active TA;\footnote{391} no disease-specific markers have been identified. Hoffman and Ahmed demonstrated that there is a poor correlation between serum markers and vascular histopathology in adult TA.\footnote{392} Angiography is the cornerstone of diagnosing and monitoring TA (see Figure 108-8A,B). Inflammatory arterial wall disease presents as arterial wall thickening, vessel stenosis, occlusion, or rarely aneurysms.\footnote{393} Different vascular imaging modalities are used in TA, each with distinct advantages and limitations.\footnote{394,395} Conventional angiography provides information about blood flow, perfusion pattern, collateralization, and degree of vessel stenosis. It reliably identifies clots or low-flow vessel segments posing a risk for subsequent artery to artery embolisms. Magnetic resonance angiography (MRA) is noninvasive and provides information about the characteristics of the vessel wall including thickening, contrast enhancement, and surrounding soft tissue inflammation.\footnote{396} CT angiography may provide similar information as MRA; however, the associated radiation exposure often limits its use in children. Recent studies highlighted the utility of 18F-fluorodeoxyglucose (18FDG) positron emission tomography scan.\footnote{397} Doppler ultrasound correlates well with angiography in delineating homogenous wall thickening in the aorta and its branches. It may be a promising tool for childhood TA.\footnote{398}

**Treatment**

Corticosteroids are the cornerstone of medical TA treatment in adults and children.\footnote{314} Combination
immunosuppression with cyclophosphamide was found to be effective in controlling disease activity in childhood TA. MTX was suggested in adult TA studies. Refractory childhood TA has been successfully treated with biologic therapies, primarily TNF inhibitors.

Surgical treatment includes stenting, dilatation, and bypass surgery. Corresponding to adult TA recommendations, the best time for vascular intervention in children is in inactive disease. A close collaboration between all treating subspecialties is mandatory.

**Outcome**

The overall outcome of TA is poor. The 5-year survival rate was reported to be 70% to 93%. Maksimowicz-McKinnon and colleagues gave a guarded prognosis of 93% attaining remission but only 28% having sustained remission of at least 6 months’ duration after prednisone was tapered to less than 10 mg daily. Angioplasty and vascular surgery were initially successful. Restenosis occurred in 78% of angioplasty and 36% of bypass/reconstruction procedures. More than two-thirds of TA patients had difficulty performing routine daily activities, and one-fourth were unable to work. Jales-Neto and co-workers recently determined that patients with childhood-onset TA had a significantly lower frequency of disease remission compared with adult-onset TA (24% vs. 56%) and more aneurysms (41% vs. 11%).

**Central Nervous System Vasculitis**

Childhood CNS vasculitis is an increasingly recognized inflammatory brain disease. CNS vasculitis is classified as secondary, when it occurs in association with a systemic illness including infection, malignancy, or rheumatic disease such as a systemic vasculitis (Table 108-9). Childhood primary CNS vasculitis solely affects the vessels of the brain and spinal cord.

**Definition and Classification**

The diagnosis of childhood primary CNS vasculitis is based on the modified Calabrese criteria for primary angiitis of the CNS (PACNS), which mandate (1) a newly acquired focal or diffuse neurologic deficit or a psychiatric manifestation in a patient 18 years of age or younger and (2) angiography and/or brain biopsy evidence of CNS vasculitis in the absence of a systemic underlying condition known to cause or mimic CNS vasculitis. Childhood PACNS is further subdivided into angiography-positive cPACNS and angiography-negative, small vessel cPACNS (SVcPACNS), with the latter being confirmed on elective brain biopsies.

**Epidemiology**

The incidence of childhood CNS vasculitis is unknown. New clinical phenotypes continue to be recognized: Angiography-positive CNS vasculitis was found to be the underlying process in a large subgroup of vascular strokes, a condition neurologists may diagnose as transient cerebral arteriopathy (TCA). Recently, new clinical phenotypes of cPACNS have been described including refractory seizure status, movement disorder, and optic neuritis. Children with cPACNS may be diagnosed as “atypical” demyelinating disease.
Clinical Presentation

Children with cPACNS can present with any focal or diffuse neurologic deficits or psychiatric symptoms. Children with angiography-positive disease typically present with headaches and strokes including acute hemiparesis, facial droop, hemisensory deficits, fine motor deficits, or dysphasia. Additional seizures and severe cognitive dysfunction are more commonly seen in progressive cPACNS.

Children with angiography-negative small vessel cPACNS may present with systemic features including fever, malaise, arthralgia, and flulike symptoms, associated with headache, neurocognitive dysfunction, behavior changes, or intractable seizures. Focal neurologic deficits, optic neuritis, and myelitis can be presenting features. Previously healthy children may have developed a rapid neurologic deterioration and present with an acute encephalitis or may have had subacute progression of symptoms such as worsening seizures or behavior changes over weeks to months.404

Diagnosis and Diagnostic Tests

The suspected diagnosis of cPACNS mandates a thorough evaluation. Systemic illnesses and other inflammatory and noninflammatory brain diseases have to be considered.404 A diagnostic algorithm was recently proposed (Figure 108-9). Inflammatory markers can be elevated in children with cPACNS, most commonly in the small vessel subtype. Von Willebrand factor antigen appears to correlate with clinical disease activity. Cerebrospinal fluid (CSF) analysis frequently reveals a mild pleocytosis with predominantly lymphocytes and occasionally elevated CSF protein. Frequently the opening pressure is raised.

The absence of laboratory markers does not exclude cPACNS. In fact, children with angiography-positive, nonprogressive cPACNS frequently have normal inflammatory markers. In contrast, angiography-positive progressive cPACNS and small vessel cPACNS patients commonly present with laboratory signs of inflammation. Serial testing may be required. Oligoclonal banding is found in children with confirmed small vessel cPACNS.409

MRI studies identify ischemic, diffusion-restricted lesions and inflammatory, fluid-attenuated inversion recovery (FLAIR)-positive parenchymal lesions.410 MRA and conventional angiography characterize vascular stenoses411,412 (Figure 108-10A,B). Gadolinium-enhanced MRA sequences can demonstrate vessel wall enhancement and thickening in more than 85% of adult and pediatric patients with active cPACNS.413

Brain biopsies confirm the diagnosis of angiography-negative cPACNS409 (Figure 108-11). In contrast to adults, biopsies are not required in children with angiography-positive disease because the diagnosis is not confounded by arteriosclerosis. Elective brain biopsies can be lesional as determined by MRI or nonlesional in the nondominant hemisphere.409 The diagnostic yield is greater than 90% in

---

**Table 108-9** Classification of Childhood Primary and Secondary Central Nervous System (CNS) Vasculitis

<table>
<thead>
<tr>
<th>Childhood Primary CNS Vasculitis (cPACNS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiography-positive, nonprogressive cPACNS (NPCPACNS)</td>
</tr>
<tr>
<td>Angiography-positive, progressive cPACNS (PCPACNS)</td>
</tr>
<tr>
<td>Angiography-negative, small vessel cPACNS (SVcPACNS)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Secondary CNS Vasculitis in Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infection or postinfectious</td>
</tr>
<tr>
<td>Bacterial infection</td>
</tr>
<tr>
<td>Mycobacterium tuberculosis</td>
</tr>
<tr>
<td>Mycoplasma pneumoniae</td>
</tr>
<tr>
<td>Streptococcus pneumoniae</td>
</tr>
<tr>
<td>Treponema pallidum</td>
</tr>
<tr>
<td>Spirochete infection</td>
</tr>
<tr>
<td>Borrelia burgdorferi</td>
</tr>
<tr>
<td>Viral infection</td>
</tr>
<tr>
<td>Cytomegalovirus</td>
</tr>
<tr>
<td>Enterovirus</td>
</tr>
<tr>
<td>Epstein-Barr virus</td>
</tr>
<tr>
<td>Hepatitis C virus</td>
</tr>
<tr>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>Influenza virus</td>
</tr>
<tr>
<td>JC virus (progressive multifocal leukoencephalopathy)</td>
</tr>
<tr>
<td>Parvovirus B19 virus</td>
</tr>
<tr>
<td>Varicella zoster virus</td>
</tr>
<tr>
<td>West Nile virus</td>
</tr>
<tr>
<td>Fungal infection</td>
</tr>
<tr>
<td>Actinomycosis</td>
</tr>
<tr>
<td>Aspergillus</td>
</tr>
<tr>
<td>Candida albicans</td>
</tr>
<tr>
<td>Rheumatic disease</td>
</tr>
<tr>
<td>Collagen vascular diseases</td>
</tr>
<tr>
<td>Behçet’s disease</td>
</tr>
<tr>
<td>Juvenile dermatomyositis</td>
</tr>
<tr>
<td>Morphea (en coup de sabre)</td>
</tr>
<tr>
<td>Sjögren syndrome</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>Systemic vasculitides</td>
</tr>
<tr>
<td>Kawasaki disease</td>
</tr>
<tr>
<td>Henoch-Schönlein purpura</td>
</tr>
<tr>
<td>Microscopic polyarteritis</td>
</tr>
<tr>
<td>Granulomatosis with polyangiitis</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>Hemophagocytic</td>
</tr>
<tr>
<td>Lymphohistiocytosis</td>
</tr>
<tr>
<td>Mitochondrial diseases</td>
</tr>
<tr>
<td>Drug-induced central nervous system vasculitis</td>
</tr>
<tr>
<td>Hemoglobinopathies</td>
</tr>
<tr>
<td>Malignancy</td>
</tr>
<tr>
<td>Radiation</td>
</tr>
</tbody>
</table>

---

**Figure 108-9** Proposed diagnostic algorithm for children with suspected central nervous system vasculitis. CSF, cerebrospinal fluid; CRP, C-reactive protein; DWI, diffusion-weighted imaging; ESR, erythrocyte sedimentation rate; FLAIR, fluid-attenuated inversion recovery; MR, magnetic resonance; MRI, magnetic resonance image; OP, opening pressure; vWF, von Willebrand factor.
Treatment

No randomized controlled trials are available for CNS vasculitis in adults and children. A recent prospective observational cohort study evaluated the efficacy and safety of a 24-month induction-maintenance protocol for small vessel cPACNS, demonstrating a full neurologic recovery in two-thirds of children. Case reports and series describe the efficacy of other immunosuppressive treatments for different types of cPACNS. In children with nonprogressive cPACNS, adjunctive corticosteroids may prevent recurrent ischemic events and improve neurologic outcome.

Outcome

There is limited information about the long-term outcome of children with CNS vasculitis. In angiography-positive cPACNS, two-thirds have a monophasic, nonprogressive course. These children often present with large ischemic lesions due to proximal large vessel stenosis and ischemia in the vascular territory. Although progression of inflammation and involvement of other vascular beds is limited in this group, the associated neurologic deficit often exceeds the other subtypes characterized by progression of inflammation. Inflammation is reversible when recognized and treated early, as recently demonstrated in the small vessel cPACNS study. Disease flares are seen in a significant number of patients. Prospective collaborative studies are ongoing to further characterize the long-term outcome of children with CNS vasculitis worldwide.

Selected References


References


56. Wu FQ, Zhao Q, Cui XD, Zhang W: C1q and anti-C1q antibody levels are correlated with disease severity in Chinese pediatric systemic lupus erythematosus, Rheumatol Int 31:501–505, 2011.


59. Wu FQ, Zhao Q, Cui XD, Zhang W: C1q and anti-C1q antibody levels are correlated with disease severity in Chinese pediatric systemic lupus erythematosus, Rheumatol Int 31:501–505, 2011.


186. Dalakas MC, Illa I, Dambrosia JM, et al: A controlled trial of high-dose intravenous immune globulin infusions as treatment for derma-


hood: clinical and immunologic features of 153 patients in an in-


Bacterial infections of the joint are usually curable with treatment, but morbidity and mortality are still significant in patients with underlying rheumatoid arthritis (RA), patients with prosthetic joints, elderly patients, and patients who have severe and multiple comorbidities. Goldenberg wrote in 1994, “Treatment and outcome [of septic arthritis] have not improved substantially over the past 20 years.” This statement is probably still true today. Incremental knowledge of the pathogenesis of septic arthritis caused by two common organisms, Neisseria gonorrhoeae and Staphylococcus aureus, and understanding of the pathobiology of prosthetic devices may lead to innovations in the management and prevention of bacterial joint infections.

The normal diarthrodial joint is resistant to bacterial infection because of local and systemic host defenses. Bacteria can reach the synovial-lined joint, however, via the hematogenous route and result in septic arthritis. The large joints are affected more commonly than the small joints, and monoarticular infection is the rule, with polyarticular infection (more than one joint involved) in less than 20% of cases. A prospective series from a community-based population in the Netherlands reflected a representative distribution of joint involvement: knee 55%, ankle 10%, wrist 9%, shoulder 7%, hip 5%, elbow 5%, sternoclavicular joint 5%, sacroiliac joint 2%, and foot joint 2%.

The incidence of septic arthritis ranges from 2 to 5/100,000/year in the general population, 5.5 to 12/100,000/year in children, 28 to 38/100,000/year in patients with RA, and 40 to 68/100,000/year in patients with joint prostheses. The incidence appears to be increasing, probably related to orthopedic procedures, an aging population, and the increased use of immunosuppressive therapy. The organisms causing bacterial arthritis depend on the epidemiologic circumstances (Table 109-1). Monoarthritis of a prosthetic joint in an elderly man is likely due to Staphylococcus, whereas a migratory arthritis in a sexually active woman with skin lesions is likely due to disseminated gonococcal infection. Septic arthritis caused by methicillin-resistant S. aureus (MRSA) is common in the elderly, in persons who use intravenous drugs, and in individuals with prosthetic joints.

**KEY POINTS**

- Acute bacterial arthritis is a medical emergency that warrants rapid, accurate diagnosis and immediate treatment along with appropriate consultation with additional specialists as needed.
- Most instances of native joint infection are the result of bacteremic seeding.
- *Staphylococcus aureus* is the most frequent microorganism in adult nongonococcal septic arthritis.
- The initial selection of an antibiotic regimen should be broad enough to take into account host factors, clinical characteristics, likely causative microorganisms, and regional antibiotic sensitivity data pending confirmation of bacteria by culture and sensitivities.
- An infected joint must be adequately drained, and an antibiotic course that is sufficiently long to cure the infection must be instituted. Surgical drainage should be considered, but only if needle aspirations are unsuccessful or impractical.
- Poor prognostic factors in bacterial joint infection include old age, underlying rheumatoid arthritis, and infection in a prosthetic joint.
- Débridement, antibiotics, and implant retention (DAIR) is a reasonable option for selected patients with early prosthetic joint infections.
- Late prosthetic joint infections require antibiotic treatment directed at the isolated microorganism and the complete removal of the infected prosthesis before reimplantation of a new prosthesis in a one-stage or two-stage operation.
- Reducing the risk of a prosthetic joint infection involves a thorough preoperative evaluation, perioperative use of antibiotics, and the careful use of antibiotic prophylaxis when a patient with a prosthesis is exposed to transient bacteremia; clinical evidence does not support the use of antibiotic prophylaxis for most dental procedures.
ETIOLOGY

Most cases of septic arthritis result from hematogenous seeding of the synovial membrane. The abundant vascular supply of the synovium and the lack of a limiting basement membrane allow organisms to target joints during bacteremia. Less common causes of septic arthritis include direct inoculation after joint aspiration or corticosteroid injection of a joint; animal or human bites; nail puncture wounds or plant thorn injury; joint surgery, especially hip and knee arthroplasties; and spread by contiguous osteomyelitis, cellulitis, or septic bursitis.

Table 109-1 lists the common organisms that cause joint infections according to the age of the patient and whether the joint is native or prosthetic. Overall, S. aureus is the most common etiologic agent among children of all age groups, followed by group A streptococci and Streptococcus pneumoniae. Neonates and infants younger than 2 months old are more susceptible to group B streptococci and gram-negative enteric bacilli than older children. Rarely, Pseudomonas, N. gonorrhoeae, and Candida albicans may be responsible in very young children. Since the introduction of the Haemophilus influenzae type B vaccine, the incidence of septic arthritis caused by H. influenzae has declined dramatically. In sexually active adolescents, N. gonorrhoeae must be considered. Pseudomonas aeruginosa and Candida are potential pathogens in adolescent intravenous drug abusers. Patients with sickle cell anemia are prone to develop Salmonella arthritis, and immunocompromised children are at higher risk for infection with gram-negative bacilli. Other unusual joint pathogens in children include Neisseria meningitidis, anaerobes, Brucella, and Kingella kingae.

The organisms causing nongonococcal septic arthritis in adults are 75% to 80% gram-positive cocci and 15% to 20% gram-negative bacilli. S. aureus is the most common organism in native and prosthetic joint infections. Staphylococcus epidermidis is common in prosthetic joint infections but is a rare cause of native joint infections. The streptococci including Streptococcus pneumoniae are the next most common group of gram-positive aerobes. Streptococcus pyogenes is followed by groups B, G, C, and F in frequency. Patients with non–group A streptococcal disease often have comorbidities such as immunosuppression, diabetes mellitus, malignancy, and severe genitourinary or gastrointestinal infections. Group B streptococcal arthritis in adults is uncommon, but it can be a serious infection in adult diabetics and patients with late prosthetic hip infections. Aggressive polyarthritis caused by group B streptococci may result in serious functional damage and permanent morbidity. Patients predisposed to gram-negative bacillary infections include patients with a history of intravenous drug abuse, very young and very old patients, and immunocompromised patients. The most common gram-negative organisms are Escherichia coli and P. aeruginosa.

Anaerobes account for 5% to 7% of septic arthritis. Common anaerobes include Bacteroides, Propionibacterium acnes, and various anaerobic gram-positive cocci. Predisposing factors include wound infections, joint arthroplasty, and immunocompromised hosts. Foul-smelling synovial fluid or air in the joint space should raise the suspicion of anaerobic infection, and appropriate cultures should be obtained and held for at least 2 weeks. Anaerobes and coagulase-negative staphylococci are more common in prosthetic joint infections.

Polyarticular septic arthritis is much less common than monoarticular infection. Many of the patients have one or more comorbidities, and some have been intravenous drug abusers. Occurrence of polyarticular septic arthritis is high in patients with RA and averages 25% (range, 18% to 35%). Although S. aureus is the most common pathogen, group G streptococci, H. influenzae, S. pneumoniae, or mixed aerobic and anaerobic bacteria have been responsible.

Table 109-1  Organisms Causing Joint Infection in Various Hosts

<table>
<thead>
<tr>
<th>Adults</th>
<th>Children ≤5 yr old</th>
<th>Children &gt;5 yr old</th>
<th>Neonates</th>
<th>Prosthetic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Common</strong></td>
<td><strong>Common</strong></td>
<td><strong>Common</strong></td>
<td><strong>Common</strong></td>
<td><strong>Common</strong></td>
</tr>
<tr>
<td>Staphylococcus aureus</td>
<td>S. aureus</td>
<td>S. aureus</td>
<td>S. aureus</td>
<td>Coagulase-negative staphylococci</td>
</tr>
<tr>
<td>Streptococcus pneumoniae</td>
<td>Haemophilus influenzae*</td>
<td>Group A streptococci</td>
<td>Group B streptococci</td>
<td></td>
</tr>
<tr>
<td>β-Hemolytic streptococci (mainly Lancefield groups A, G, and B)</td>
<td>S. pneumoniae</td>
<td>Enterobacteriaceae</td>
<td>Pseudomonas aeruginosa</td>
<td></td>
</tr>
<tr>
<td>Neisseria meningitidis</td>
<td>N. meningitidis</td>
<td>N. gonorrhoeae</td>
<td>N. gonorrhoeae</td>
<td></td>
</tr>
<tr>
<td>Pasteurella</td>
<td>Kingella kingae</td>
<td>H. influenzae</td>
<td>H. influenzae</td>
<td></td>
</tr>
<tr>
<td>Anaerobes</td>
<td>M. tuberculosis</td>
<td>M. tuberculosis</td>
<td>M. tuberculosis</td>
<td></td>
</tr>
<tr>
<td>Mycoplasma/Ureaplasma</td>
<td>B. burgdorferi</td>
<td>B. burgdorferi</td>
<td>B. burgdorferi</td>
<td></td>
</tr>
<tr>
<td>Fungi (Sporothrix, dimorphic fungi, Cryptococcus)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borrelia burgdorferi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Rare in children immunized with Hib vaccine.
PATHOGENESIS

Acute bacterial arthritis is usually designated gonococcal or nongonococcal. In the case of gonococcal arthritis, *N. gonorrhoeae* possesses a variety of virulence factors on the cell surface. *N. gonorrhoeae* is able to attach to cell surfaces via filamentous outer-membrane appendages, or pili. Another outer membrane protein, protein I, has forms IA and IB. Protein IA binds the host factor H and inactivates complement component, C3b, circumventing the host’s complement system.²² Protein IA also prevents phagolysosomal fusion in neutrophils, enabling survival of the organism within the phagocytes. Lipo-oligosaccharide is a gonococcal molecule similar to lipopolysaccharide of other gram-negative bacteria and possesses endotoxin activity, which contributes to the joint damage seen in gonococcal arthritis.²³

*S. aureus* is the most common organism that causes nongonococcal arthritis. The virulence of *S. aureus* is associated with its ability to attach to host tissue within the joint, evade host defenses, and cause damage to the joint.

Table 109-2 lists some of these virulence factors and their mechanisms of action. The attachment of *S. aureus* to the joint tissues is facilitated by microbial surface components recognizing adhesive matrix molecules (MSCRAMMs). MSCRAMMs are embedded in the cell wall peptidoglycan of *S. aureus* (Figure 109-1).²⁴ They bind to host matrix proteins including collagen, fibrinogen, elastin, vitronectin, laminin, and fibronectin. Gene knockout experiments in animal models showed that the gene coding for the protein that binds collagen is an important virulence factor for *S. aureus* joint infections.²⁵ Most *S. aureus* isolates also express the fibronectin-binding proteins, FnbpA and FnbpB. Disruption of the respective genes, fnbpA and fnbpB, by knockout gene experiments completely obliterates adherence of *S. aureus* to fibronectin-coated surfaces (e.g., prosthetic joints).²⁶

The genes of several *S. aureus* cell surface proteins (e.g., protein A, fibronectin-binding proteins, coagulase) and

for polyarticular infections. Involvement of more than one joint also can occur in certain patient populations such as neonates and patients with sickle cell anemia, or with certain organisms, such as *N. gonorrhoeae*, *N. meningitidis*, and *Salmonella*.²⁴

Polymicrobial (two or more bacterial species), polyarticular (two or more joints) septic arthritis is a rare clinical entity.²⁵ Large joints are usually affected. Among five reported cases, the knee was affected in four cases (bilaterally in two); the elbow and wrist were affected in three cases, and the shoulder was affected in two cases. The mean number of joints infected was three. Bacteremia was present in all but one case (80%) and always involved the same organisms that were in the synovial fluids. Most bacterial species isolated were the usual organisms seen in septic arthritis. Combinations of gram-positive aerobic and anaerobic organisms were common. A characteristic of most cases (80%) was the extension of locally destructive processes as a result of the contiguous spread of infection from the affected joints such as osteomyelitis, fasciitis with compartment syndrome, and abscess or sinus tract formation. Systemic complications including septic shock, multiorgan failure, and toxic shock syndrome were noted in 60% of cases. The mortality rate of polyarticular, polyarticular septic arthritis in this small series was 60%.²⁶

Arthrocentesis is a common procedure frequently used in conjunction with corticosteroid administration in patients with various forms of joint diseases. Septic arthritis after joint aspiration and injection is extremely rare, occurring in 0.0002% of patients.²⁷ Arthroscopic surgery is also a common procedure that is complicated by a low incidence of septic arthritis (<0.5% of procedures).²⁸ Coagulase-positive and coagulase-negative staphylococci account for more than 87% of these infections. In rare cases of septic arthritis of the knee related to anterior cruciate ligament repair, the tissue allografts were identified as the source of the infection.²⁹ Cultures yielded gram-negative organisms such as *Pseudomonas aeruginosa*, *Citrobacter*, *Klebsiella oxytoca*, and mixed infection with *S. aureus*, *Enterococcus faecalis*, and *P. aeruginosa*.

### Table 109-2: Virulence Factors of *Staphylococcus aureus* and Their Mechanisms of Action

<table>
<thead>
<tr>
<th>Virulence Factor</th>
<th>Mechanism of Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collagen-binding protein</td>
<td>Binds collagen</td>
</tr>
<tr>
<td>Clumping factor A and B</td>
<td></td>
</tr>
<tr>
<td>Fibronectin-binding protein A and B</td>
<td>Binds fibrinogen</td>
</tr>
<tr>
<td>Capsular polysaccharide</td>
<td></td>
</tr>
<tr>
<td>Protein A</td>
<td>Antiphagocytic</td>
</tr>
<tr>
<td>Toxic shock syndrome toxin-1</td>
<td>Superantigen</td>
</tr>
<tr>
<td>Enterotoxins</td>
<td>Superantigens</td>
</tr>
</tbody>
</table>

**Figure 109-1** Schematic diagram of *Staphylococcus aureus*. Many of the cell surface proteins are regulated by the agr locus (see text). At low cell concentrations, agr facilitates the production of the cell-surface proteins, which facilitate attachment to tissue. At higher cell concentrations, as occurs with establishment of infection, agr downregulates production of the cell surface proteins and activates genes coding for exotoxins. TSST-1, toxic shock syndrome toxin-1.
exotoxins (e.g., toxic shock syndrome toxin-1 [TSST-1], enterotoxin B, proteases, and hemolysins) are regulated by the accessory gene regulator \(agr\). At low cell numbers such as at the time of infection, production of cell surface proteins for attachment to host tissues is facilitated by the \(agr\) gene. When the cells have attached to tissue or an orthopedic device and have passed from exponential to stationary phase of growth, \(agr\) represses the expression of genes coding for cell surface proteins and activates genes coding for exotoxins and tissue-destructing exoenzymes. Because of this complex effect on the different stages of infection, inhibitors of \(agr\) may reduce tissue destruction but enhance tissue colonization. This effect could have implications for chronic infections such as occur with prosthetic joints.

Adherence receptors may allow intracellular movement of \(S. aureus\) into host cells (e.g., osteoblasts, endothelial cells, neutrophils). When internalized, the organism is protected from the host’s immune system and from antimicrobial agents. After adherence to the joint tissue, the bacteria activate the host immune response. Opsonization and phagocytosis are key defenses to eradicate the organism. \(S. aureus\) possesses two virulence factors, protein A and capsular polysaccharide, which interfere with these defenses. Protein A interferes with binding of complement by binding to the fragment crystallizable (Fc) portion of IgG. Protein A has been termed a superantigen for B cells because 30% of human B cells show Fab-mediated binding of the protein A molecule. Binding of protein A by B cells leads to activation and subsequently to depletion of B cells through apoptosis. This process may have implications regarding the ability of the immune system to control infection with \(S. aureus\). The gene coding for protein A had been experimentally disrupted, and joint infection caused by the altered strain in a mouse model resulted in less joint destruction than infection caused by the wild-type strain.

Capsular polysaccharide interferes with opsonization and phagocytosis. Of the 11 reported capsule serotypes of \(S. aureus\), types 5 and 8 account for 85% of clinical infections. The capsule of these two serotypes is thinner, which facilitates the attachment to host fibronectin and fibrin. When attached to these host proteins, capsule production is upregulated to form a thicker capsule, which makes the bacteria more resistant to opsonization and phagocytosis. The thicker capsule can also conceal the highly immunogenic adherence proteins (MSCRAMMs). A mutant of the type 5 capsule in a murine model had a lower rate of infection and resulted in less severe arthritis compared with mice infected with the wild-type strain. A vaccine consisting of types 5 and 8 polysaccharide reduced \(S. aureus\) bacteremia by more than half in hemodialysis patients. The duration of protection was approximately 40 weeks after a single vaccination.

\(S. aureus\) exotoxins (e.g., TSST-1 and enterotoxins) act as superantigens that bind to host major histocompatibility complex (MHC) class II molecules and T cell receptors, resulting in clonal expansion and activation of some T cells. This activation triggers the release of numerous cytokines including interleukin (IL)-2, interferon-\(\gamma\), and tumor necrosis factor (TNF). Induction of these cytokines results in systemic toxicity and joint damage. The stimulated T cells initially proliferate but later disappear, likely due to apoptosis, and result in immunosuppression. Internalized organisms that had been protected from this inflammatory response may cause fulminant or persistent infection. Mice injected with strains of \(S. aureus\) lacking TSST-1 and enterotoxins rarely develop arthritis; when arthritis is induced, it is much milder compared with arthritis in animals injected with the wild-type strain. Vaccination of mice with a mutated, recombinant form of enterotoxin A devoid of superantigen function was associated with a significant reduction in mortality.

In response to bacterial infection of the joint space, the host releases a variety of cytokines and inflammatory mediators. Initially, IL-1\(\beta\) and IL-6 are released into the joint space, leading to an influx of inflammatory cells. These neutrophils and macrophages engulf invading bacteria and release additional cytokines including TNF, IL-1, IL-6, and IL-8. Blocking TNF with a monoclonal antibody and IL-1 with an IL-1 receptor antagonist inhibited leukocyte infiltration into the joint by 80% in a rabbit model of \(S. aureus\)–induced arthritis when the cytokine inhibitors were given simultaneously with \(S. aureus\). When the same inhibitors were given 24 hours after infection, however, there was no effect on leukocyte infiltration, suggesting the crucial roles of TNF and IL-1 in the early stages of \(S. aureus\)–induced arthritis. Release of interferon-\(\gamma\) is associated with the influx of T cells, which occurs a few days after infection. In a mouse model of \(S. aureus\) septic arthritis, interferon-\(\gamma\) has been associated with a worsening of the severity of arthritis while protecting the animals from sepsis. The host’s early cytokine response may aid the clearance of organisms and limit infection in the host. A late cytokine response may amplify the destructiveness of an established infection.

**CLINICAL FEATURES**

Acute bacterial arthritis is most commonly monoarticular. Polyarticular infection occurs in 5% to 8% of pediatric cases and in 10% to 19% of adult nongonococcal cases. The differential diagnosis of acute monoarthritis overlaps with many causes of polyarthritis because virtually any form of arthritis can initially manifest as a single swollen joint. The three main etiologies to consider when a patient presents with acute monoarticular arthritis are trauma, infection, and crystal-induced synovitis such as gout or pseudogout. Polyarticular septic arthritis is usually seen in patients with systemic inflammatory disorders such as the spondyloarthropathies, RA, systemic lupus erythematosus, and other connective tissue diseases or patients with overwhelming sepsis.

Disseminated gonococcal infection occurs in 1% to 3% of patients infected with \(N. gonorrhoeae\). Gonococcal arthritis is the most common cause of acute monoarthritis in sexually active young adults. In the preantibiotic era, gonococcal arthritis was a well-recognized illness in neonates. Disseminated gonococcal infection is three times more common in women than men. Women are more commonly affected because they are more likely to have asymptomatic and untreated primary infections. Bacterial dissemination has been associated with intrauterine devices and has occurred during menstruation, pregnancy, and pelvic operation.
Patients with gonococcal joint disease typically present with one of two forms. The first form is characterized by fever, shaking chills, vesiculopustular skin lesions, tenosynovitis, and polyarthralgias. Blood cultures are frequently positive, whereas synovial fluid cultures are rarely positive. N. gonorrhoeae can be cultured from the genital, rectal, and pharyngeal sites. Tenosynovitis of multiple tendons of the wrist, fingers, ankle, and toes is a unique feature of this form of disseminated gonococcal infection and distinguishes it from other forms of infectious arthritis. In the second form of gonococcal infection, patients have purulent arthritis, most commonly of the knee, wrist, or ankle, and more than one joint can be infected simultaneously. N. gonorrhoeae can frequently be cultured from the synovial fluid.45

The classic presentation of nongonococcal septic arthritis is the acute onset of pain, swelling, and decreased range of motion in a single joint. Large joints are affected most commonly. In adults, the knee is involved in more than 50% of cases; hip, ankle, and shoulder infections are less common.31 In infants and small children, the hip is more often involved.46 Patients with septic arthritis often have underlying illnesses and predispositions to infections. Many are immunocompromised; are intravenous drug abusers; have prosthetic joints; and have diseases such as neoplasia, renal failure, and RA. Table 109-3 lists the risk factors that predispose to septic arthritis.3,5,47

Most patients with bacterial arthritis are febrile, although chills are unusual. Fever may be absent in elderly patients. In children, septic arthritis usually is accompanied by fever, malaise, poor appetite, irritability, and progressive reluctance to use the affected limb. Physical examination typically reveals warmth and tenderness of the affected joint, joint effusion, and limited active and passive range of motion. Septic arthritis among patients with RA has been a special challenge to clinicians because of the high incidence of infection and the poor outcome. Septic arthritis in patients with RA is associated with poor joint outcome and high mortality.42,46 In many cases, it is difficult to differentiate septic arthritis in a joint already affected by RA from rheumatoid flare. Whenever bacterial arthritis is suspected, the most important diagnostic procedure is arthrocentesis and examination of the synovial fluid. For joints that are deep and more difficult to aspirate, ultrasound-guided or fluoroscopy-guided needle aspiration should be done.

### DIAGNOSIS AND DIAGNOSTIC TESTS

Arthrocentesis and synovial fluid analysis should be performed for all patients who present with an inflamed joint. Normal joints contain a small amount of synovial fluid that is clear, is highly viscous, and has few white blood cells (WBCs). The protein concentration is approximately one-third that of plasma, and the glucose concentration is similar to that of plasma. Infected synovial fluid is usually purulent with an elevated leukocyte count typically greater than 50,000 WBC/mm³ and often exceeding 100,000 WBC/mm³ with polymorphonuclear cell predominance. Synovial fluid levels of glucose, lactate dehydrogenase, and total protein have limited value in the diagnosis of septic arthritis. Although a low synovial fluid glucose (<40 mg/dL or less than half the serum glucose concentration) and an elevated lactate dehydrogenase suggest bacterial infection, they are not sufficiently sensitive or specific for the diagnosis.49

Figure 109-2 is an algorithm for synovial fluid analysis; Table 109-4 lists the differential diagnoses of septic arthritis and the known causes of pseudoseptic arthritis.50

A definite diagnosis of bacterial arthritis can be made only by visualizing bacteria on a gram-stained smear or by culturing bacteria from the synovial fluid. In patients not previously treated with antibiotics, synovial fluid cultures are positive in 70% to 90% of cases of nongonococcal bacterial arthritis.51 Blood cultures are positive in 40% to 50% of cases of septic arthritis and are the only method of identifying the pathogen in about 10% of cases.52,53 Occasionally, an extra-articular site of infection offers a clue to the etiologic organism infecting the joint. Examples include septic arthritis in association with pneumococcal pneumonia, E. coli urinary tract infection, and cellulitis caused by staphylococci or streptococci. Gram-positive

### Table 109-3 Risk Factors for Development of Septic Arthritis

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &gt;80 yr</td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td></td>
</tr>
<tr>
<td>Presence of a prosthetic joint in the knee or the hip</td>
<td></td>
</tr>
<tr>
<td>Recent joint surgery</td>
<td></td>
</tr>
<tr>
<td>Skin infection</td>
<td></td>
</tr>
<tr>
<td>Previous septic arthritis</td>
<td></td>
</tr>
<tr>
<td>Recent intra-articular injection</td>
<td></td>
</tr>
<tr>
<td>HIV or AIDS</td>
<td></td>
</tr>
<tr>
<td>Intravenous drug abuse</td>
<td></td>
</tr>
<tr>
<td>End-stage renal disease on hemodialysis</td>
<td></td>
</tr>
<tr>
<td>Advanced hepatic disease</td>
<td></td>
</tr>
<tr>
<td>Hemophilia with or without AIDS</td>
<td></td>
</tr>
<tr>
<td>Sickle cell disease</td>
<td></td>
</tr>
<tr>
<td>Underlying malignancy</td>
<td></td>
</tr>
<tr>
<td>Hypogammaglobulinemia (susceptible to <em>Mycoplasma</em> infections)</td>
<td></td>
</tr>
<tr>
<td>Late complement-component deficiency (susceptible to <em>Neisseria</em> infections)</td>
<td></td>
</tr>
<tr>
<td>Low socioeconomic status with high rate of comorbidities</td>
<td></td>
</tr>
</tbody>
</table>

### Table 109-4 Differential Diagnosis of Septic Arthritis and Reported Causes of Pseudoseptic Arthritis

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Partially treated septic arthritis</td>
<td></td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td></td>
</tr>
<tr>
<td>Juvenile rheumatoid arthritis</td>
<td></td>
</tr>
<tr>
<td>Gout</td>
<td></td>
</tr>
<tr>
<td>Pseudogout</td>
<td></td>
</tr>
<tr>
<td>Apatite-related arthropathy</td>
<td></td>
</tr>
<tr>
<td>Reactive arthritis</td>
<td></td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
<td></td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td></td>
</tr>
<tr>
<td>Sickle cell disease</td>
<td></td>
</tr>
<tr>
<td>Dialysis-related amyloidosis</td>
<td></td>
</tr>
<tr>
<td>Transient synovitis of the hip</td>
<td></td>
</tr>
<tr>
<td>Plant thorn synovitis</td>
<td></td>
</tr>
<tr>
<td>Metastatic carcinoma</td>
<td></td>
</tr>
<tr>
<td>Pigmented villonodular synovitis</td>
<td></td>
</tr>
<tr>
<td>Hemarthrosis</td>
<td></td>
</tr>
<tr>
<td>Neuropathic arthropathy</td>
<td></td>
</tr>
<tr>
<td>Synovitis after injection of hylan</td>
<td></td>
</tr>
</tbody>
</table>

*Extremely inflammatory synovitis with negative culture is referred to as pseudoseptic arthritis. Typically, synovial fluid analysis shows ≥250,000 white blood cells (WBC/mm³). Often the WBC count is >100,000 WBC/mm³.

Data from references 6, 52, 53.
Algorithm for synovial fluid analysis in septic arthritis.

OA, osteoarthritis; RA, rheumatoid arthritis; WBC, white blood cells.

Figure 109-2 Algorithm for synovial fluid analysis in septic arthritis. OA, osteoarthritis; RA, rheumatoid arthritis; WBC, white blood cells.

cocci are identified in 50% to 75% of synovial fluid gram-stained smears, but gram-negative bacilli are identified less than 50% of the time in culture-proven cases.51

Inflammatory markers such as erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), and WBC are usually raised, but the sensitivity is low and their absence does not exclude the diagnosis of septic arthritis.47,52 Serum and joint fluid procalcitonin levels as a marker of septic arthritis have been studied, but the results are inconclusive.53 A study that evaluated neutrophil-derived circulating free deoxyribonucleic acid (j-cf-DNA) in synovial fluid of 42 patients found that at a cutoff of 300 ng/mL, j-cf-DNA had a sensitivity of 89%, a specificity of 100%, a positive predictive value of 100%, and a negative predictive value of 97%.54 If validated and standardized, this could be a valuable additional test to diagnose septic arthritis.

Culture for N. gonorrhoeae is almost always negative in skin lesions and is positive in less than 50% of synovial fluid samples and in less than one-third of blood cultures; this may be the result of the fastidious growth requirements of N. gonorrhoeae. The organism can often be easily recovered from other sites such as urethral, cervical, rectal, or pharyngeal specimens (i.e., the genital tract). In culture-negative septic arthritis where N. gonorrhoeae is suspected, polymerase chain reaction techniques can be used to detect gonococcal DNA in the synovial fluid. Unfortunately, the technique is not standardized and is not widely available.55

When culturing the synovial fluid, it should be brought directly to the laboratory and placed on conventional broth and solid media or into aerobic and anaerobic blood culture bottles. Inoculating blood culture bottles with 5 to 10 mL of joint fluid or smaller volumes into isolator tubes may increase the yield of positive cultures beyond that of standard techniques.56,57 Synovial fluid culture using the BACTEC Peds Plus/F bottle and the BACTEC 9240 instrument (Becton Dickinson Diagnostic Systems, Sparks, Md) detected significantly more pathogens and fewer contaminants than culture by the agar-plate method.60

Plain radiographs in septic arthritis are usually normal early in the course of the infection, but baseline films should be obtained to look for evidence of other disease and contiguous osteomyelitis. Radiographs often show nonspecific changes of inflammatory arthritis including periarticular osteopenia, joint effusion, soft tissue swelling, and joint space loss. In more advanced infection, periosteal reaction, marginal or central erosions, and destruction of subchondral bone may be seen. Bony ankylosis is a late sequela of septic arthritis. Dislocation or subluxation of the femoral head is unique to hip infection of neonates.55

Ultrasound of the hip is the modality of choice to detect fluid collections in this deep joint and can serve as a guide in its aspiration. Ultrasound can be similarly used in other joints such as the popliteal cyst of the knee, shoulder, acromioclavicular, or sternoclavicular joints. Triple-phase bone scan using technetium 99m is often done in children to identify an associated metaphyseal osteomyelitis or avascular necrosis of the femoral head. Whole-body bone scan is preferred in young children because, despite focal symptoms, septic arthritis and osteomyelitis may be multifocal in this age group.56 In septic arthritis of all age groups, the periarticular distribution of increased uptake is seen on
the early “blood-pool” phase and the delayed images of the joint. Bone scans provide only nonspecific information, however, and cannot differentiate septic from noninfectious causes of joint inflammation. Bone scans are more sensitive than standard radiography in the diagnosis of arthritis because radionuclide uptake precedes morphologic bone changes that are seen on radiograph. A suggestive bone scan must be interpreted in the proper clinical context and supported by microbiologic data for a definitive diagnosis of joint or bone infection.

In joints that are difficult to evaluate otherwise or that have complex anatomic structures, computed tomography (CT) and magnetic resonance imaging (MRI) can provide useful images to delineate the extent of the infection. MRI is highly sensitive in early detection of joint fluid and is superior to CT in the delineation of soft tissue structures and soft tissue abscess. These images can show early bone erosion; reveal soft tissue extension; and facilitate arthrocentesis of joints such as shoulders, hips, acromioclavicular, sternoclavicular, sacroiliac, and facet joints of the spine. MRI findings such as reactive bone marrow can suggest presence of secondary osteomyelitis, which can complicate septic arthritis. When multiple joint involvement is suspected, triple-phase bone scintigraphy is the preferred modality of investigation.

**TREATMENT**

Treatment of septic arthritis must begin immediately after the clinical evaluation is complete and all appropriate cultures are taken. A serious clinical suspicion of a joint infection warrants the initiation of antibiotic therapy before culture confirmation is available. Delays in treatment allow the infection to become more established in the joint and permanently damage the articular cartilage. Untreated, there is the opportunity for the joint infection to spread to other body sites via the hematogenous route and become more widespread and more difficult to cure.

The principles of treatment of an infected joint, whether natural or prosthetic, follow those of treatment of an infected body cavity in which antibiotics must be used in conjunction with adequate drainage of the infected closed space. The clinical circumstances and the preliminary laboratory data aid the selection of antibiotic agents. Host factors, any extra-articular sites of infection, and the gram-stained smear of the synovial fluid are the best early guides for the antibiotic agents with which to start. Table 109-5 lists current antibiotic agents for adults, and Table 109-6 lists agents for children.

Narrow antibiotic coverage is indicated if gram-positive cocci are found in the synovial fluid, and the clinician suspects a primary source of staphylococcal infection from the skin. Appropriate monotherapy in this case may be a penicillinase-resistant penicillin or vancomycin if methicillin resistance is likely. If gram-negative bacilli are noted in the synovial fluid and the patient has a kidney infection, specific agents (e.g., ampicillin or a cephalosporin) against *E. coli* and other common urinary tract pathogens may be used. In healthy, young, sexually active individuals with community-acquired septic arthritis and a negative synovial fluid gram-stained smear, ceftriaxone is a reasonable option to cover *N. gonorrhoeae*. If synovial fluid gram stain shows gram-positive cocci, vancomycin should be the empiric therapeutic option considering the fact that a significant proportion of community-acquired *S. aureus* infections are now methicillin resistant. A reasonable initial empiric therapy to cover gonococci, *S. aureus*, and streptococci is ceftriaxone plus vancomycin pending final culture results. In elderly debilitated patients or adults with low risk for sexually transmitted disease, as well as a negative gram-stained smear of synovial fluid, broad antibiotic coverage against a wide variety of organisms including *S. aureus*, streptococci, and gram-negative bacilli should be given initially. A typical regimen includes an antistaphylococcal agent (e.g., vancomycin) plus a third-generation cephalosporin (e.g., ceftriaxone).

When the identity and the sensitivities of the organism are known, antibiotic therapy should continue with the most efficacious agent that has the best safety profile and narrowest spectrum. The parenteral route of antibiotic administration is the preferred initial treatment. Continued antibiotic therapy may be switched to oral agents if adequate blood levels can be achieved and maintained by this route. There is no evidence that the direct intra-articular instillation of drugs is necessary or preferable in septic arthritis because there is no barrier against the free diffusion of antibiotic agents from the blood to the synovial fluid. In cases in which uncertainty exists, serum and synovial fluid levels of antibiotic drugs can be measured to ensure that therapeutic levels are reached.

Although some clinicians feel that patients with native-joint septic arthritis need urgent surgical drainage, the medical literature suggests otherwise. Most individuals with septic arthritis respond adequately to appropriate antimicrobial agents after initial joint aspiration for fluid analysis. In experimental infectious arthritis cases, early antibiotic therapy was shown to reduce the loss of collagen and erosion of articular surface, which minimizes the need for open surgical drainage. It is generally accepted that prompt and adequate drainage of the septic joint is essential to decrease the risks of substantial loss of articular function; however, the best approach to drain the joint remains controversial.

From retrospective studies, daily aspiration of an infected joint showed better functional outcome than open surgical drainage, although the former had higher overall mortality. An explanation for higher mortality could be the higher comorbid conditions of patients who had daily aspirations than the ones who were more fit and underwent open surgical drainage. If the synovial fluid cell count and polymorphonuclear percentage decrease with successive aspiration, the antimicrobial therapy is probably effective. If needle aspiration is technically difficult (as in the hip or the shoulder) or does not provide thorough drainage of the joint, if the joint effusion does not resolve promptly, if sterilization of the joint fluid is delayed, if the infected joint is already damaged by pre-existing rheumatoid disease, or if infected synovial tissue or bone needs débridement, surgical drainage should be considered sooner rather than later. Arthroscopy is emerging as an alternative to arthrotomy with the advantage of reduced surgical morbidity. Wound healing is faster, and rehabilitation time is shortened.

A recent retrospective analysis from the United Kingdom suggested that most patients can be treated...
**Table 109-6  Antibiotic Agents Used in Children**

<table>
<thead>
<tr>
<th>Age</th>
<th>Likely Pathogen</th>
<th>Antibiotic</th>
<th>Dosage (mg/kg/day)</th>
<th>Doses/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neonate</td>
<td><em>Staphylococcus aureus</em>; group B streptococci; gram-negative bacilli</td>
<td>Nafcillin/oxacillin or Cefazolin or Vancomycin or Clindamycin or Linezolid</td>
<td>100</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Nafcillin or Gentamicin</td>
<td>150</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Cefotaxime or Ceftriaxone</td>
<td>50-7.5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Piperacillin or Gentamicin or Ceftriaxone or Cefotaxime or Imipenem</td>
<td>150</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Ciprofloxacin</td>
<td>100-150</td>
<td>3-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Ceftriaxone</td>
<td>50</td>
<td>1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Cefuroxime or Cefazolin</td>
<td>150-200</td>
<td>3-4</td>
</tr>
<tr>
<td>Child &lt;5 yr old</td>
<td><em>S. aureus</em>; <em>Haemophilus influenzae</em>; group A streptococci; <em>Streptococcus pneumoniae</em></td>
<td>Nafcillin or Ceftriaxone or Cefuroxime or Cefotaxime or Cefazolin</td>
<td>150</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Gentamicin or Nafcillin</td>
<td>150</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Cefotaxime or Ceftriaxone</td>
<td>50</td>
<td>1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Ciprofloxacin</td>
<td>50</td>
<td>3-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Ceftriaxone</td>
<td>50</td>
<td>1-2</td>
</tr>
<tr>
<td>Child &gt;5 yr old</td>
<td><em>S. aureus</em>; group A streptococci</td>
<td>Nafcillin or Ceftriaxone</td>
<td>150</td>
<td>3-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Cefuroxime or Cefazolin</td>
<td>150</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Gentamicin or Nafcillin</td>
<td>150</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Cefotaxime or Ceftriaxone</td>
<td>50</td>
<td>1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Ciprofloxacin</td>
<td>50</td>
<td>3-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Ceftriaxone</td>
<td>50</td>
<td>1-2</td>
</tr>
<tr>
<td>Adolescent (sexually active)</td>
<td>Previous organisms; <em>Neisseria gonorrhoeae</em></td>
<td>Nafcillin or Ceftriaxone</td>
<td>150</td>
<td>3-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Cefuroxime or Cefazolin</td>
<td>150</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Gentamicin or Nafcillin</td>
<td>150</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Cefotaxime or Ceftriaxone</td>
<td>50</td>
<td>1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Ciprofloxacin</td>
<td>50</td>
<td>3-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or Ceftriaxone</td>
<td>50</td>
<td>1-2</td>
</tr>
</tbody>
</table>

*Decreased incidence in children fully immunized with Hib vaccine.
If patient is penicillin allergic, alternatives include vancomycin (40 mg/kg/day divided into four doses) or clindamycin (20-40 mg/kg/day divided into four doses).
medically (with repeated aspirations) and do not require surgical drainage (either by arthroscopy or arthrotomy). 

Though statistically not significant, medical therapy resulted in more complete cure and less deterioration of functional status at time of discharge. Another study, which included 20 adults with native hip joint septic arthritis, concluded that symptom duration, especially if it is longer than 3 weeks before presentation, was a statistically significant predictor of the need for excision arthroplasty. These results highlight the need for careful case selection for surgical intervention.

The optimal duration of antibiotic treatment has not been prospectively studied. For native joint infections caused by N. gonorrhoeae, a 1-week course of ceftriaxone should be adequate. For septic arthritis caused by organisms other than N. gonorrhoeae, therapy ranging from 2 weeks to 6 weeks is recommended depending on type, sensitivity of microorganism, and presence of osteomyelitis. If long-term antibiotics are chosen (4 to 6 weeks’ duration), parenteral antibiotics may be switched to oral antibiotics after 2 weeks provided that there is clinical improvement, inflammatory markers are trending down, and oral antibiotics are available to which the microorganism is susceptible. The duration of antibiotic administration can be 2 weeks for uncomplicated infection by susceptible microorganisms or 4 to 6 weeks for more extensive infection in an immunocompromised host. For septic arthritis caused by H. influenzae, streptococci, or gram-negative cocci, 2 weeks of antibiotic therapy is usually adequate. Staphylococcal septic arthritis usually requires 3 to 4 weeks of therapy, and for pneumococcal or gram-negative bacillary infections, therapy should be continued for at least 4 weeks.

During the first few days of management, immobilization of the infected joint by external splinting and adequate analgesic administration ensure patient comfort. Physical therapy, starting with passive then graduating to active motion, should be instituted as soon as the patient can tolerate mobilization of the inflamed joint because early active range-of-motion exercises are beneficial for ultimate functional recovery. Involving the orthopedic surgeon and the physical therapist early on in the course of treatment facilitates the best choice of drainage procedure and results in the best functional outcome.

PROSTHETIC JOINT INFECTIONS

Total joint replacement for advanced arthritis is one of the major advances in medicine in the 20th century and continues to improve in the 21st century. Infection of prosthetic joints is an uncommon but devastating complication of joint replacement surgery. Nearly 800,000 total knee and total hip replacements were done in the United States in 2006 (www.cdc.gov/nchs/fastats/insurg.htm), with an infection rate of 1% to 3%. The infection rate is higher for knee arthroplasty (1% to 2%) compared with hip and shoulder arthroplasty (0.3% to 1.3%) and is much higher in patients undergoing reimplantation because of infection of the initial prosthesis (3% for hips and 6% for knees). The risk of infection is about twofold higher in patients with RA compared with patients with osteoarthritis.

The risk of infection is related to many factors. In a retrospective study of 462 infected orthopedic implants, the most important risks for infection included (1) a surgical site infection at a site other than the prosthesis (odds ratio [OR], 35.9), (2) a score of 2 on the National Nosocomial Infections Surveillance System surgical patient risk index (OR, 3.9), (3) the presence of a malignancy (OR, 3.1), and (4) a history of joint arthroplasty (OR, 2.0). Certain patient populations are at increased risk of infection because of comorbid conditions (e.g., diabetes mellitus and RA). Other surgical risk factors include simultaneous bilateral arthroplasty, long operative time (>2.5 hours), blood transfusion, urinary tract infection, and Staphylococcus aureus bacteremia.

Orthopedic implants adversely affect host defenses. Prosthetic devices impair opsonic activity and diminish the ability of neutrophils to kill bacteria. Polymorphonuclear leukocytes release lysosomal enzymes and superoxide into the area surrounding the prosthesis, resulting in tissue damage and local devascularization. Phagocytes may be focused on removal of the foreign body such that fewer cells are available to fight infection. Finally, polymethyl methacrylate bone cement can inhibit neutrophil and complement functions, and the heat produced by the polymerization of polymethyl methacrylate can damage adjacent cortical bone and result in a devascularized necrotic area, which is ideal for bacterial growth. After implantation, prosthetic joints are immediately coated by host proteins including albumin, fibrinogen, and fibronectin. S. aureus, which possesses numerous host protein binding receptors (MSCRAMMs), is a common pathogen in infection of prosthetic joints. Patients with a prosthetic joint who develop S. aureus bacteremia have an approximate one in three chance of developing an infection of the implant.

Another phenomenon crucial to development of infection is the ability of organisms to form biofilms on the surface of the prosthetic device. A biofilm is defined as “an assemblage of microbial cells that is irreversibly associated with a surface and enclosed in a matrix of primarily polysaccharide material.” Biofilm formation is a natural process. Organisms grow on indwelling medical devices, potable water system pipes, and living tissues. S. epidermidis is particularly adept at attaching to and forming biofilms on foreign bodies such as prosthetic joints. Small numbers of these organisms from the patient’s skin or mucous membranes, or from the hands of the surgeons or clinical staff, contaminate and colonize the orthopedic device at the time of implantation. Staphylococcal surface proteins, SSP-1 and SSP-2, are fimbria-like polymers that facilitate adherence of S. epidermidis to polystyrene. S. epidermidis produces a polysaccharide/adhesin substance crucial to the formation of this extracellular matrix known as slime. Polysaccharide/adhesin mutants have been shown to be less virulent than the wild-type strain in a rabbit model of endocarditis.

Prosthetic joint infections are divided into early onset (<3 months after placement), delayed (3 to 24 months postsurgery), and late onset (>24 months after placement). Early and delayed infections are usually related to surgical contamination at the time of the implantation, whereas late infections usually result from hematogenous seeding of the joint. Owing to its high virulence, S. aureus accounts for most early and late infections (see Table 109-1). So-called small colony variants of S. aureus may be responsible for persistent and recurrent infections of prosthetic implants.
These subspecies of *S. aureus* are difficult to treat because they grow slowly and are relatively resistant to cell-wall active antimicrobial agents and aminoglycosides. Delayed infections are usually caused by less virulent microorganisms such as coagulase-negative staphylococci and *P. acnes*. Because these low-virulence organisms are common skin contaminants, it is important to interpret culture results carefully.

Clinically, the most common symptom in patients with prosthetic joint infection is pain of the affected joint. Differentiating pain from mechanical loosening of the prosthesis from pain related to infection can be difficult. Typically, a patient with mechanical loosening but no infection has pain only with motion, whereas a patient with infection experiences pain at rest and with motion. Warmth at the implant site, effusion, erythema, and fever are frequently associated with early and late, but not delayed, infections. This difference in clinical presentation likely represents the virulence of the most common organisms associated with the three categories of infection. The presence of a sinus tract with purulent discharge suggests involvement of the implant and is an indication for removal of the prosthesis.

Inflammatory blood laboratory markers have variable sensitivity and specificity in the diagnosis of prosthetic joint infection. In a recent meta-analysis, elevated levels of interleukin-6 (IL-6) and CRP had a much higher sensitivity (97% and 88%, respectively) than an elevation of the WBC count or ESR (45% and 75%, respectively). The specificity of an elevated level of IL-6 (91%) was the highest of these markers with elevation of WBC (87%), CRP (74%), and ESR (70%) somewhat lower. Inflammatory markers increase following arthroplasty, and the return to normal levels is quite variable. IL-6 will return to normal within a matter of a few days following surgery, whereas the CRP level may remain elevated for up to 3 weeks and the ESR for several months following arthroplasty. Therefore a low or normal CRP or IL-6 level in a patient with suspected prosthetic joint infection has a good negative-predictive value. Serial plain radiographs may be helpful; the presence of subperiosteal bone growth and transcortical sinus tracts is specific for infection. Bone scans using technetium 99m–labeled methylene diphosphonate are sensitive but lack specificity because the bone scan is typically positive for 6 to 12 months after the original implantation. Bone scans may be a useful screening test for patients with suspected late prosthetic joint infection. CT has limitations because of the imaging artifacts caused by the metal implant. MRI can be performed only in patients with titanium or tantalum implants.

Aspiration of the joint may be helpful in differentiating infection from noninfectious causes of joint pain, particularly in patients without RA. In one study, a synovial fluid leukocyte count of greater than 1700/mm³ had a sensitivity of 94% for determining infection, whereas a differential count of greater than 65% neutrophils had a sensitivity of 97%. The specificities of these two measurements were 88% and 98% in patients without underlying inflammatory diseases such as RA. Gram-stained smear of the synovial fluid has a low sensitivity (<20%) but a high specificity (>97%). Cultures of drainage from a sinus tract are not helpful, unless the culture grows *S. aureus*. Generally, at least three tissue specimens should be taken at the time of surgery including tissue from the joint capsule, synovial lining, bone-cement interface, and samples from purulent material or sequestrum. Swabs of the joint have a low sensitivity and should be avoided. Unless the patient is septic or otherwise systemically ill, antimicrobial therapy should be discontinued a minimum of 2 weeks before the revision surgery and perioperative antibiotics should not be administered until all of the tissue cultures have been obtained. Using this methodical approach, there is a direct correlation between the number of tissue specimens positive for a particular microorganism and the probability of infection. The probability of infection has been estimated to be less than 5% if all tissue specimens are negative and greater than 94% if three or more tissue specimens are positive for growth. Finally, the location of the prosthesis is helpful in the interpretation of the positive culture. The isolation of *P. acnes* from a single tissue culture from a knee prosthesis is more likely to be a contaminant than if the same organism is obtained from a shoulder prosthesis.

For patients who have received recent antibiotics, there may be a role for sonication of the prosthesis at the time of explantation. A recent study determined that culturing of the explanted prosthesis that has been sonicated to remove adherent bacteria had a higher sensitivity (75%) as compared with culture of periprosthetic tissue (45%) in patients who had received antibiotics in the past 14 days.

Medical therapy of patients with prosthetic joint infections is challenging. Organisms existing in biofilms are much more resistant to antimicrobial agents for several reasons. First, the drugs have difficulty penetrating the biofilm layer. The biofilm-associated organisms also grow much slower than organisms in suspension. As a result, antimicrobial agents such as vancomycin, penicillins, and cephalosporins, which act on rapidly dividing organisms, are not effective in treating device-related infections. Rifampin and fluoroquinolones may be more effective because they are active against organisms in the stationary phase of growth. Although most experts recommend rifampin in combination with another agent for the treatment of rifampin-susceptible strains of *S. aureus*, there are little clinical data to support the benefit of combination therapy versus monotherapy in the treatment of prosthetic joint infections. The role of newer antibiotics such as linezolid, daptomycin, and tigecycline is unclear at this time. In a rabbit experimental model of *S. aureus* osteomyelitis, the combination of tigecycline and rifampin eradicated infection in 100% of 14 rabbits.

Treatment of late prosthetic joint infections is complex. In most patients, effective therapy requires a combination of antibiotics with the removal of the orthopedic device. Failure to remove the infected prosthesis is frequently associated with an unacceptably high rate of relapse, probably related to biofilm formation on the orthopedic implant. Removal of the joint prosthesis, débridement of infected bone, and placement of a new prosthesis during the same operation has been associated with a high rate of recurrence of infection, but studies indicate that single-stage revision or débridement with retention of the prosthesis may be effective in certain situations. Patients whose symptoms of pain and swelling of the joint have been less than 8 days or less than 3 weeks and who have a stable
JOINT INFECTIONS

There is consensus that preoperative evaluation of a patient for occult infection such as periodontal disease or bacteriuria is warranted, and corrective steps to eradicate any infection are essential before joint replacement. There is also consensus that perioperative antibiotic prophylaxis significantly reduces the rate of early postoperative infection, and this practice is routine. The role of antibiotic prophylaxis to prevent late prosthetic joint infection before diagnostic or therapeutic procedures that lead to transient bacteremia, especially dental treatment, is controversial. A position paper of the American Academy of Oral Medicine in 2010 stated that “the risk of patients’ experiencing drug reactions or drug-resistant bacterial infections and the cost of antibiotic medications alone do not justify the practice of using antibiotic prophylaxis in patients with prosthetic joints.” Many orthopedic and oral and maxillofacial surgeons have argued that there is “no scientific evidence to support the view that patients with arthropathies, even in the high-risk groups, require antibiotic prophylaxis during dental treatment.”

The incidence of late infection of a prosthetic joint as a result of procedure-related bacteremia is extremely low—10 to 100 cases per 100,000 patients with total joint replacement per year. The cost of providing antibiotic prophylaxis to all patients with prosthetic joints before all procedures that are associated with transient bacteremia is substantial. The efficacy of such antibiotic prophylaxis is unknown. Cost-effective analyses have shown mixed results. These discrepancies are due to the lack of reliable data and the different assumptions used in the calculations. In the patient with the greatest risk of infection, an invasive procedure that leads to bacteremia sometimes can result in an infected total joint replacement. Counseling these patients on the risks and benefits of antibiotic prophylaxis would lead to an informed decision on which the patient and the physician can agree.

As we gain experience with the use of biologic agents in the management of RA and other inflammatory arthritides, there remains the conundrum of whether the increased risk of infection from TNF inhibitor and methotrexate warrants holding them before an elective orthopedic procedure. A retrospective analysis of 10 cases of postoperative infections showed the use of a TNF inhibitor was significantly associated with the development of a serious infection (OR, 4.4). A meta-analysis of the clinical literature on TNF inhibitors through 2005 found an increased risk of serious infection. The risks and benefits must be weighed carefully, and the patients must be fully informed on how these agents should be used on a case-by-case basis.

PREVENTION OF PROSTHETIC JOINT INFECTIONS

There is consensus that preoperative evaluation of a patient for occult infection such as periodontal disease or bacteriuria is warranted, and corrective steps to eradicate any infection are essential before joint replacement. There is also consensus that perioperative antibiotic prophylaxis

![Figure 109-3](https://example.com/fig109-3.png)

Table 109-7  Factors That Might Portend a Poor Outcome in Septic Arthritis

<table>
<thead>
<tr>
<th>Factor</th>
<th>Odds Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Older age</td>
<td>2.5 (1.8–3.4)</td>
</tr>
<tr>
<td>Pre-existing arthritis, especially rheumatoid arthritis, but also osteoarthritis and tophaceous gout</td>
<td>1.5 (1.1–2.1)</td>
</tr>
<tr>
<td>Presence of synthetic material (e.g., total joint replacement)</td>
<td>1.4 (1.0–1.9)</td>
</tr>
<tr>
<td>Delay in diagnosis or long duration of symptoms before seeking medical attention</td>
<td>1.6 (1.1–2.4)</td>
</tr>
<tr>
<td>Polyrarticular infection, especially if &gt;3 joints and small hand joints are affected</td>
<td>2.0 (1.4–2.9)</td>
</tr>
<tr>
<td>Presence of bacteremia</td>
<td>1.8 (1.3–2.5)</td>
</tr>
<tr>
<td>Infection caused by virulent or difficult-to-treat organisms (e.g., <em>Staphylococcus aureus</em>, <em>Pseudomonas aeruginosa</em>, or some gram-negative bacilli)</td>
<td>2.2 (1.6–2.9)</td>
</tr>
<tr>
<td>Patients receiving immunosuppressive therapy</td>
<td>2.1 (1.5–2.9)</td>
</tr>
<tr>
<td>Serious underlying comorbidities (e.g., liver, kidney, or heart diseases)</td>
<td>1.4 (1.0–1.9)</td>
</tr>
<tr>
<td>Peripheral leukocytosis at presentation</td>
<td>1.3 (1.0–1.7)</td>
</tr>
<tr>
<td>Worsening renal function</td>
<td>1.5 (1.1–2.1)</td>
</tr>
</tbody>
</table>

Data from references 36, 43, 49, 120.

or as short-term and long-term outcomes. Among the survivors, loss of articular cartilage, loss of motion, or increase in pain in the affected joint would be considered poor functional outcomes. Loss of the limb to infection and need for surgery to fuse the joint or restore function are also poor outcomes. Most studies report the outcome at the time of hospital discharge, and long-term data on adults with septic arthritis are unavailable. The rate of development of degenerative joint disease, the rate of relapse or recurrence of infection, and the rate of progression of functional impairment in the affected joint over time have not been well studied.

Many retrospective studies have characterized features that may increase the chance of a poor outcome at the time of hospital discharge (Table 109-7). One prospective community-based study of adults and children found poor joint outcome in 33% of survivors among 154 patients with bacterial arthritis and noted older age, pre-existing joint disease, and an infected joint containing synthetic material as negative prognostic factors by univariate analysis. These investigators noted no association between poor outcome and young age, comorbidity, immunosuppressive medication, functional class, multiple infected joints, type of organism, or treatment delay. In a large retrospective study from the United Kingdom on the outcome of 243 patients, 11.5% died secondary to septic arthritis and additional morbidity was noted in 31.6% of patients. Multivariate analysis suggests that important predictors of death are confusion at presentation, age 65 years or older, multiple joint sepsis, and involvement of the elbow joint. Predictors of morbidity were age 65 years or older, diabetes mellitus, open surgical drainage, and gram-positive infections other than *Staphylococcus aureus*.

References

43. Goldenberg DL: Septic arthritis and other infections of rheumato-
47. Perez-Ruiz F, Testillano M, Gastaca MA, et al: Pseudoseptic pseudo-
Lyme Disease

LINDA K. BOCKENSTEDT

KEY POINTS

Lyme disease is caused by infection with tick-transmitted spirochetes of the genus *Borrelia burgdorferi sensu lato* and has a worldwide distribution.

Lyme disease has a characteristic pattern of signs and symptoms and usually begins with the expanding macular skin lesion erythema migrans.

Earlier recognition and treatment has led to a decline in the incidence of carditis, acute neurologic disease, and late disease manifestations.

Musculoskeletal manifestations occur in more than 50% of patients and at all stages of infection, but frank arthritis is a sign of late disease and is uncommon (<10% of patients).

The diagnosis of Lyme disease should be suspected when a patient who lives, works, or vacations in an endemic area presents with signs and symptoms of *B. burgdorferi* infection.

Two-tiered serologic tests (enzyme-linked immunosorbent assay and immunoblot) can be negative with early infection but become positive in most patients with infection of greater than 1 month’s duration.

Most patients are cured with 2 to 4 weeks of antibiotic therapy, although the time to disease resolution may extend beyond the duration of therapy and irreversible tissue damage may occur.

Antibiotic-refractory arthritis occurs in less than 10% of patients with Lyme arthritis, responds to disease-modifying antirheumatic drugs, and typically resolves within 4 to 5 years.

Post-Lyme disease syndrome (persistent debilitating complaints of fatigue, mild cognitive dysfunction, and musculoskeletal pain) after antibiotic treatment for Lyme disease occurs in a minority of patients. *B. burgdorferi* cannot be detected in these patients, and controlled treatment trials show no benefit of prolonged antibiotic therapy over placebo.

Lyme disease is a multisystem disorder caused by the tick-borne spirochete *Borrelia burgdorferi*. The disease first came to medical attention in the United States in the late 1970s with the investigation of a clustering of cases of juvenile arthritis in the region of Lyme, Connecticut. A characteristic skin rash described as single or multiple expanding red macules often heralded the onset of arthritis. This rash, termed *erythema migrans* (EM), had been linked in Europe to the bite of *Ixodes* ticks and the subsequent development of neurologic abnormalities. Further investigation revealed that arthritis was one manifestation of a systemic disorder affecting the skin, heart, joints, and nervous system. In 1982 Burgdorfer isolated the causative agent, the spirochete *B. burgdorferi*, from *Ixodes* ticks. Demonstration that patients with Lyme disease developed antibodies to this organism and its eventual culture from skin, cerebrospinal fluid (CSF), and synovial tissue confirmed the infectious etiology of the disorder. It is now the most common vector-borne disease in the United States.

ECOLOGY AND EPIDEMIOLOGY OF LYME DISEASE

Lyme disease has a worldwide distribution, with most cases reported in North America, Europe, and Asia. On each of these continents, hard-shelled ticks of the *Ixodes* family are the only known vectors for the disease. Other arthropods and blood-sucking insects such as mosquitoes cannot transmit the infection. The incidence of Lyme disease varies geographically and is determined by the prevalence of *B. burgdorferi*-infected ticks. In the United States, cases of Lyme disease have been reported in all 50 states and the District of Columbia, but most are clustered in the Northeast and mid-Atlantic region, upper Midwest, and northern California. In 2009, 29,959 confirmed and 8509 probable cases were reported to the Centers for Disease Control and Prevention, with 93% of confirmed cases originating from 11 states: Pennsylvania, New Jersey, New York, Massachusetts, Connecticut, Wisconsin, Maryland, New Hampshire, Delaware, and Maine.

The spirochetes associated with Lyme disease reside within the genus *B. burgdorferi sensu lato* (sl), and the vast majority of cases are caused by *B. burgdorferi sensu stricto* (ss), *Borrelia garinii*, and *Borrelia afzelii*. All three genospecies can be found in Europe, whereas *B. burgdorferi ss* is the main species found in North America. Variation among the genospecies may account for the differences in clinical expression of Lyme disease between the two continents, with *B. garinii* associated with neurologic disease, *B. afzelii* associated with late skin involvement, and *B. burgdorferi ss* associated with arthritis. Because of the prominence of musculoskeletal manifestations with *B. burgdorferi ss* infection, this chapter focuses mainly on Lyme disease in North America.

TICKS AND LYME DISEASE

Lyme disease is found primarily in temperate climates where humans can have incidental exposure to questing ticks. *Ixodes* ticks have a 2-year life span in which they pass...
through three developmental stages—larva, nymph, and adult—feeding only once per stage. B. burgdorferi is not passed transovarially and is maintained by passage between reservoir hosts and ticks. Small rodents are the main reservoirs for B. burgdorferi ss and B. afzelii, whereas birds are the principal haven for B. garinii. In the southern United States, ticks feed preferentially on lizards, which are not competent reservoirs for B. burgdorferi; this may explain in part the rarity of Lyme disease in this region.

Larvae acquire B. burgdorferi after feeding on an infected reservoir host in early spring and then molt to nymphs, which lay dormant until the following late spring and summer. The peak incidence of Lyme disease is in the summer months, when humans come in contact with questing nymphs, which have more promiscuous feeding patterns. Engorged nymphs molt into adult ticks, which feed almost exclusively on deer. B. burgdorferi does not persist in deer, which serve to maintain and propagate the tick population.

PATHOGENESIS

*Borrelia burgdorferi* Invasion of the Mammalian Host

During tick feeding, B. burgdorferi migrate from the tick midgut to the salivary glands, where they are deposited with saliva into the blood meal host. Migration takes about 24 hours, during which time spirochetes multiply and undergo phenotypic changes that permit their survival in mammals. Spirochetes multiply first at the tick bite site in the skin, and if not eliminated by the cutaneous immune response, they can disseminate through tissues and the bloodstream to infect any organ system at least transiently. The degree to which B. burgdorferi cause disease in tissues depends on spirochete virulence, growth conditions that allow persistence at a particular site, and host factors that modulate the inflammatory response.

Analysis of the B. burgdorferi genome has revealed no known virulence factors common to other bacterial pathogens to help explain the pathogenesis of Lyme disease. Instead, the genome is remarkably rich in genes encoding putative lipoproteins, only a handful of which have been studied in detail. Outer surface protein (Osp) A is a midgut adhesin required for spirochete infection of ticks. Osp C is essential for initial infection of the mammal but is dispensable after spirochetes have disseminated and colonized other tissue. To do so, B. burgdorferi harnesses host plasmin to move through tissues and expresses adhesins including decorin binding proteins A and B, BBK32, and p66, which allow it to bind to extracellular matrix proteins and integrins on cells. Expression of VlsE, an Osp that undergoes antigenic variation, is required for infection to persist in immunocompetent hosts.

Pathology of Lyme Disease

Because intact spirochetes are seen only rarely in tissue specimens and the spirochete genome reveals no known toxins, the pathology of Lyme disease is believed to be due to the host inflammatory response to B. burgdorferi components rather than tissue destruction by the spirochete itself. Histopathologic studies of EM lesions, cardiac tissue, synovial biopsy specimens, and limited nervous system tissue (meninges, spinal cord, and nerve roots) reveal varying degrees of mononcytic and lymphoplasmacytic infiltrates, especially perivascular, that stain positively for cell surface markers for macrophages, T cells, and B cells. The joint effusions of patients with Lyme arthritis reveal acute inflammation with elevated leukocyte counts, whereas the synovium resembles that of rheumatoid arthritis, with chronic inflammation mediated by mononuclear cell infiltration and pseudolymphoid follicles formed by T cells, B cells, and plasma cells. In the synovium and less commonly the epineural area, perivascular infiltrates can be associated with endarteritis obliterans.

**Immune Response to Borrelia burgdorferi**

Innate immune cells respond to B. burgdorferi through engagement of the Toll-like receptor (TLR) family of pattern recognition receptors, especially TLR2/TLR1 heterodimers (lipoproteins), TLR5 (flagellin), and TLR9 (spirochete DNA). As a consequence, proinflammatory cytokines (including interleukin-1β [IL-1β] and tumor necrosis factor), chemokines (IL-8), nitric oxide, and prostaglandins that recruit inflammatory cells to the site of infection are produced. B. burgdorferi also induces matrix metalloproteinase expression in tissues through TLR-dependent and non–TLR-dependent pathways that contribute to pathology. Other pattern recognition receptors may be engaged by B. burgdorferi after its ingestion by phagocytes, including the intracellular NOD2 receptors, which respond to peptidoglycan and have been shown to potentiate the inflammatory response in vitro. Humoral immunity is a key host defense against B. burgdorferi infection. B. burgdorferi lipoproteins are B cell mitogens, and antibodies that arise in the absence of T cell help are sufficient to resolve inflammation and prevent challenge infection in the mouse model of Lyme borreliosis. With the induction of adaptive immunity, IgG-containing immune complexes and cryoglobulins can be found in the serum of patients with Lyme disease and are concentrated in the joints of patients who develop Lyme arthritis. B cell–recruiting chemokines such as CXCL13 and pathogen-specific antibody production can be found in the CSF of patients with neuroborreliosis, some of these antibodies can also bind neutral antigens. B. burgdorferi infection primes CD4+ and CD8+ T cells, and the predominance of T helper type 1 responses correlates with more severe arthritis and neuroborreliosis. Th17 cells are also involved, as demonstrated by the finding that the B. burgdorferi neutrophil-activating protein A (NapA) can elicit IL-17 from synovial fluid T cells ex vivo. There is an association between T cell and B cell responses to Osp A and the development of antibiotic-refractory Lyme arthritis. Although evidence has been presented to suggest an autoimmune etiology (see later section on antibiotic-refractory arthritis), the self-limited nature of Lyme arthritis also raises the possibility that the immune responses detected are appropriate and directed toward eliminating persisting antigens rather than viable organisms. Alternatively, prolonged arthritis may be due to abnormal or delayed regulation of the host immune response.
when the pathogen and its inflammatory products have been eliminated. Deficiency in CD25+ T regulatory cells prolongs murine Lyme arthritis, and synovial fluid γδ T cells isolated from patients with Lyme arthritis can modulate B. burgdorferi–specific CD4+ T cell responses by inducing apoptosis in a Fas-dependent fashion. 

Mechanisms of Spirochete Persistence

When visualized in vivo, B. burgdorferi resides primarily in the extracellular matrix in connective tissue. Despite occasional sightings of spirochetes inside cells, an intracellular phase of the B. burgdorferi life cycle has not been shown. B. burgdorferi employs immune evasion strategies of an extracellular pathogen, which are directed toward deterring phagocyte ingestion and antibody- and complement-mediated lysis. B. burgdorferi expresses Erp and complement regulator–acquiring surface proteins that bind host factor H to prevent complement-mediated lysis. To impede antibody-mediated clearance, B. burgdorferi undergoes antigenic variation and reduces expression of lipoproteins as infection progresses. The vsbE gene undergoes random rearrangement of its expression locus, producing antigenically distinct variants of VlsE, a protein essential for spirochete survival in vivo. In the chronic phase of B. burgdorferi infection in mice, spirochetes can be visualized in the extracellular matrix of connective tissue, especially in the skin, without an associated inflammatory response.

CLINICAL FEATURES OF LYME DISEASE

Lyme disease occurs in stages that reflect the immune response to the spirochete as it establishes infection in the skin and later disseminates to distant organ sites (Table 110-1). Presenting clinical manifestations depend on the stage of the illness in which patients first seek medical attention. A characteristic feature of Lyme disease is that clinical signs can resolve without specific therapy, and patients may present in later stages of the illness without exhibiting signs of early disease.

Early Localized Infection

The hallmark of Lyme disease is the skin lesion EM, which is present in about 80% of patients (Figure 110-1). The lesion arises within 1 month (median, 7 to 10 days) at the tick bite site, especially in skin folds or where clothes bind in adults and around the hairline in children. EM begins as a red macule that expands at the rate of 2 to 3 cm/day, enlarging to more than 70 cm in diameter. Characteristic lesions greater than 5 cm in diameter in an appropriate clinical setting are sufficient for establishing the diagnosis of Lyme disease. EM most often manifests with uniform erythema, but central clearing can occur in larger lesions, producing a classic “bull’s eye” appearance (see Figure 110-1B). Vesicular or necrotic centers are rarer (see Figure 110-1D), but even these EM lesions have relatively few symptoms other than a tingling or burning sensation. Intense pruritus or pain is unusual and should raise concern for alternative diagnoses.

EM may be accompanied by systemic flulike symptoms including low-grade fever, malaise, neck pain or stiffness, arthralgias, and myalgias. Particularly severe systemic symptoms should alert the physician to possible coinfection with another tick-borne pathogen such as Babesia microti or Anaplasma phagocytophilum (the agent of human granulocytic anaplasmosis, formerly known as human granulocytic ehrlichiosis). Lyme disease can also manifest with systemic symptoms alone. Absence of upper respiratory or gastrointestinal symptoms may help distinguish Lyme disease from common viral infections. Musculoskeletal complaints and debilitating fatigue associated with Lyme disease should be distinguished from fibromyalgia and chronic fatigue syndrome, which are typically more insidious in onset and are not associated with objective findings or laboratory abnormalities.

A newly recognized southern tick–associated rash illness (STARI) can produce a skin lesion similar to the bull’s eye form of EM. The rash is associated with the bite of the Lone Star tick, Amblyomma americanum, which is endemic to the southeastern and south-central states, but which also can be found as far north as Maine or west as central Texas and Oklahoma. Similar to EM, systemic symptoms can accompany the rash of STARI, but disease in organs other than the skin does not occur. The etiology of STARI is unknown. Although a noncultivable spirochete named Borrelia lonesi has been found in A. americanum, STARI

### Table 110-1 Clinical Manifestations of Lyme Disease

<table>
<thead>
<tr>
<th>Stage</th>
<th>Clinical Manifestations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Localized Infection</td>
<td>Occurs 3 to 30 days after tick bite&lt;br&gt;EM in 80%-90% of patients; single lesion, occasionally associated with fever, malaise, neck pain or stiffness, arthralgias and myalgias&lt;br&gt;Systemic symptoms noted above in the absence of EM during summer months&lt;br&gt;Borrelial lymphocytoma (rare, seen primarily in Europe)</td>
</tr>
<tr>
<td>Early Disseminated Infection</td>
<td>Occurs weeks to months after tick bite&lt;br&gt;Profound malaise and fatigue common&lt;br&gt;Multiple EM lesions with systemic symptoms similar to early localized infection&lt;br&gt;Musculoskeletal&lt;br&gt;Migratory polyarthralgias and myalgias&lt;br&gt;Carditis (&lt;3% of untreated patients)&lt;br&gt;Varying degrees of atrioventricular nodal block&lt;br&gt;Mild myopericarditis&lt;br&gt;Neurologic (&lt;10% of untreated patients)&lt;br&gt;Cranial neuropathies (especially facial nerve palsy)&lt;br&gt;Lymphocytic meningitis&lt;br&gt;Radiculoneuropathies&lt;br&gt;Encephalomyelitis</td>
</tr>
<tr>
<td>Late Disease</td>
<td>Occurs months to years after tick bite&lt;br&gt;Arthritis (&lt;10% of patients)&lt;br&gt;Acute monoarticular or migratory pauciarticular inflammatory arthritis, usually involving the knee&lt;br&gt;Chronic antibiotic-refractory arthritis (&lt;10% of patients with arthritis)&lt;br&gt;Neurologic (rare)&lt;br&gt;Peripheral neuropathies&lt;br&gt;Mild encephalopathy&lt;br&gt;Encephalomyelitis (primarily seen in Europe)&lt;br&gt;Skin&lt;br&gt;Acrodermatitis chronica atrophicans (primarily seen in Europe)</td>
</tr>
</tbody>
</table>

EM, erythema migrans.
Cardiac Disease

The incidence of cardiac involvement has declined to 1% in recent years, possibly owing to earlier recognition and treatment of *B. burgdorferi* infection. It most often occurs within the first 2 months of infection and manifests as varying degrees of atrioventricular block, occasionally accompanied by mild myopericarditis. Electrophysiologic studies have mapped the conduction defect to the area above the bundle of His and involving the atrioventricular node, although multiple levels can be affected. Overt congestive heart failure is rare, and chronic cardiomyopathy, reported in Europe, has not been documented to occur in the United States. Patients with Lyme carditis often have a history of EM and may have concomitant arthralgia and myalgia at the time of presentation. Absence of valvular heart disease helps distinguish Lyme carditis from acute rheumatic fever, and prominent myocardial dysfunction or pericardial involvement should suggest other infectious etiologies.

Skin Disease

Fifty percent of patients with untreated Lyme disease develop multiple EM lesions, a sign of disseminated infection (see Figure 110-1C). Secondary lesions are typically smaller and can occur anywhere on the body, but they are most noticeable on the trunk. The lesions usually appear as flat macules and can develop partial central clearing. EM lesions may be accompanied by migratory muscle, joint, and periarticular pain that lasts hours to days, but frank arthritis is now considered a late manifestation of the disease.

Early Disseminated Infection

Within weeks of the onset of infection, *B. burgdorferi* can disseminate through the skin, blood, and lymphatics to infect multiple tissues. Clinically apparent disease at this stage, however, is usually seen in the skin, heart, or nervous system. Patients with disseminated infection have debilitating fatigue and appear ill. Specific localizing signs and symptoms may fluctuate, but profound fatigue is a consistent complaint.

Nervous System Involvement

Acute neurologic Lyme disease occurs in less than 10% of patients and most commonly manifests as cranial nerve palsy or meningitis, although radiculopathy and encephalomyelitis are also occasionally seen. Cranial palsy occurs in 8% of cases and usually affects the seventh nerve,
resulting in unilateral or bilateral facial palsy. Even in endemic areas, however, onset of seventh nerve palsy in the nonwinter months is due to *B. burgdorferi* infection in only 25% of cases. Bilateral facial palsy is seen in only a few other conditions—Guillain-Barré syndrome, human immunodeficiency virus infection, sarcoidosis, and other causes of chronic meningitis—all of which are readily distinguished from Lyme disease. Rarely, other cranial nerves (III, IV, V, VI, or VIII) may be involved. Lyme meningitis manifests with fever, headache, and stiff neck similar to viral meningitis, along with a CSF lymphocytosis and elevated protein. In children, meningitis may occur with EM, cranial nerve involvement, and increased intracranial pressure (papilledema), which is rare in adults. Lyme radiculopathy typically manifests as pain, weakness, numbness, and reflex loss in a dermatomal distribution, resembling mechanical radiculopathies. Lyme disease should be considered when there is no obvious precipitating factor for disk-related symptoms, and imaging studies do not delineate pathology at the appropriate root level. Untreated Lyme radiculopathy can progress to become bilateral, which helps distinguish it from mechanical disease. When truncal involvement causes unilateral chest or abdominal pain, Lyme radiculopathy is often mistaken for visceral disease or early herpes zoster before the development of vesicular lesions.

**Other Organ System Involvement**

A variety of other organs can exhibit pathology with disseminated *B. burgdorferi* infection including the eye (keratitis), the ear (sensorineural hearing loss), the liver (hepatitis), the spleen (necrosis), skeletal muscle (myositis), and subcutaneous tissue (panniculitis). In general, other, more classic manifestations of Lyme disease are present concurrently or have been present in the recent past to suggest the diagnosis.

**Late Disease**

Months after the onset of infection, untreated patients can develop late manifestations of Lyme disease, usually involving the joints (discussed separately later), nervous system, and the skin. At this stage of infection, two-tier (enzyme-linked immunosorbent assay [ELISA] and IgG immunoblot) serologic testing for *B. burgdorferi* should be positive.

**Late Neurologic Disease**

Late neurologic Lyme disease is now rare; patients may present with encephalomyelitis, peripheral neuropathy, or encephalopathy. *E. coli* Encephalomyelitis, seen predominantly in Europe with *B. garinii* infection, is a slowly progressive, unifocal or multifocal inflammatory disease of the central nervous system, with increased T2 signals in the white matter on magnetic resonance imaging (MRI). CSF examination often reveals a lymphocytic pleocytosis, elevated protein, and normal glucose, and serum IgG to *B. burgdorferi* and intrathecal antibody production can be found. These findings help distinguish Lyme encephalomyelitis from multiple sclerosis, which may rarely be associated with positive IgG reactivity to *B. burgdorferi* in serum and CSF samples, but there is no intrathecal antibody production. Multiple sclerosis patients with positive Lyme serologies do not respond to antibiotics used for neurologic Lyme disease.

Late peripheral nervous system involvement manifests as a mild sensorimotor neuropathy in a “stocking and glove” distribution, with evidence of a mild confluent mononeuritis multiplex on electrophysiologic studies. Patients may have intermittent limb paresthesias and occasionally radicular pain. The most common finding on physical examination is reduced vibratory sensation in the lower extremities. Serum IgG to *B. burgdorferi* should be present, but CSF examination is normal, consistent with disease confined to the peripheral nervous system. Patients with this form of neuropathy should be evaluated for other infectious diseases (syphilis, human immunodeficiency virus, and hepatitis C virus); metabolic disorders (especially vitamin B12 deficiency, diabetes mellitus, and thyroid disease); and autoimmune diseases (antinuclear antibody [ANA] or rheumatoid factor associated).

Patients with Lyme encephalopathy complain of memory impairment and cognitive dysfunction that are best shown by formal neuropsychological testing. Occasionally, patients may have CSF abnormalities with elevated protein, lymphocytic pleocytosis, and intrathecal antibody to *B. burgdorferi*, but CSF examination can also be normal. Serum IgG to *B. burgdorferi* should be present, however, to consider the diagnosis. The mild cognitive dysfunction seen in patients with Lyme encephalopathy must be distinguished from neurocognitive deficits secondary to chronic stress, sleep deprivation, fibromyalgia, chronic fatigue syndrome, or aging. As for any chronic encephalopathy, toxic-metabolic causes should be excluded. Brain imaging studies are generally normal or show only nonspecific abnormalities and are not useful in establishing a diagnosis of encephalopathy associated with Lyme disease.

**Late Skin Disease**

The late skin lesion acrodermatitis chronica atrophicans is found mainly in Europe because of its association with *B. afzelii* infection, although any *B. burgdorferi* species can cause the lesion. Acrodermatitis chronica atrophicans develops insidiously over years and is most often found on the dorsum of the hands or feet. It begins as a unilateral bluish red discoloration and swelling, which evolves to atrophic, cellophane-like skin with prominent appearance of the blood vessels. About 60% of patients also have a peripheral sensory neuropathy affecting the involved extremity. A prominent lymphoplasmacytic infiltrate is shown on the skin biopsy specimen. Antibiotics can lead to improvement in pain and swelling, but atrophic skin remains.

**LYME ARTHRITIS AND OTHER MUSCULOSKELETAL MANIFESTATIONS OF LYME DISEASE**

Musculoskeletal symptoms are common in all stages of Lyme disease and include migratory pain in joints, tendons, bursae, and muscles. Typically, musculoskeletal pain affects one or two sites at a time, lasts only hours to a few days at any one location, and is associated with significant fatigue.
The incidence of frank arthritis has declined from 50% in early studies to less than 10% in more recent years. ELISA and IgG immunoblot for *B. burgdorferi* are positive when arthritis appears, and *B. burgdorferi* DNA can be detected by polymerase chain reaction (PCR) in synovium and synovial fluid even though cultures are usually negative. Although Lyme arthritis can resemble pauciarticular juvenile arthritis or reactive arthritis, patients generally test negative for ANA, rheumatoid factor, and anticitrullinated protein antibodies (ACPA) and do not have an increased frequency of HLA-B27 alleles. Joint fluid analysis and synovial histopathology cannot distinguish these entities. Axial and sacroiliac joint involvement is not a feature of Lyme disease, but enthesis can be seen. Most patients with Lyme arthritis have positive two-tier serologic tests for *B. burgdorferi* infection.

Arthritis usually begins months or years after *B. burgdorferi* infection and is predated by migratory arthralgias in half of patients. The most typical pattern is a monoarticular or oligoarticular arthritis involving one or a few large joints (fewer than five total), with the knee affected in 80% of cases. Joints are warm with large effusions, often greater than 100 mL in the knee, but comparatively little pain. Synovial fluid is inflammatory with white blood cell counts ranging from approximately 2000 to 70,000/mm$^3$ (median = 24,000/mm$^3$), with a predominance of neutrophils. Depending on the chronicity of the arthritis, synovial biopsy specimens reveal only mononuclear cell infiltration or more advanced changes consistent with rheumatoid synovium. Large effusions can lead to Baker’s cyst formation and rupture. The temporomandibular joint is also frequently involved and in one study was the first joint to be affected in 25% of patients with arthritis. Other joints commonly affected include the shoulder, ankle, elbow, wrist, and hip. Lyme arthritis is often intermittent, with episodes lasting a few weeks to months. Recurrent episodes are notable for smaller effusions and progressive synovial hypertrophy, bony erosion, and cartilage destruction. A small percentage (<10%) of patients with intermittent arthritis settle into a pattern of chronic arthritis, generally affecting only a single joint and often the knee. Inflammation of a single joint that persists for more than 12 months would be an unusual presenting manifestation of Lyme arthritis, as is the prominent involvement of small joints.

The natural history of Lyme arthritis suggests that it is a self-limited disorder. In the late 1970s, before the use of antibiotics for Lyme disease, 21 patients who presented with EM and later developed Lyme arthritis were followed for 1 to 8 years without antimicrobial therapy. Six patients had only a single episode of arthritis, and the remaining 15 had recurrent episodes that decreased in frequency over the study period. On average, the number of patients who continued to experience episodes of arthritis decreased by 10% to 20% each year. Similar results were found in children in whom antibiotic treatment for arthritis was delayed 4 years.

**Antibiotic-Refractory Lyme Arthritis**

A few patients treated with standard antibiotic regimens for Lyme arthritis have persistent joint inflammation and proliferative synovitis that does not respond to further antimicrobial therapy. The pathogenesis of “antibiotic-refractory” Lyme arthritis is unknown but may be due to persistent spirochetes or their antigens, infection-induced autoimmunity, or inadequate regulation of the inflammatory response. Spirochete virulence may play a role as a retrospective study of archived tissue samples obtained before treatment revealed that patients who went on to develop antibiotic-refractory arthritis had a higher prevalence of infection with the more invasive RST1 *B. burgdorferi* strains. Although patients with antibiotic-refractory Lyme arthritis no longer have PCR evidence for spirochete DNA in tissues, experiments in the mouse model of Lyme borreliosis suggest that spirochete debris, including *B. burgdorferi* DNA, can persist near cartilage and in the entheses for extended periods after infectious spirochetes have been killed with antibiotics, particularly when the initial pathogen burden was high. Recently, a single nucleotide polymorphism in TLR1 (TLR1 1805 GG) that impairs the innate immune response to *B. burgdorferi* was found to be more prevalent among patients who developed antibiotic-refractory Lyme arthritis. A genetic predisposition had been suggested in earlier studies that found an increased frequency of the rheumatoid arthritis–related alleles HLA-DRB1*0401, HLA-DRB1*0101, and HLA-DRB1*0404 in patients with antibiotic-refractory Lyme arthritis. Because of the high prevalence of B cell and T cell responses to *B. burgdorferi* Osp A in patients with antibiotic-refractory arthritis, it had been proposed that immune responses to Osp A triggered by infection may be perpetuated by a self-antigen after the pathogen has been eliminated. An Osp A peptide corresponding to amino acids 163 through 175 (Osp A 163-175) was found to share an epitope with human leukocyte function–associated antigen 1α, an adhesion molecule expressed on inflamed tissues. The leukocyte function–associated antigen 1α peptide stimulated Osp A-specific T cells only weakly, however, and did not promote production of the T helper (Th1) cytokine interferon-γ normally found in antibiotic-refractory arthritis. Antibodies to cytokeratin 10, a constituent of synovial capillaries, have been found in the blood and synovial tissue of patients with antibiotic-refractory Lyme arthritis. These antibodies also react with Osp A and may contribute to ongoing inflammation when infection is cleared. Linked T and B cell responses to peptides of epidermal cell growth factor have been found in about 50% of patients with antibiotic-refractory Lyme arthritis, but their role in perpetuating the inflammatory response is unclear. If autoimmunity is responsible for antibiotic-refractory Lyme arthritis, it must eventually succumb to immune regulation because even this form of Lyme arthritis generally resolves within 4 to 5 years. In this regard, the presence of a higher percentage of T regulatory cells correlates with more rapid resolution of joint inflammation after treatment for antibiotic-refractory Lyme arthritis.

**DIAGNOSIS**

The diagnosis of Lyme disease should be considered in individuals who present with an appropriate clinical history and who have a reasonable risk of exposure to *B. burgdorferi*–infected ticks (Figure 110-2). Supporting serologic evidence is necessary to secure the diagnosis for all stages of
infection except for early localized disease in which EM can be recognized by morphologic features alone. Routine laboratory tests are nonspecific, with some patients exhibiting mildly elevated white blood cell (neutrophil) counts, erythrocyte sedimentation rates, and liver function tests. Culture or microscopic visualization of spirochetes in clinical samples is not sensitive enough for routine use in diagnosis. Culture of a skin biopsy specimen taken from the leading margin of an EM lesion is an exception, with B. burgdorferi detected in more than 40% of samples, but is rarely necessary to identify EM.

**Serologic Testing**

Detection of antibodies to B. burgdorferi is the mainstay of laboratory testing for Lyme disease. Presence of antibodies to B. burgdorferi at best indicates previous exposure to the organism, however, and should not be considered evidence of active infection. In nonendemic areas, about 5% of normal human serum samples yield positive results on serologic tests for Lyme disease. In endemic areas, asymptomatic IgG seroconversion to B. burgdorferi has been found in about 7% of subjects.

A two-tiered approach is recommended for detection of B. burgdorferi–specific antibodies. An ELISA or an indirect immunofluorescence assay to detect IgM and IgG reactivity to B. burgdorferi should be used as an initial screening test, followed by an immunoblot (Western blot) to confirm that positive or equivocal results are due to antibodies that bind B. burgdorferi antigens. ELISA and immunofluorescence assays are highly sensitive tests but lack specificity because of cross-reactivity of antibodies to B. burgdorferi with other bacterial pathogens. A positive or equivocal ELISA or immunofluorescence assay should be confirmed by immunoblot analysis of B. burgdorferi proteins (Table 110-2). Banding patterns characteristic of early infection include antibodies to the 41-kD flagellin protein and OspC, which has a molecular weight ranging from 21 to 24 kD depending on the B. burgdorferi strain used (see Figure 110-2) With disseminated infection and especially with late Lyme disease, IgG reactivity to an expanding array of B. burgdorferi proteins can be seen (see Figure 110-2).

Some commercial laboratories have developed assays using recombinant antigens or employ criteria for the interpretation of immunoblots that have not been validated and published in peer-reviewed literature. For this reason, the Centers for Disease Control and Prevention advises using only validated tests approved by the U.S. Food and Drug Administration (FDA) for serologic diagnosis of Lyme disease.

Two-tier testing for IgM and IgG should be performed for individuals with suspected Lyme disease and signs and symptoms of less than 1 month’s duration, whereas only IgG results should be considered for illnesses of longer duration. Positive IgM serologies alone after 1 month of illness are most often false-positive tests, which occur in the setting of other infectious diseases (especially infectious mononucleosis and other spirochetal and tick-borne infections), rheumatoid arthritis (with or without rheumatoid factor), and conditions associated with a positive ANA (systemic lupus erythematosus). No further testing is recommended if ELISA or immunofluorescence assay results are negative. Two-tier testing has overall sensitivities of 29% to 40% in EM during the acute phase; 29% to 78% for EM in the first month of infection; 39% to 69% for EM in the first month of infection; and 59% to 78% for EM in the first month of infection.

<table>
<thead>
<tr>
<th>Duration of Disease</th>
<th>Isotype Tested</th>
<th>Criteria for Positive Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>First month of infection</td>
<td>IgM</td>
<td>2 of the following 3 bands are present: 23 kD (OspC), 39 kD (BmpA), and 41 kD (Fia)</td>
</tr>
<tr>
<td>After first month of infection</td>
<td>IgG</td>
<td>5 or 10 bands are present: 18 kD, 21 kD, 28 kD, 30 kD, 39 kD, 41 kD, 45 kD, 58 kD (not GroEL), 66 kD, and 93 kD</td>
</tr>
</tbody>
</table>

convalescent phase; and greater than 95% in neurologic, arthritis, and other manifestations of late disease.36

A peptide-based ELISA that uses a highly conserved invariant region of the VlsE protein, termed C6 (IR6), is now commercially available.131,132 The C6 peptide ELISA has a high degree of sensitivity and specificity in all stages of Lyme disease and may be particularly useful in early Lyme disease.102 A positive or equivocal test should be confirmed with IgM and IgG immunoblots.

After antibiotic therapy, IgM and IgG titers to B. burgdorferi measured by either whole cell ELISA or the C6 peptide ELISA (IgG only) generally decrease slowly but can remain positive for years.103,104 Repeat serologic testing is not recommended as a means for assessing response to treatment.

Detection of Antibodies to Borrelia burgdorferi in Cerebrospinal Fluid

In suspected cases of neuroborreliosis, intrathecal antibody production is usually assessed by measuring the ratio of IgG to B. burgdorferi in CSF and serum.36,105 Intrathecal antibody production is more commonly found in European neuroborreliosis than in North American Lyme disease, which may be due to the higher prevalence of B. garinii in central nervous system (CNS) infection than B. burgdorferi ss. Antibodies to B. burgdorferi may persist in CSF after treatment for Lyme disease and should not be used to assess efficacy of therapy.

Polymerase Chain Reaction

PCR has been used to detect B. burgdorferi DNA in a variety of clinical specimens with variable success.106 The greatest utility of PCR clinically is in Lyme arthritis, in which the sensitivity of PCR for detection of B. burgdorferi DNA is 85%.107 In contrast, the sensitivity of PCR for detecting B. burgdorferi DNA in CSF is low (<40%) and most often positive in patients with a CSF pleocytosis. PCR of urine specimens is not recommended because of inconsistent sensitivity and documented nonspecific amplification of non–B. burgdorferi DNA targets.36 Although certain commercial laboratories currently offer PCR tests for B. burgdorferi DNA in blood or urine specimens, these have not been validated.39 There are no FDA-approved tests for PCR-based molecular techniques for detecting B. burgdorferi DNA in patient specimens.

Other Tests for Lyme Disease

A urine antigen test, immunofluorescent staining for cell wall–deficient forms of B. burgdorferi, and lymphocyte transformation assays are offered by some commercial laboratories to aid in the diagnosis of Lyme disease. These tests have not been adequately validated for accuracy or clinical usefulness, and the Centers for Disease Control and Prevention cautions against their use.99

Diagnostic Imaging

Imaging studies have a limited role in the evaluation of patients with Lyme disease because no feature is sufficiently distinctive to confirm the diagnosis. Plain radiographs of arthritic joints show changes consistent with an inflammatory arthropathy including joint effusions, synovial hypertrophy, periarticular osteoporosis, cartilage loss, bony erosions, and calcified entheses.108 MRI of arthritic joints can confirm the radiographic findings and reveal associated myositis and adenopathy, which may be useful in distinguishing Lyme arthritis from septic arthritis in children.109

Cranial and spinal MRI findings in neuroborreliosis can reveal focal nodular lesions or patchy white matter lesions on T2-weighted images, consistent with inflammatory or demyelinating processes.110,111 These lesions typically resolve after treatment for Lyme disease,111 in some cases only after several years.112 In patients with post–Lyme disease syndrome, cerebral MRI and the more sensitive technique of fluid-attenuated inversion recovery are normal in about 50% of cases or show nonspecific findings of small white matter lesions.113 Positron emission tomography and single-photon emission computed tomography studies are often normal or show only nonspecific changes with subcortical and cortical hypoperfusion.114,115

TREATMENT AND OUTCOME

Updated guidelines for the clinical assessment (Figure 110-3) and treatment (Table 110-3) of Lyme disease have been published.116 Because many of the manifestations of Lyme disease resolve without specific therapy, the goal of antibiotic treatment is to hasten resolution of signs and symptoms and to prevent later clinical manifestations. Generally, oral antibiotics are sufficient therapy for EM, disseminated EM, uncomplicated facial palsy, mild carditis (first-degree atrioventricular block), and arthritis. Disseminated infection and late manifestations of Lyme disease may require longer courses of antibiotics, and there is often a greater lag time to symptom resolution compared with early disease. Doxycycline is the antibiotic of choice in nonpregnant adults and children 8 years old and older because it is also effective against A. phagocytophilum, which may occur with early Lyme disease.116 Amoxicillin and cefuroxime axetil are acceptable alternatives for the treatment of EM, facial palsy, and other non-neurologic manifestations of Lyme disease. Macrolide antibiotics are less effective than other antimicrobials and should be used only in individuals who cannot take doxycycline, amoxicillin, or cefuroxime axetil. First-generation cephalosporins are not effective therapy for Lyme disease.

Documented nervous system involvement (other than isolated facial palsy) and symptomatic cardiac involvement are the two main indications for intravenous antibiotic therapy.116 Ceftriaxone administered intravenously for 2 to 4 weeks is the preferred antimicrobial, with parenteral cefotaxime or penicillin G acceptable alternatives. There is increasing evidence, however, that oral doxycycline, which is well absorbed and has a high CNS penetration, may be effective for meningitis or radiculopathy.117-119 Lumbar puncture is recommended in individuals with cranial nerve palsies who have symptoms of meningeal irritation because a CSF pleocytosis would be an indication to treat with intravenous therapy. Asymptomatic CSF pleocytosis can occur in the setting of facial palsy and is not an indication
for intravenous therapy. Repeat treatment is not recommended for chronic neurologic abnormalities, unless objective signs of relapse are present.

Patients with symptomatic cardiac involvement (chest pain, shortness of breath, syncope) or with significant conduction system disease (first-degree atrioventricular block with P-R intervals ≥0.3 msec, or second-degree or third-degree block) should be hospitalized for cardiac monitoring and intravenous antibiotic therapy. Consultation with a cardiologist is recommended, and placement of a temporary pacemaker may be necessary. Oral antibiotics can be substituted for intravenous antibiotics at the time of hospital discharge to complete the course of therapy. For arthritis, a 1-month course of oral doxycycline or amoxicillin is recommended, with a repeat course of oral therapy if inflammation does not resolve within 3 months of treatment. For patients with moderate-to-severe joint swelling after a 1-month course of oral antibiotics, intravenous ceftriaxone for 2 to 4 weeks can be used; when inflammation is mild, an additional 1-month course of oral antibiotics can be considered, although arthritis usually resolves without additional therapy. Longer courses of
**Table 110-3**  **Recommended Treatment of Lyme Disease**

<table>
<thead>
<tr>
<th>Manifestation</th>
<th>Drug</th>
<th>Adult Dosage</th>
<th>Pediatric Dosage</th>
<th>Duration (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erythema migrans (recommended)</td>
<td>Doxycycline(^1)</td>
<td>100 mg orally (PO) bid</td>
<td></td>
<td>14 days (10-21 days)</td>
</tr>
<tr>
<td></td>
<td>Amoxicillin</td>
<td>500 mg PO tid</td>
<td></td>
<td>14 days (14-21 days)</td>
</tr>
<tr>
<td></td>
<td>Cefuroxime axetil</td>
<td>500 mg PO bid</td>
<td></td>
<td>14 days (14-21 days)</td>
</tr>
<tr>
<td>Erythema migrans (alternative)(^a)</td>
<td>Azithromycin</td>
<td>500 mg PO qd</td>
<td></td>
<td>7-10 days</td>
</tr>
<tr>
<td></td>
<td>Clarithromycin</td>
<td>500 mg PO bid (if patient is nonpregnant)</td>
<td></td>
<td>14-21 days</td>
</tr>
<tr>
<td></td>
<td>Erythromycin</td>
<td>500 mg PO qd (if patient is nonpregnant)</td>
<td></td>
<td>14-21 days</td>
</tr>
<tr>
<td>Acute neurologic disease</td>
<td>Penicillin G</td>
<td>18-24 million U/day, divided q4h</td>
<td></td>
<td>14 days (14-21 days)</td>
</tr>
<tr>
<td>Cranial nerve palsy(^b)</td>
<td>Same as oral regimens for erythema migrans Ceftaxone</td>
<td>2 g IV qd</td>
<td>14 days (14-21 days)</td>
<td></td>
</tr>
<tr>
<td>Meningitis or radiculopathy(^c)</td>
<td>Cefotaxime</td>
<td>2 g IV q8h</td>
<td>150-200 mg/kg/day IV in 3-4 divided doses (max 6 g/day)</td>
<td>14 days (10-28 days)</td>
</tr>
<tr>
<td>(Alternative IV)</td>
<td>Penicillin G</td>
<td>200,000-400,000 U/kg/day divided q4h (max 18-24 million U/day)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac disease(^d)</td>
<td>Same as for erythema migrans OR IV regimen as for neurologic disease</td>
<td></td>
<td>14 days (14-21 days)</td>
<td></td>
</tr>
<tr>
<td>Late disease</td>
<td>Penicillin G</td>
<td>200,000-400,000 U/kg/day divided q4h (max 18-24 million U/day)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arthritis without neurologic involvement</td>
<td>Same as for erythema migrans OR IV regimen as for neurologic disease</td>
<td></td>
<td>14 days (14-21 days)</td>
<td></td>
</tr>
<tr>
<td>Recurrent arthritis after oral regimen</td>
<td>Penicillin G</td>
<td>200,000-400,000 U/kg/day divided q4h (max 18-24 million U/day)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central or peripheral nervous system disease</td>
<td>Penicillin G</td>
<td>200,000-400,000 U/kg/day divided q4h (max 18-24 million U/day)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Complete response to treatment may be delayed beyond the treatment period, regardless of the clinical manifestation, and relapse may recur. Patients with objective signs of relapse may need a second course of treatment.

\(^1\)Tetracyclines are relatively contraindicated in pregnant or lactating women and in children younger than 8 yr of age. \(^2\)Due to their lower efficacy, macrolides are reserved for patients who are unable to take or who are intolerant of tetracyclines, penicillins, and cephalosporins. \(^3\)Patients without clinical evidence of meningitis may be treated with an oral regimen. The recommendation is based on experience with seventh cranial nerve palsy. Whether oral therapy would be as effective for patients with other cranial neuropathies is unknown; the decision between oral and parenteral therapy should be individualized. \(^4\)For nonpregnant adult patients intolerant of \(\beta\)-lactam agents, doxycycline 200-400 mg/day orally (or IV if unable to take oral medications) in 2 divided doses may be adequate. For children 8 yr of age and older, the dosage of doxycycline for this indication is 4-8 mg/kg/day in two divided doses (maximum daily dosage of 200-400 mg). \(^5\)A parenteral antibiotic regimen is recommended at the start of therapy for patients who have been hospitalized for cardiac monitoring; an oral regimen may be substituted to complete a course of therapy or to treat outpatients. A temporary pacemaker may be required for patients with advanced heart block.

Antibiotics provide no additional benefit when PCR for *B. burgdorferi* in joint fluid is negative. In this situation, nonsteroidal anti-inflammatory drugs and hydroxychloroquine are recommended for treatment of antibiotic-refractory Lyme arthritis. In rare patients who fail to respond to this approach, other disease-modifying antirheumatic drugs such as methotrexate and tumor necrosis factor inhibitors have been used anecdotally with success. Arthroscopic synovectomy is curative in most patients who fail to respond to medical management. Intra-articular corticosteroids may be associated with a higher rate of antibiotic unresponsiveness and are rarely used.

**Pregnancy and Lyme Disease**

Pregnant and lactating women with Lyme disease can be treated with the same antibiotic regimens recommended for nonpregnant patients except that doxycycline should be avoided. Maternal-fetal transmission of *B. burgdorferi* does occur, but in contrast to syphilis in pregnancy, there is no evidence that the organism causes a congenital syndrome. Pregnant patients should be reassured that with recommended therapy for Lyme disease, *B. burgdorferi* infection in the mother should not cause harm to the fetus.

**Expected Outcomes**

Most patients treated for Lyme disease with recommended courses of antibiotics experience resolution of all signs and symptoms of the disorder. About 15% of patients treated for Lyme disease may experience a Jarisch-Herxheimer reaction, a self-limited worsening of symptoms within 24 to 48 hours of initiation of antibiotic therapy. Within the first week of treatment, patients may rarely show...
evolution of disease such as the development of new EM lesions or facial nerve palsy, but these signs should improve as therapy progresses. Most patients with Lyme arthritis experience resolution of joint inflammation after a 1-month course of oral antibiotics, and less than 10% of patients progress to antibiotic-refractory arthritis, which nevertheless resolves within 4 years.13

Subjective complaints of fatigue and musculoskeletal pain may persist for months after treatment for Lyme disease.127 When patients complain of persistent pain and fatigue, evaluation for co-infection with B. microti or A. phagocytophylum should be performed.128 Patients with coinfection tend to be more symptomatic at presentation and can have a delayed resolution of symptoms compared with patients with Lyme disease alone. Objective, nonprogressive signs such as mild facial weakness after facial palsy are likely due to irreversible tissue damage, and further antibiotic therapy does not seem to be beneficial.129-132

**Chronic Lyme Disease and Post–Lyme Disease Syndrome**

There is a great deal of controversy over the potential for Lyme disease to cause life-altering chronic morbidity in patients. It is unusual to have objective signs after recommended antibiotic regimens for Lyme disease diagnosed according to the guidelines discussed earlier, and when such signs (e.g., Lyme arthritis) are present, further antibiotic therapy does not alter outcome.65 Even when objective signs are present, they are usually nonprogressive (e.g., residual facial weakness after facial palsy) or resolve over time (as is the case for Lyme arthritis). Use of the term chronic Lyme disease, which implies ongoing infection, for nonprogressive signs and symptoms after Lyme disease is not valid.

A few patients treated for Lyme disease may have fatigue, musculoskeletal pain, and complaints of memory impairment despite conventional or prolonged courses of antibiotic therapy, a condition referred to as post–Lyme disease syndrome.129,133 In several controlled, population-based cohort studies that used validated standardized measures of outcomes (e.g., SF-36), patients with Lyme disease had more joint pain, symptoms of memory impairment, and worse functional status because of pain compared with controls.126,134,135 These complaints could not be documented by abnormalities on physical examination or by neurocognitive testing, however, and a follow-up study showed that quality-of-life measures improved with time.136 Psychological factors and the presence of psychiatric comorbidity in patients with Lyme disease correlate with poor functional outcomes after treatment.137 More recently, a European study demonstrated that the frequency of nonspecific symptoms at 6 and 12 months after treatment for early Lyme disease was no greater than that of the control group, which included family members without a history of Lyme disease.138 Children are less likely than adults to have persistent complaints after treatment for Lyme disease.139

Two randomized, double-blind, placebo-controlled trials of antibiotic therapy were conducted on seropositive and seronegative patients with chronic symptoms (>6 months) after treatment for Lyme disease.130 Patients were randomly assigned to receive either intravenous ceftriaxone for 1 month followed by 2 months of oral doxycycline or matched intravenous and oral placebos. An interim analysis of the first 129 subjects enrolled (78 seropositive, 51 seronegative) resulted in termination of the study because no differences in outcome between groups receiving antibiotics or placebo were found, and evidence of ongoing infection could not be documented. Another trial of antibiotics for post-treatment Lyme disease symptoms found that fatigue, as assessed by the Fatigue Severity Scale–11, improved in the group receiving intravenous ceftriaxone, but cognitive dysfunction did not.131 Individuals who had positive IgG immunoblots for Lyme disease and who had not received prior treatment with intravenous antibiotics were more likely to have improvement in fatigue. Subsequently, a randomized, placebo-controlled trial of a 10-week course of intravenous ceftriaxone for memory impairment after treatment for Lyme disease did not show sustained improvement in cognitive function.132 An open pilot study provided evidence that gabapentin may be effective in the treatment of chronic pain syndromes after Lyme disease.133

A growing number of patients with similar subjective complaints are being treated for months or years with antibiotics for presumed B. burgdorferi infection.140-144 These patients are diagnosed with chronic Lyme disease even though they often have no serologic evidence of B. burgdorferi exposure or are considered as testing positive on the basis of IgM reactivity on immunoblot despite years of not feeling well. They typically experience only partial resolution of their symptoms with antibiotics. Occasionally, patients may have other conditions such as rheumatoid arthritis or fibromyalgia, for which therapy has been delayed because of a misdiagnosis of Lyme disease.145 Musculoskeletal pain is common in the general population; 20% to 30% of adults complain of chronic fatigue.140 In the absence of a clinical history with objective manifestations of Lyme disease or positive two-tiered serologic tests, definitive attribution of symptoms to B. burgdorferi infection cannot be made. Caution should be exercised when attributing the response of symptoms to the antimicrobial effects of antibiotics. Ceftriaxone and other β-lactam antibiotics can modulate neurotransmitter activity,146 and tetracyclines inhibit matrix metalloproteinases.147 Prolonged antibiotic use is not without risk. Minor side effects are common, and serious adverse events such as biliary complications from ceftriaxone therapy or indwelling catheter–related infections occur at high enough rates to warrant only judicious use of antibiotics.110,132,142

**PREVENTION**

The most effective way to prevent Lyme disease is to reduce exposure risk to B. burgdorferi–infected ticks through personal protective measures and environmental controls.148 These measures include avoidance of tick habitats such as wooded areas, stone fences, woodpiles, and tall grass; wearing protective clothing; and performing daily surveillance and prompt removal of ticks (within 24 hours of feeding). Other effective measures include use of DEET-containing insecticide sprays, yearly application of acaricides to property to kill ticks, construction of four-poster bait stations that apply acaricides onto deer as they feed, and tall fences to prevent deer from incidentally transporting ticks to an area.
A single 200-mg dose of doxycycline (or 4 mg/kg up to 200 mg for children 8 years old and older) has been shown to reduce the incidence of Lyme disease after a recognized tick bite\(^{116}\) but is not routinely recommended because of the low rate of infection.\(^{116}\) An FDA-approved recombinant Osp A–based vaccine to prevent Lyme disease was withdrawn because of low market demand and concern for potential vaccine-related side effects.\(^{151,152}\)

**SUMMARY**

Lyme disease is a localized or systemic infection that usually manifests with skin and musculoskeletal signs and symptoms, but it can involve other organ systems, especially the heart and nervous system. The diagnosis should be based on objective clinical findings consistent with Lyme disease and supporting serologic tests. Most patients are cured with 2 to 4 weeks of antibiotic therapy, although the time to disease resolution may be prolonged, especially for individuals in whom therapy was delayed; irreversible tissue damage may occur. A poor response to antibiotic therapy should raise concern for alternative diagnoses or co-infection with other tick-borne pathogens. Arthritis becomes refractory to antibiotics in less than 10% of patients with Lyme disease. Treatment with nonsteroidal anti-inflammatory drugs and hydroxychloroquine usually resolves arthritis within 4 to 5 weeks. Some patients treated for Lyme disease develop a post-Lyme disease syndrome of fatigue, headaches, mild memory impairment, and musculoskeletal pain. Ongoing infection cannot be shown, and controlled treatment trials show no benefit of prolonged antibiotic therapy over placebo. Referral to an academic medical center with experience in the diagnosis and treatment of Lyme disease should be considered when patients do not respond as expected to therapy.

**Selected References**


Full references for this chapter can be found on www.expertconsult.com.
References


Recognition of tuberculosis (TB) and other mycobacterial infections of the musculoskeletal system has become a major challenge for rheumatologists in the United States and in other developed countries. Before 1999, most rheumatologists could easily go an entire year without seeing a single case of mycobacterial infection. Such cases were rare even at academic centers, where they would likely be presented as unusual teaching cases at clinical conferences.

However, 1999 introduced the routine clinical use of anti–tumor necrosis factor (TNF) therapy in the United States, and with it came an unexpected increase in TB cases seen by rheumatologists. Fortunately, the initiation of routine screening with tuberculin skin tests (TSTs) and other methods has led to a sharp decline in the number of new cases, particularly those associated with reactivation of latent infection. However, the risk of primary infection with new exposure mandates continued vigilance in these patients and potentially in patients on other biologic therapies. Moreover, for non-TB mycobacterial infections, no screening tests are available, making vigilance on immunosuppressive therapy even more important.

Another major force driving the increase in mycobacterial infection is the human immunodeficiency virus (HIV) epidemic that continues to be a worldwide problem. By 1991, 21% of extrapulmonary TB cases in the United States were associated with acquired immunodeficiency syndrome (AIDS). In developing countries, the HIV pandemic has led to marked increases in osteoarticular TB co-infection. These cases are distinguished by disseminated multifocal disease suggestive of hematogenous spread, rapid progression, and more common coexistence with pulmonary infection. HIV testing should be a routine part of the workup of any patient presenting with a musculoskeletal infection secondary to Mycobacterium tuberculosis. Since the introduction of effective antiretroviral therapy, the incidence of TB and nontuberculous mycobacterial infection in patients co-infected with HIV has sharply declined in the United States.

Although TB is uncommon among the non–HIV-infected population in the developed world (in the United States, TB is steadily declining among the general population), it remains a major problem in developing countries. There, TB continues to ravage the population, with 9 million new cases of active disease and 1.6 million deaths each year. Worldwide, it is the number two infectious disease killer after HIV/AIDS. Someone in the world is newly infected with TB bacilli every second. About one-third of the world’s population is infected with TB, providing a reservoir that will continue to complicate its global control. More alarming is the emergence of extensively drug-resistant TB (XDR-TB). These strains fail to respond to all first-line and most second-line TB agents. An outbreak of XDR-TB in South Africa was associated with a mortality rate of 100%, with a median survival of only 16 days after diagnosis. Globalization of the world economy is encouraging increased contact between populations of the developed and developing worlds. Recent immigrants to the United States from endemic areas constitute an expanding reservoir of patients with latent TB.

The challenge of diagnosing musculoskeletal mycobacterial infection reaches beyond the rarity of the disease. Such infections are often indolent and may lack the pain, fevers, chills, and other prominent symptoms that typically accompany bacterial infections of the musculoskeletal system. In addition, unless the diagnosis of TB is a consideration from the outset, routine culture techniques will not isolate the organism. Delay in the diagnosis of mycobacterial infection of the musculoskeletal system is common and will continue without a heightened index of suspicion for these types of infections.

For all these reasons, 21st century rheumatologists need to be well informed about a group of diseases with which they may have had little or no experience during their formal medical training.
CLINICAL SCENARIOS

To appreciate the entire spectrum of clinical problems that a rheumatologist may encounter in dealing with mycobacterial infection, it is useful to think in terms of several distinct categories. Franco-Paredes and colleagues provided a clinically useful division of mycobacterial infections into four major categories. The following sections reflect that division.

Direct Involvement of the Musculoskeletal System

Musculoskeletal infections caused by mycobacteria typically present as chronic, indolent, localized involvement of the bones, spine, peripheral joints, or soft tissues that produces a focus of nonspecific pain and, less often, swelling. In infections initiated by direct tissue inoculation, the traumatic event is often trivial or remote in time from the onset of clinical disease. Diagnosis may be delayed for months to years, in part because of minimal early symptoms and attribution of those symptoms to a noninfectious disorder until disease progression and disability prompt a more aggressive diagnostic investigation.

Constitutional symptoms are typically subtle or absent, and laboratory indicators of inflammation are often normal. Synovial effusion is frequently minimal, and the fluid, if it is attainable, shows nonspecific inflammation. Radiographic abnormalities may be delayed, although newer imaging techniques have allowed the earlier detection of abnormalities and the distinction of TB from other infections and neoplasms. Characteristic pulmonary or extrapulmonary findings are not always present; for example, less than 50% of osteoarticular TB presents with evidence of active or past pulmonary disease. Tuberculin skin tests (TSTs) and interferon-γ-release assay (IGRA) testing may provide useful clues to the cause, but results are not invariably positive, especially in debilitated or immunosuppressed patients. Correct diagnosis is highly dependent, in most cases, on demonstration of the infectious agent by microscopic examination and culture of affected tissue.

In HIV-infected patients, mycobacterial infections are often diagnosed before patients’ HIV-positive status is known, sometimes leading to its recognition. Atypical pulmonary TB and extrapulmonary (often multifocal) infection are common; extrapulmonary infection occurs in 60% to 70% of such cases, compared with 16% of all TB patients.

The clinical patterns of musculoskeletal TB include spondylitis, osteomyelitis, peripheral joint infection, and soft tissue abscesses. In a series of 230 consecutive cases of TB from the preantibiotic era, 5.2% had skeletal involvement; the spine was affected in 60% of these cases. TB of the bones and joints is spread hematogenously. The sites most commonly affected are the spine and hips, followed by the knees and wrists; other joint involvement is rare. Constitutional symptoms are unusual in musculoskeletal TB and, when present, suggest TB in other organs. Vertebral collapse due to spinal TB may initially be attributed to the more common osteoporosis-caused spinal compression fracture. TB only rarely involves skeletal muscle but must be considered in the differential diagnosis of an enlarging muscle lesion. Isolated cases involving tendons, trochanteric bursae, and fasciae latae illustrate the variety of possibilities. Biopsy and culture are required for diagnosis. Imaging studies do not reliably distinguish TB from neoplasm.

In nonendemic areas, skeletal TB usually occurs in elderly, debilitated patients, most often in the form of solitary osteolytic lesions in the axial skeleton. The development of skeletal disease is often remote from the initial infection, which strongly implies reactivation of previous subclinical disease. Patients may have false-negative TSTs for several reasons, including long-term corticosteroid use or coexisting debilitating disease, such as rheumatoid arthritis or chronic renal failure, which compromises resistance and produces anergy (Table 111-1). Spinal disease in children in nonendemic regions has largely been eliminated by effective medical therapy of pulmonary infection. In contrast, in endemic areas with high infectivity rates, those infected are more commonly children and young to middle-aged adults. These individuals have a higher incidence of multifocal skeletal involvement in the ribs, pelvis, vertebral appendages, cervical spine, feet, and long bone diaphyses, and they show positive TST reactivity. Bone seeding occurs through hematogenous spread, sometimes secondarily from another extrapulmonary site. When pulmonary findings are present, a miliary pattern is typical. Spread to bone may also occur from infected nodes, by direct extension or through draining lymph channels.

Spondylitis

The spine is the dominant site of involvement in skeletal TB, accounting for 50% to 60% of cases. Between 48% and 67% of lesions occur in the lower thoracic and thoracolumbar spine in HIV-negative patients, whereas the lumbar spine is most commonly involved in HIV-positive patients. Infection usually begins in the anterior subchondral bone of a single vertebra adjacent to the intervertebral disk (Figures 111-1 and 111-2). Progression to bone takes 2 to 5 months and begins with extension from cancellous to cortical bone, and then across the disk space to adjacent vertebrae (Figure 111-3). Bone destruction may lead to vertebral collapse, which typically occurs anteriorly, resulting in gibbus deformity. Isolated neural arch involvement and intraspinal abscess may also occur.

Paravertebral abscess begins with extension of infection under the anterior longitudinal ligament. In the thoracic spine, it may extend into the pleural space and the lung parenchyma. In the cervical region, it may present in the posterior cervical triangle or the retropharyngeal space. In the lumbar spine, a cold abscess characteristically produces lateral displacement of the psoas muscle and may dissect.

<table>
<thead>
<tr>
<th>Table 111-1</th>
<th>Causes of False-Negative Purified Protein Derivative Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased age (&gt;70 yr)</td>
<td></td>
</tr>
<tr>
<td>Steroid use (prednisone ≥15 mg/day)</td>
<td></td>
</tr>
<tr>
<td>Hypoalbuminemia (&lt;2 g/dL)</td>
<td></td>
</tr>
<tr>
<td>Azotemia</td>
<td></td>
</tr>
<tr>
<td>Impaired cellular immunity</td>
<td></td>
</tr>
<tr>
<td>Human immunodeficiency virus infection</td>
<td></td>
</tr>
</tbody>
</table>
Myobacterial infections of bones and joints

Symptoms. Paraparesis and paraplegia have been reported in 1% to 27% of patients in various series. In comparison with pyogenic vertebral osteomyelitis, spinal TB more often presents with a prolonged clinical course, thoracic segment involvement, absence of fever, spinal deformity, neurologic deficit, and paravertebral or epidural masses. On occasion, tuberculous spondylitis may present with chronic inflammatory-type back pain more typical of a spondyloarthopathy.

Mycobacterial colony counts in bone biopsy specimens are relatively low. Only 40% of smears and cultures from psoas abscesses are positive. Among patients meeting strict clinical and radiographic criteria in one series, between 73% and 82% had compatible histologic features on biopsy; of these, 80% to 95% had positive cultures. The differential diagnosis, which is extensive, includes pyogenic and fungal osteomyelitis, primary and metastatic tumors, sarcoidosis, multiple myeloma, and eosinophilic granuloma.

Cervical spine involvement is relatively rare, accounting for only 0.4% to 1.2% of cases of extrapulmonary TB in the United States. The most common presenting symptoms are neck pain and stiffness, although hoarseness, dysphagia, torticollis, fever, anorexia, and neurologic disorders may also occur. Spinal involvement can progress to myelopathy because of delays in diagnosis. Radiographs may show characteristic osteolysis of the anterior vertebral body with sparing of the posterior portion, gibbus deformity, disk involvement, and a partially calcified paraspinous mass. Computed tomography (CT) or magnetic resonance imaging (MRI) is useful for assessing compromise of the spinal canal. Retropharyngeal infection may extend into the craniocervical junction and, if not promptly recognized, may cause atlantoaxial dislocation and neurologic complications.

The sacroiliac joint is involved in up to 10% of cases of skeletal TB, often without other evidence of disease. Infection, TB in particular, should be considered in all cases of unilateral sacroiliitis, particularly when other features of a spondyloarthropathy are absent. Emigration from an endemic area and a past history of TB increase the likeli-

Figure 111-1 Tuberculous spondylitis: sites of involvement. Tuberculous lesions can localize in the vertebral body (1) or, more rarely, in posterior osseous or ligamentous structures (2). Extension to the intervertebral disk (3) or prevertebral tissues (4) is not infrequent; subligamentous spread (5) can lead to erosion of the anterior vertebral surface. (From Resnick D: Diagnosis of bone and joint disorders, ed 3, Philadelphia, 1995, WB Saunders, p 2464.)

Figure 111-2 Tuberculous spondylitis with discovertebral lesion. A, The initial radiograph reveals subchondral destruction of two vertebral bodies, with mild surrounding eburnation and loss of intervertebral disk height. The appearance is identical to that in pyogenic spondylitis. B, Several months later, an osseous response is evident. Note the increased sclerosis, osteophytosis and improved definition of the osseous margins can be seen. (From Resnick D: Diagnosis of bone and joint disorders, ed 3, Philadelphia, 1995, WB Saunders, p 2465.)
Osteomyelitis may develop in a bone or a joint that has been previously exposed to trauma. Although any bone can be involved, the femur and tibia are most commonly affected. Dactylitis may also occur in children. In one large series from an endemic area, the tibia are most commonly affected. Dactylitis may also occur in children. In one large series from an endemic area, the tibia are most commonly affected. Dactylitis may also occur in adults.

Tuberculous Osteomyelitis

Tuberculous osteomyelitis occurs in both children and adults. Although any bone can be involved, the femur and the tibia are most commonly affected. Dactylitis may also occur in children. In one large series from an endemic area, such cases represented 19% of bone and joint TB and 15% of cases of osteomyelitis of hematogenous origin. Bone pain was the most common presentation; a draining sinus, abscess formation, and local swelling and tenderness were also common. The average delay before diagnosis was 28 months.

Multifocal osteoarticular TB is a less common variant of the disease, but TB should be considered in all patients from endemic areas who present with multiple destructive skeletal lesions.

For a definitive diagnosis of osteoarticular TB, a biopsy specimen of an affected site must be obtained. Soft tissue lesions characteristically demonstrate rim enhancement on CT examination. CT may also facilitate percutaneous needle biopsy or abscess drainage. Histologic examination generally reveals granulomatous inflammation. In one series of 121 cases, biopsy showed a positive culture in 33%, granulomatous histologic features in 46%, and both in 21%. Radiographic findings include cavity formation with a thin adjacent layer of sclerosis in about 50% of cases, sometimes containing a sequestrum. The true extent of bone involvement may be difficult to detect because of clinically silent lesions. Bone imaging with technetium 99, although more sensitive than conventional radiographs, may provide false-negative information in early, indolent, or highly destructive disease. Tuberculin test reactions are positive in more than 80% of cases.

Treatment with chemotherapy is generally effective. In a minority of cases, surgical débridement is required for healing. Initiation of therapy on the basis of histologic findings is appropriate pending culture results. Sinus cultures are commonly positive for pyogenic bacteria both before and after antituberculous therapy, but these are presumed to be contaminants. Healing is associated with sclerosis at the margin of lesions. Misdiagnosis of the condition as pyogenic osteomyelitis may lead to unnecessary surgery or to delayed antituberculous treatment, resulting in extension of infection into the joint and chronic disability.

Septic Arthritis

Tuberculous joint involvement is second in frequency to vertebral infection. The typical pattern is a monoarticular arthritis involving the large joints, most commonly the hip and knee (Figures 111-4 and 111-5). Other joints less commonly involved include the sacroiliac, shoulder, elbow, ankle, carpal, and tarsal joints. Infection begins in the synovium, with slower progression of destructive changes than in pyogenic septic arthritis. Prosthetic joint infection with M. tuberculosis has been reported and usually is caused by local reactivation of latent disease.

The diagnosis of tuberculous arthritis is often missed. A consecutive series spanning the years 1970 to 1984 emphasized typical features. Of 23 cases of musculoskeletal TB, 9 involved the spine, 1 the hip, and the remaining 13 the peripheral joints. Most patients were men older than 50 years. The history of TB or exposure was generally forgotten. In all cases, presenting symptoms included joint pain and swelling. Four patients had evidence of active pulmonary TB, and in two patients with sterile pyuria, M. tuberculosis grew from the urine. Only 5 of 10 patients tested had a positive TST. Radiographs showed changes of erosive bone lesions begin with hematogenous implantation of organisms in the medullary area. Metaphyseal involvement is most common, and lesions may spread through the growth plate to involve the adjacent joint, usually late in the disease course. Lesions are typically destructive. Lytic lesions in unusual areas, such as the pubic symphysis, sacroiliac joint, and elbow, can be misdiagnosed as malignancy. Osteomyelitis may develop in a bone or a joint that has been previously exposed to trauma.
arthritis in 7 and no changes in 4 of 11 joints studied. The median delay in diagnosis was 8 months.

Arriving at a correct diagnosis requires vigorous pursuit and usually a synovial biopsy and culture. Initial studies are often misleading and may contribute to delayed diagnosis or misdiagnosis. Synovial fluid findings are variable and do not distinguish this arthropathy from other inflammatory or septic arthritides. Cell counts more often suggest bland inflammatory, rather than septic, arthritis and may contain a preponderance of neutrophils. The diagnosis may be facilitated if the organism is observed on an acid-fast smear of synovial fluid, but only 10% to 20% of reported cases are positive. In contrast, synovial fluid cultures are frequently positive. Radiographic changes are similar to those seen in other septic arthritides, beginning with juxta-articular bone demineralization and progressing to marginal bone erosion and articular cartilage destruction (see Figure 111-4).

With the open biopsy technique, granulomatous histologic features and positive cultures are present in more than 90% of cases. No data are available for direct amplification tests in mycobacterial arthritis, but their use can be considered in suspected cases to arrive at an earlier diagnosis. The histologic examination alone may be confusing, because granulomatous synovitis can also be found in nontuberculous mycobacterial infection, sarcoidosis, erythema nodosum, brucellosis, Crohn’s disease, and foreign body reaction. As noted, synovial acid-fast smear is of limited value. Tuberculous arthritis is also reported in children, sometimes early in the disease course; synovial biopsy and culture are recommended in children with monoarthritis and a positive tuberculin reaction. TB must be considered among the possible causes of septic arthritis in patients with pre-existing rheumatoid arthritis, although it is not widely seen in developed countries. Conversely, rheumatoid factor may be present in TB, leading to diagnostic confusion in the presence of chronic monoarthritis.

Emergence of Tuberculosis during Treatment of Rheumatic Disease

Many patients with systemic rheumatic disease have dysregulated immune systems that are treated with immunosuppressive drugs. Such an impaired immune response may permit the reactivation of latent TB. TNF plays a key role in granuloma formation and stabilization, which promotes the containment of M. tuberculosis. The rate of TB among patients with rheumatoid arthritis before the widespread use of anti-TNF drugs was approximately 6 cases per 100,000, although some studies have suggested that the rate in rheumatoid arthritis is higher than in the general population. In one large patient registry, the estimated incidence of TB associated with infliximab in rheumatoid arthritis patients was in excess of 1000 per 100,000 person-years of exposure during the years 2000-2001. In the United States, the rate was 52 per 100,000 person-years during the same time period. Etanercept may be associated with a lower risk for TB reactivation than infliximab: 10 cases per 100,000 person-years of exposure with etanercept, versus 41 per 100,000 person-years with infliximab, according to an analysis of Food and Drug Administration data. More recently, both British and French registries of biologic use have identified a greater risk for TB associated with the use of adalimumab and infliximab than with etanercept. Fewer data are available for the newer TNF antagonists, and comparisons are complicated by the widespread use of pre-treatment screening instituted after the first agents in this class were approved for use. Monoclonal antibodies to TNF may play a greater role in destabilizing granulomata than...
TNF receptors. Nonetheless, similar precautions should be taken with any TNF antagonist, and all should be presumed to increase risk of new infection and reactivation of latent infection.

Other biologic agents may pose a lower risk for reactivation of TB because the mechanisms of action are not intimately involved with host defenses related to intracellular organisms. Animal studies have suggested that abatacept does not impair the ability of mice to respond to M. tuberculosis, but this has not been confirmed in humans. Patients treated with abatacept and tocilizumab in clinical trials were pre-screened with a TST and excluded if positive, thus making a direct comparison difficult. Little or no evidence indicates that B cells play a major role in containing TB, and no recommendation for TB screening is included in the labeling for rituximab.

Glucocorticoid use poses a significant hazard for patients with latent TB. Many mechanisms account for this effect, such as impairment of cellular immune responses and monocyte chemotaxis and function, including the monocyte's production of TNF. The use of glucocorticoids has been associated with a five times increased risk for the development of TB in a case-control study based on a large general practice database in the United Kingdom. The risk appeared to be dose related but was seen even at the physiologic dose of prednisone 7.5 mg/day. Just as with anti-TNF therapy, the risk for TB was greatest early in the course of treatment. In a more recent study from France, prednisone given at 10 mg or more per day, but not at lower doses, was associated with an increased risk for developing TB.

Reactivation of TB in the face of anti-TNF therapy typically occurs within 6 months of treatment and often presents as extrapulmonary disease. Institution of a TB screening protocol in rheumatoid arthritis patients and treatment of latent TB before TNF antagonist administration resulted in a 78% decrease in active TB. Strategies to treat latent TB infection that are tailored to the at-risk population can effectively and safely lessen the likelihood of active TB in patients treated with TNF antagonists. Indeed, it has been suggested that much of the blame for the development of active TB in this situation can be attributed to incomplete compliance with screening protocols. Conversely, in a report of 84 patients at a single center treated with etanercept after demonstrating a positive purified protein derivative (PPD), no cases of active TB were seen during a mean of 2 years of follow-up, despite the fact that only 78 patients received prophylaxis, and 26 of those failed to complete the prescribed course. Although screening for latent TB infection has reduced the incidence of active disease, false-negative TSTs can undermine such good intentions (see Table 111-1).

The approach to a patient with latent TB who needs anti-TNF therapy has not been standardized. Many authorities recommend that patients with positive TSTs and normal chest radiographs who have never been treated for TB receive at least 1 to 2 months of therapy before beginning an anti-TNF drug, if for no other reason than to be certain that they will be able to tolerate the full course of latent TB therapy. Appropriate regimens include 9 months of isoniazid or 4 months of rifampin. Patients discovered to have active TB should receive a complete course of standard antituberculous therapy before any consideration is given to using an anti-TNF drug.

Although anti-TNF therapy and steroids stand out as particular risk factors for the development of TB, all rheumatic disease patients treated with immunosuppressive therapy that impairs cellular immunity should be considered at risk. This is especially true for the elderly, the malnourished, and immigrants from countries with high endemic rates of TB.

### Rheumatic Disorders Precipitated by Treatment of Tuberculosis

A variety of rheumatic conditions may be precipitated by drugs used in the treatment of TB. These include drug-induced lupus caused by isoniazid (INH) and rifampin (RIF). As with other cases of drug-induced lupus, they are associated with positive antinuclear antibodies and the presence of antihistone antibodies. Typically, these patients follow a benign course, with reversal of disease after the drug is discontinued.

Arthropy and tendinopathy have been described with the use of fluoroquinolones, especially ciprofloxacin and levofloxacin. Risk of tendon rupture (usually the Achilles tendon) is greatest in patients older than 50 years and increases with concomitant use of corticosteroids.

Pyrazinamide interferes with the renal tubular excretion of uric acid and has been associated with the development of hyperuricemia and gout in adults.

Some patients develop a paradoxical worsening of their condition upon initiation of antituberculous therapy. Such a development may raise questions about a flare of the underlying disease, especially if immunosuppressive therapy has been withdrawn in the face of infection. Symptoms include fever, malaise, weight loss, and increasing respiratory symptoms. The mechanism for such reactions is not completely understood but has been categorized within the spectrum of immune reconstitution inflammatory syndromes. Such reactions are more common in HIV patients and have also been seen in patients treated with infliximab following cessation of anti-TNF therapy. Corticosteroids may ameliorate this response.

### Reactive Immunologic Phenomenon in the Setting of Tuberculosis

A variety of reactive immunologic phenomena have been associated with M. tuberculosis infection. These are uncommonly found in clinical practice.

Poncet's disease is an aseptic inflammatory polyarthritis that occurs in the presence of active TB. Although any joints can be involved, the most commonly affected are the knees, ankles, and elbows. The mechanism is thought to be similar to other forms of reactive arthritis secondary to remote infection. Most cases resolve after satisfactory treatment of the TB. A reactive arthritis has been described following intravesicular instillation of bacille Calmette-Guérin (BCG) vaccine for bladder cancer.

Other seldom encountered immune reaction patterns described with M. tuberculosis include erythema nodosum, erythema induratum, and amyloidosis (AA type).
Tuberculin Skin Test

The TST—also called the purified protein derivative (PPD)—has been in use for nearly a century and remains the most widely used screening test for TB. It is routinely administered in rheumatology offices that prescribe anti-TNF therapy. The test represents a crude mix of antigens from M. tuberculosis and is plagued by both false-positive and false-negative results. It is unable to distinguish latent infection from active disease and may be negative in the face of severe active TB. Corticosteroids (≥15 mg/day prednisone) may render the PPD negative in the face of latent TB. Elderly and malnourished patients may not exhibit a positive TST. Table 111-1 provides a partial list of causes of a false-negative TST that are of special interest to rheumatologists.

False-positive results may likewise occur in the case of infection with nontuberculous mycobacteria or previous BCG vaccine. PPD positivity degrades over time in BCG-vaccinated patients at a variable rate. The degree of positivity is influenced by a number of factors, including the number of BCG vaccinations and the number of subsequent PPD tests performed. Some patients may retain a response as long as 15 years after BCG vaccine. However, a PPD result of 20 mm or greater is rarely due to BCG. In addition, if faced with a high-risk situation, such as initiation of anti-TNF therapy, one probably should make a presumption of latent TB even if the diameter of induration measures as little as 5 mm.

The important message for rheumatologists is that the TST is an important but imperfect screening tool for latent TB infection, with sensitivity and specificity in the range of 70%. A negative TST should not eliminate the clinician’s vigilance in monitoring patients who are being treated with anti-TNF therapy for reactivated or new-onset TB, especially in high-risk populations.

Imaging

Although imaging patterns suggestive of TB have been discussed, no pathognomonic skeletal radiographic features can establish the diagnosis. Early features on radiographs may be equivocal or nonexistent. Chest radiographs often are normal or fail to show features characteristic of TB.

Conventional radiography is generally a useful approach for defining bone destruction, extent of disease, and adjacent soft tissue lesions. MRI is more effective in identifying early disease; it may help distinguish TB from other infections and neoplasms and can aid in estimating the extent of disease. Scintiscans with technetium and gallium may also be helpful in localizing bone and soft tissue lesions, but early false-negative findings are not uncommon. CT can be helpful in guiding diagnostic needle biopsy. Fine-needle biopsy is an acceptable alternative to core-needle biopsy and open biopsy for the diagnosis of osteoarticular TB in both the axial and the peripheral skeleton, and it has the advantage of obviating general anesthesia. Both CT and MRI may be helpful in monitoring therapy.

Culture

Nearly all species of Mycobacteria are slow growing, with M. tuberculosis being the slowest. Other bacteria may rapidly outgrow mycobacteria if the specimen is not inoculated on special isolation media. The small number of mycobacteria found in clinically infected areas further challenges the ability to confirm mycobacterial infection and accounts for the low yield of positive Ziehl-Neelsen staining (10% to 20%) in synovial fluid and other biologic fluids. Synovial fluid and other bodily fluids are less likely than tissue to yield a positive culture. If a joint is suspected of harboring TB, a synovial biopsy should be obtained. Arthroscopically derived tissue yields higher positive cultures than needle biopsies do. CT-guided needle aspiration and biopsy can provide invaluable information in the case of spinal involvement. Characteristic features on tissue pathology, including caseating and noncaseating granulomas, may allow an early presumptive diagnosis of TB pending culture results, which may take 4 to 6 weeks to be finalized.

Advanced Diagnostic Testing

Interferon-γ Release Assays

The limitations of the traditional PPD skin test in terms of specificity and sensitivity in identifying latent TB have led to the development of new T cell–based testing. These assays measure the production of interferon-γ by whole blood mononuclear cells stimulated by specific M. tuberculosis antigens. The interferon-γ release assays (IGRAs) have become a useful test for latent TB and have good sensitivity and specificity for latent TB infection. These assays may prove particularly helpful in distinguishing TB from nontuberculous mycobacteria and in assessing patients who have had recent BCG vaccines. They also offer the opportunity to circumvent operator error with TST administration and do not require patients to return for a reading. IGRAs have replaced the TST in some centers for routine screening for latent TB infection.

The role of IGRAs in the diagnosis of latent TB infection in patients with immune-mediated inflammatory disorders before initiation of anti-TNF therapy remains unsettled. The weight of evidence suggests that IGRAs are more sensitive and more specific than TSTs in this patient population. Indeterminate results from IGRAs in this setting have ranged from 1.2% to 28.6%, raising concerns about the performance of the assay from center to center. Overall, agreement of IGRAs with the TST is poor owing to a lower proportion of positive TSTs. Positive IGRAs are more closely related to TB risk factors than are positive TSTs. Discordant IGRA-negative/TST-positive results are associated with prior BCG vaccination, and discordant IGRA-positive/TST-negative results are associated with corticosteroid therapy, suggesting that the former represents a false-positive TST and the latter represents a false-negative TST. Nevertheless, the meaning of a discordant result is uncertain enough that patients with a positive test result with IGRA or TST should be considered for therapy of possible latent TB infection. Additional experience will
be required to provide confidence that a negative IGRA in
the setting of a positive TST reliably excludes latent TB
infection. If only one testing modality is utilized, one recent
analysis found that use of an IGRA was more cost effective
than use of the TST for patients being screened before
receiving anti-TNF therapy.77

IGRAs have also been studied in patients receiving anti-
TNF therapy to identify previously undiagnosed (“hidden”) latent TB infection or newly acquired TB infection.
However, currently no guidelines are available for monitor-
ing for tuberculosis infection in patients on anti-TNF
therapy. Rates of TST conversion in patients on anti-TNF
therapy have been reported to be 33% and 37% from centers
in Korea and Taiwan; corresponding rates of IGRA conver-
sion were 14% and 11%, respectively, and appeared to
better correlate with the risk of active tuberculosis.78,79
Pending formal guidelines, in regions where ongoing expo-
sure to infectious tuberculosis may occur, routine monitor-
ing of patients with TB on therapy—perhaps annually—is
probably wise.

**Nucleic Acid Amplification**

Molecular diagnostics using nucleic acid amplification may
be helpful in patients with low mycobacterial loads and may
provide more rapid diagnosis. The presence of polymerase
chain reaction (PCR) inhibitors, especially in extrapulmo-

nary specimens, can lead to false-negative results. Despite
more than a decade of experience, the role of nucleic acid
amplification tests in diagnosing TB infection is still being
defined. These assays must be interpreted with caution in
extrapulmonary tissue specimens, and when the clinical
suspicion of infection is low.80-82 In a small cohort of patients
with vertebral osteomyelitis, multiplex PCR testing of bone
biopsy specimens demonstrated 90% sensitivity and 100%
specificity for tuberculous infection, illustrating the poten-
tial for this technology to provide rapid and accurate
diagnosis.83

**TREATMENT**

Appropriate management of tuberculous infection of bones
and joints is a complex and evolving process. Proper selec-
tion of antibiotic regimens and ongoing disease monitoring
should involve co-management with an infectious disease
consultant. Nontheless, rheumatologists should be familiar
with the basic treatment principles.

Treatment of M. tuberculosis infections that involve the
musculoskeletal system consists of the same combination
chemotherapy regimens that are effective in pulmonary TB.
Treatment regimens for infections of the bone and spine are
longer than those recommended for lung disease and other
extrapulmonary sites. The longer courses for musculoskel-
etal disease have been based on poor tissue penetration into
osseous tissues and higher rates of relapse.

Current guidelines on the treatment of TB published
jointly by the Centers for Disease Control and Prevention,
the American Thoracic Society, and the Infectious Diseases
Society of America recommend a 6-month course of therapy
for all sites except the bone (6 to 9 months) and the central
nervous system (9 to 12 months).84 Rifampin is the critical
component that allows shorter-course therapy.

The standard approach to TB therapy (both pulmonary
and extrapulmonary) in the United States currently includes
drugs to start: isoniazid, rifampin, ethambutol, and
pyrazinamide—known as IREZ therapy. Once the bacillus
is confirmed to be sensitive to INH, ethambutol can be
discontinued. Pyrazinamide is administered for 2 months,
and rifampin and INH are continued for the duration of
therapy.

Longer courses of therapy are required for patients who
are slow to respond. Radiographic features of mycobacterial
disease may not change much after 6 months of treatment,
and response to treatment is based mainly on clinical fea-
tures, including reduction in pain, resolution of constitu-
tional symptoms, and emergence of increased mobility.
Longer courses are advised for cases of relapse and for resis-
tant organisms.

Surgical intervention is seldom indicated for the initial
management of osteoarticular TB. Possible exceptions at
presentation include patients with advanced or progressive
disease or spinal kyphosis of 40 degrees or greater. Multidrug-
resistant tuberculosis may also provide a relative indication
for surgical débridement. Patients who have extensive joint
destruction and immobility after an adequate course of che-
motherapy may be candidates for surgery.

Following successful antibiotic therapy, arthroplasty
of the hip and knee may be undertaken and is usually suc-
cessful. Recurrence of disease in the prosthetic joint is less
likely if the surgery is performed years after the infection,
and if the tissue obtained at surgery is culture negative. This
may be impractical, however, when a patient is unable to
ambulate following an adequate course of therapy. In such
cases, TB therapy is continued through surgery and for
at least 3 months postoperatively. If TB recurs in the
prosthetic joint, it can sometimes be managed with anti-
biotic therapy alone. However, in many cases, removal
of the prosthesis is necessary for complete resolution of
infection.

The increasing incidence of drug-resistant TB compli-
cates the selection of appropriate drugs.85,86 Primary mono-
resistance to isoniazid occurs in about 7% of TB isolates
in the United States. When identified, this does not sub-
stantially impact the treatment outcome. Multidrug-
resistant TB (MDR-TB) refers to isolates that are resistant
to isoniazid and rifampin. The rate of MDR-TB in the
United States has been relatively stable for the past decade,
at less than 1%. In other parts of the world, MDR-TB rates
may exceed 6% for new cases and 30% in previously treated
individuals. The treatment of MDR-TB is complex and
often requires the use of multiple (often toxic) second-line
agents for 18 to 24 months or longer. XDR-TB—that is,
strains resistant to all first-line TB drugs and to at least three
second-line agents—has been reported in at least 17 coun-
tries, with particularly high rates in Kazakhstan, Iran,
and South Africa.87 Options for treatment of XDR-TB are
very limited, and risks of treatment failure and death are
high.88

Incomplete adherence to treatment is a particularly
important risk factor for secondary drug resistance, and the
likelihood of drug resistance increases following relapse
after treatment. Directly observed therapy is strongly advo-
cated to reduce the spread of infection and the frequency
of drug-resistant TB.89

---

85 86 87 88 89
OSTEOARTICULAR INFECTIONS CAUSED BY NONTUBERCULOUS MYCOBACTERIA

Nontuberculous mycobacteria (NTM) are ubiquitous in the environment, including soil, water, and animal reservoirs. They are not typically spread from human to human, and infections with these organisms have been reported with increasing frequency. Although a vast majority of infections caused by these organisms are pulmonary, skin and soft tissue infections may be seen in normal hosts; children may develop a localized lymphadenitis. Osteoarticular infections are usually caused by direct inoculation of the organism or by spread of contiguous infection. Blunt trauma has been implicated as a risk factor in cases of vertebral osteomyelitis due to NTM. Immunosuppressed hosts may develop nontuberculous mycobacterial infections of the musculoskeletal system, but these infections are much less common than TB. In contrast to TB, nontuberculous infection is more likely to cause tenosynovitis, synovitis, or osteomyelitis and is less likely to cause spinal infection. A review identified only 31 cases of vertebral osteomyelitis due to NTM reported between 1965 and 2003. Although more than 120 species of nontuberculous mycobacteria are known, most musculoskeletal infections are caused by Mycobacterium marinum, Mycobacterium kansasii, and Mycobacterium avium-intracellulare (also called M. avium complex, or MAC), but cases of musculoskeletal infection with Mycobacterium haemophilum, Mycobacterium chelonae, and Mycobacterium xenopi have also been reported.

Three distinct patterns of musculoskeletal involvement have been reported: tenosynovitis, synovitis, and osteomyelitis. Tenosynovitis typically presents as chronic unilateral hand and wrist swelling (Figure 116A-C). Synovitis typically presents as chronic indolent asymmetric swelling in a knee, hand, or wrist. A number of species have been associated with these syndromes, and the number isolated from immunosuppressed patients is growing. Predisposing factors, in addition to immunosuppression and direct inoculation, include environmental exposure and pre-existing joint disease.

Musculoskeletal infections with atypical mycobacteria may be indistinguishable from M. tuberculosis infections, so that correct diagnosis usually requires tissue biopsy and culture. Synovial fluid, when attainable, is typically inflammatory, and a culture may be helpful only if specific mycobacterial cultures are requested. Identification of acid-fast bacilli on smear and granulomatous inflammation from a tissue biopsy specimen often provides direction for an appropriate microbiologic investigation, but histologic features do not consistently demonstrate granuloma formation. With a compatible clinical presentation and histologic findings, mycobacterial culture, including special techniques for M. marinum, should be requested. Direct amplification testing may be useful for more rapid identification of mycobacterial species in tissue specimens, but data from musculoskeletal cases are limited. In addition to mycobacteria, causes of granulomatous synovitis include fungi, brucellosis, sarcoidosis, inflammatory bowel disease, and nonmetallic foreign bodies. Mycobacteria other than M. tuberculosis are responsible for a significant proportion of these cases.

MAC has become the most common mycobacterial infectious agent affecting patients with HIV/AIDS, in whom it has a tendency to cause disseminated disease. Fortunately, a sharp decline in the incidence of MAC has been noted following the introduction of effective antiretroviral therapy and MAC prophylaxis. Isolation of M. tuberculosis is always clinically significant. In contrast, when dealing with a nontuberculous mycobacterial isolate, the clinician must judge whether it is a contaminant, represents insignificant colonization, or is the cause of disease. Certain guidelines have proved useful in this respect:

- The illness should be consistent with one or more syndromes associated with mycobacterial infection.
EMERGENCE OF NONTUBERCULOUS MYCOBACTERIAL INFECTION DURING THE TREATMENT OF RHEUMATIC DISEASE

No published data exist on the risk of nontuberculous mycobacterial infection in the setting of therapy with nonbiologic disease-modifying antirheumatic drugs, including methotrexate. The decision to interrupt or continue therapy in such instances should be made on a case-by-case basis and should involve an infectious disease specialist, with recognition that alternative treatments, such as steroids and biologics, may carry an even greater risk.

Nontuberculous mycobacteria are increasingly recognized as important pathogens in patients receiving anti-TNF therapy. Indeed, these infections now occur twice as frequently as TB in the United States, presumably because of the use of screening tests for the latter.\textsuperscript{107} The complexity of these infections was illustrated by a recent review of 239 cases of nontuberculous mycobacterial infection reported to the MedWatch database, in which only 105 were established to be probable or confirmed according to established disease criteria.\textsuperscript{108} The most commonly reported presentation was seen in elderly women with rheumatoid arthritis (RA), possibly because both nontuberculous mycobacterial infection and RA are common in this group; most patients were also taking concomitant steroids and/or methotrexate. Infections with several different species within this group have been reported, with MAC being the most common. As with TB, the risk appears to be associated with all drugs in the class, rather than with a specific agent, although more infections have been reported with infliximab.\textsuperscript{109} Extrapulmonary disease is a common finding and is unusual for these organisms. A paradoxical response to therapy, similar to that seen with TB, has been reported with treatment of MAC following infliximab.\textsuperscript{109} Although cessation of anti-TNF therapy is appropriate following the development of an atypical mycobacterial infection, the safe reintroduction of etanercept during treatment of M. marinum has been reported.\textsuperscript{110}

Managing the risk of nontuberculous mycobacterial infection in the setting of anti-TNF therapy can be challenging. Although screening has been helpful in reducing the risk of TB infection in this setting, no method of screening for atypical mycobacteria is currently available. Because these infections are more common in individuals with pre-existing pulmonary disease, clinicians may wish to consider more extensive evaluation of RA patients with bronchiectasis about to initiate anti-TNF therapy. Although MAC prophylaxis is recommended for HIV-positive patients with CD4 counts of 50 or lower, no experience with and no guidelines for such prophylaxis with anti-TNF therapy have been put forth. The complex nature of these infections has been further demonstrated by a case of M. marinum mimicking RA, with synovitis and subcutaneous nodules developing during infliximab therapy.\textsuperscript{111}

References


The references for this chapter can also be found on www.expertconsult.com.
Fungal Infections of Bones and Joints

ERIC M. RUDERMAN • JOHN P. FLAHERTY

Fungal Infection is a relatively infrequent but important cause of osteomyelitis and arthritis. Fungal diseases that commonly cause osteomyelitis include coccidioidomycosis, blastomycosis, cryptococcosis, candidiasis, and sporotrichosis (Table 112-1). Fungal arthritis is less common and is most often associated with sporotrichosis, cryptococcosis, coccidioidomycosis, blastomycosis, candidiasis, and, occasionally, other species. The epidemiology of these fungal infections, their musculoskeletal presentations, and their treatment are considered in this chapter.

The epidemiology and clinical features of individual deep mycoses may suggest the diagnosis in some cases, but their indolent presentation, which often resembles that of other noninfectious diseases, may be misleading. Travel and immigration have blurred their geographic localization. Infection may be acute and overwhelming in immunocompromised patients, for whom disseminated fungal infections are a major risk. Anticytokine and other immunosuppressive treatments for rheumatic diseases, particularly those targeting tumor necrosis factor (TNF), are associated with disseminated fungal infection, as are acquired immunodeficiency syndrome (AIDS), pregnancy, and treatments for transplantation and malignancies, in some cases. For rheumatologists, disseminated fungal infections are an important diagnostic consideration in some patients and must be considered before starting biologic treatments in those at risk; they may also complicate the clinical course of other arthritides. The presentation of fungal infections as a consequence of rheumatic disease therapy will also be considered in this chapter.

Fungal infections are generally diagnosed by histologic examination or culture of involved tissues. Improved biopsy techniques may assist the diagnosis, provided the possibility of fungal infection is considered and proper studies are requested. Synovial fluid leukocyte counts and culture results vary among fungal infections and in individual cases and may be misleading. Serologic testing may also assist in diagnosing and staging several fungal infections. Detecting fungal antigens and DNA in blood and tissue is now possible in some cases, but the clinical use of these methods is still under investigation.

COCCIDIOIDOMYCOSIS

The soil fungi Coccidioides immitis and Coccidioides posadasii generally cause a primary respiratory illness after spores are inhaled. A self-limited acute pneumonia may result, associated with systemic manifestations such as arthralgia and erythema nodosum (valley fever), but infection is often unapparent and only infrequently becomes chronic or disseminated. Coccidioidomycosis is endemic to the southwestern United States and areas of Central and South America, but cases are increasingly diagnosed in nonendemic areas because of travel, infection from fomites, and reactivation of remote infection. Cases increase when soil is disturbed and in windy conditions. Direct human-to-human transmission is rare. Extrapulmonary infection is almost always caused by hematogenous spread from an initial pulmonary focus. The bones and joints are frequent sites of dissemination, particularly in immunocompromised hosts.

Septic arthritis of the knee is common, generally arising from direct infection of the synovium. Other joint infections are caused by spread from a contiguous osteomyelitis involving the vertebrae, wrists, hands, ankles, feet, pelvis, and long bones. The onset is characterized by gradually increasing pain and joint stiffness, with little swelling but early radiographic changes. In one series, arthritis was the only manifestation of disseminated coccidioidomycosis in 51 of 57 patients and was an aspect of more generalized disease in the remaining 6 patients.

Diagnostic confusion is common in osteoarticular coccidioidomycosis because of the delayed dissemination (months to years) after primary infection and because of atypical clinical presentations. The criteria for diagnosis
include compatible clinical features, serologic studies, histologic examination, and culture. Early infections, often before systemic spread, are associated with a positive antibody precipitin test that detects immunoglobulin (Ig) M antibody. Complement fixation serologic values detecting IgG antibodies are in a range indicative of disseminated disease in a majority of patients and show a significant decrease with effective treatment. Serologic testing for Coccidioides may have negative results early in the infection and in immunosuppressed patients. A specific enzyme immunoassay (EIA) to Coccidioides galactomannan antigen in urine has shown promise in the diagnosis of severe coccidioidomycosis infection. The definitive diagnosis is most commonly made by the demonstration of granulomatous synovitis and typical spherules in a biopsy specimen, confirmed in some cases by positive culture and direct amplification testing. Synovial fluid, when obtainable, does not necessarily demonstrate septic leukocyte counts and may have a lymphocyte predominance. Synovial fluid is rarely culture positive; culture of synovial tissue may be more helpful. Radioisotope bone scans may be helpful to identify areas of infection.

With early diagnosis of effusive synovitis, antifungal treatment alone is appropriate. Treatment is usually initiated with oral azole antifungal agents, most commonly fluconazole or itraconazole. Amphotericin B is recommended for alternative therapy, especially if lesions are appearing to worsen rapidly and are in particularly critical locations such as in vertebral osteomyelitis. Lipid formulations of amphotericin B have demonstrated less nephrotoxicity and infusion-related side effects than conventional deoxycholate amphotericin B and may be given at doses higher than those tolerated with conventional amphotericin, but they have never been formally studied in clinical trials. With more widely disseminated infection or involvement of critical areas such as the spine, as well as in high-risk hosts, the choice and duration of treatments are often complicated. Factors that favor surgical intervention are large size of abscesses, progressive enlargement of abscesses or destructive lesions, presence of bony sequestrations, instability of the spine, or impingement on critical organs or tissues (e.g., epidural abscess compressing the spinal cord). Newer antifungal agents such as voriconazole and posaconazole show promise as alternative therapy. Long-term prophylaxis with fluconazole can limit the risk for reactivation in immunosuppressed patients.

Coccidioidal synovitis may also occur as a consequence of immune complex–mediated inflammation, a presentation that may complicate either primary pulmonary or disseminated disease and is typically polyarticular. It is accompanied by fever, erythema nodosum or multiforme, eosinophilia, and hilar adenopathy. It abates in 2 to 4 weeks.

### BLASTOMYCOSIS

Blastomycosis, caused by Blastomyces dermatitidis, is endemic in the north-central and southern United States. Infection most commonly produces sporadic or clustered cases of pulmonary disease and is induced by exposure to soil or dust containing decomposed wood and, presumably, contaminated with the organism. Affected individuals do not appear to have any distinguishing or predisposing characteristics except for exposure to the organism during work or recreation. Clinical presentation includes high fever and other constitutional symptoms, pulmonary and skin involvement, and a significant mortality rate; leukocytosis and an elevated sedimentation rate are frequently seen. Bone pain, swelling, and soft tissue abscesses are the most common manifestations of osteoarticular disease. Hematogenous dissemination is common; skin disease and osteoarticular disease occur most frequently. Bone involvement occurs in 25% to 60% of disseminated cases, and arthritis is estimated to occur in 3% to 5%. In a study of 45 patients with skeletal blastomycosis, 41 had osteomyelitis while 12 presented with septic arthritis. The skeletal areas most commonly affected are the long bones, vertebrae, and ribs (Figure 112-1).

Arthritis is usually monoarticular in the knee, ankle, or elbow but may rarely be polyarticular. Joint infection is an isolated skeletal disorder in only a few cases; joint radiographs more commonly show punched-out bone lesions (Figure 112-2A). Synovial fluid is commonly purulent, and organisms are evident on microscopic examination, as well as by culture. The synovial histologic examination shows epithelioid granulomas with budding yeast forms (Figure 112-2B). The diagnosis is also commonly made from involved nonarticular sites. Urinary antigen testing appears to be sensitive but may be falsely positive in patients with other endemic fungal infections. For moderately severe or severe disease, treatment with amphotericin B for 1 to 2 weeks or until improvement is noted, followed by oral itraconazole for a total of at least 12 months, is recommended. For mild to moderate disease, oral itraconazole for 12 months is recommended. Serum levels of itraconazole should be determined after the patient has been on treatment for at least 2 weeks, to ensure adequate drug exposure.

### CRYPTOCOCCOSIS

Cryptococcus neoformans, the fungus causing cryptococcosis, is geographically ubiquitous and is found in pigeon feces; a related species, Cryptococcus gattii, is associated with certain types of eucalyptus trees in tropical climates and has been associated with an ongoing outbreak of disease on Vancouver Island and surrounding areas of Canada and the northwestern United States. It is a common pathogen in immunocompromised patients and is associated with defects in cell-mediated host defense including human immunodeficiency virus (HIV) infection, transplantation, lymphoreticular malignant

<table>
<thead>
<tr>
<th>Organism</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspergillus</td>
<td>1, 49, 101</td>
</tr>
<tr>
<td>Candidiasis</td>
<td>1</td>
</tr>
<tr>
<td>Coccidioidomycosis</td>
<td>99, 100</td>
</tr>
<tr>
<td>Cryptococcus</td>
<td>29, 102</td>
</tr>
<tr>
<td>Histoplasmosis</td>
<td>69, 72</td>
</tr>
<tr>
<td>Pneumocystis</td>
<td>110, 119</td>
</tr>
<tr>
<td>Scedosporiosis</td>
<td>103</td>
</tr>
<tr>
<td>Sporotrichosis</td>
<td>49</td>
</tr>
</tbody>
</table>
neoplasms, TNF antagonist treatment,\textsuperscript{29,30} and corticosteroid therapy.

Cryptococcosis varies in acuity, usually affecting the lungs in its primary form, but it sometimes disseminates hematogenously to a wide variety of sites including the central nervous system and skin. Although bone infection is common, causing osteolytic lesions in 5\% to 10\% of cases, articular involvement is rarely reported.\textsuperscript{31,32} Bone lesions may be confused with metastatic neoplasm (Figure 112-3).

Cryptococcal arthritis is an indolent monoarticular arthritis in about 60\% of reported cases and a polyarthritis in the remainder.\textsuperscript{31,34} The knee is most commonly involved. A single case of tenosynovitis with carpal tunnel syndrome has been recognized. The majority of cases reported from the pre-AIDS era also demonstrated radiographic evidence of periarticular osteomyelitis. These patients were young adults, did not have debilitating disease or other evidence of dissemination, and had pulmonary involvement in only 50\% of cases. Synovial tissue showed acute and chronic synovitis, multinucleate giant cells, prominent granuloma formation, and large numbers of budding yeast with special stains. Most recently reported cases are associated with immunosuppression and disseminated infection. Interestingly, osteoarticular cryptococcal infections have been linked to sarcoidosis, although it is unclear whether this association relates to the immunologic impact of the sarcoidosis or to the immunosuppressive therapies used to treat it.\textsuperscript{35} Serum cryptococcal antigen testing appears to be sensitive, in part because osteoarticular infection results from hematogenous dissemination. The choice of treatment for cryptococcal disease depends on both the anatomic sites of involvement and the host’s immune status, with amphotericin B and fluconazole being considered most effective.\textsuperscript{34,36} 5-Flucytosine is typically added to amphotericin B or fluconazole for the first 2 to 4 weeks (induction therapy) in cases of severe cryptococcal infection.\textsuperscript{37}

**CANDIDIASIS**

Candida species are widely distributed yeasts. *Candida albicans* is a normal commensal of humans, and other species can probably live in nonanimate environments such as soil. Since the advent of antibiotic therapy in the 1940s, and related to the common use of immunosuppression and parenteral lines, candidiasis has been responsible for an increasing incidence of mucocutaneous and deep-organ infections.\textsuperscript{38} Osteomyelitis, though rarely reported, is a potentially serious complication of hematogenous dissemination in both adults and children.\textsuperscript{39,40} It may also occur from direct tissue inoculation during surgery or by injection of contaminated heroin,\textsuperscript{41} and bone infection may emerge after successful amphotericin B treatment of other sites. Infection is commonly located in two adjacent vertebrae or in a single long bone. Surgical inoculation has occurred in the sternum, spine, and mandible. A few patients have had multiple sites of involvement. Candidal prosthetic joint infections may also occur as a late consequence of total joint replacement.\textsuperscript{42}

The clinical presentation is localized pain. Other symptoms and laboratory abnormalities vary. Bone changes of osteomyelitis are commonly demonstrated in radiographs of the symptomatic site. The diagnosis is established when culture of involved bone obtained by either open or needle biopsy has identified a variety of *Candida* species. Use of direct amplification testing has been reported.\textsuperscript{43} Treatment with azoles (fluconazole, itraconazole, voriconazole, or posaconazole); echinocandins (caspofungin, micafungin, or anidulafungin); or amphotericin B formulations may be effective.\textsuperscript{44,45} Species identification and susceptibility testing assist antifungal therapy selection. For example, *Candida glabrata* frequently demonstrate reduced susceptibility to azole antifungal agents and amphotericin but remain susceptible to echinocandins; *Candida krusei* are often resistant.
to azoles and may show reduced susceptibility to amphotericin but also remain susceptible to echinocandins; Candida lusitaniae are often resistant to amphotericin; and Candida parapsilosis may demonstrate reduced susceptibility to echinocandins. The use of surgical débridement must be individualized. With vertebral involvement but no neurologic complications, medication alone has been effective.

Candidiasis is an uncommon cause of monoarticular arthritis. C. albicans is the most common pathogen, although septic arthritis due to other candidal species has been reported. Reported cases commonly involve a knee, occur in the context of multifocal extra-articular Candida infection, and are accompanied by constitutional symptoms. Both children and adults have been affected. Predisposing conditions include gastrointestinal and pulmonary disorders, narcotic addiction, intravenous catheters, leukopenia, immunosuppressive treatment (including TNF antagonists), broad-spectrum antibiotics, and corticosteroids. Some involved joints were previously affected by arthritis, and infection has followed arthrocentesis in isolated cases. In most cases, radiographs reveal coincident osteomyelitis. Synovial fluid leukocyte counts may vary; Candida species have been cultured from synovial fluid in all cases but are not commonly identified on smear. Histologic studies of synovium show nonspecific chronic inflammation rather than granulomas.

**SPOROTRICHOSIS**

Sporotrichosis is caused by Sporothrix schenckii, a saprophyte found widely in soil and plants. Infection in humans occurs through inoculation of the skin or, rarely, inhalation into the respiratory tract; it is a source of infection among agricultural workers in tropical and subtropical areas. It most commonly involves the skin and lymphatics but may disseminate from the lungs to the central nervous system, eyes, bones, and joints. In immunocompetent hosts, a single site is typically involved; in immunocompromised hosts including patients on anticytokine therapy, multifocal disease may occur.

In contrast to the relatively common occurrence of skin infection, articular sporotrichosis is a rare disorder. In 84% of patients in one series, there was no accompanying skin involvement, suggesting entry through the lungs. Sporotrichosis most often occurs in individuals with a chronic illness that alters host defense such as alcoholism or a myeloproliferative disorder. Sporothrix arthritis is most often indolent and infects a single joint or multiple joints in equal proportions. The knee, hand, wrist, elbow, and shoulder are most frequently involved; hand and wrist involvement distinguishes this from other fungal arthritides. Articular infection shows a propensity to spread to adjacent soft tissues.
forming draining sinuses. Constitutional symptoms are unusual.

Radiographic changes vary from juxta-articular osteopenia to the commonly observed punched-out bone lesions. When it can be obtained, synovial fluid is inflammatory. Synovitis is characterized on gross evaluation by destructive pannus and on microscopic examination by granulomatous histologic features or, less frequently, by nonspecific inflammation. Organisms are difficult to identify in tissue, and the diagnosis is often made by positive culture of joint fluid or involved tissue. Incubation at room temperature assists growth of the mycelial phase of S. schenckii. Serologic testing is not useful in the diagnosis of sporotrichosis. In a small number of cases, sporotrichosis may disseminate to cause a potentially fatal infection characterized by low-grade fever, weight loss, anemia, osteolytic bone lesions, arthritis, skin lesions, and involvement of the eyes and central nervous system.52-55 These infections occur in immunosuppressed patients with either hematologic malignancies or HIV infection.

In 44 cases reported in 1979, treatment was optimal with combined joint débridement and high-dose intravenous amphotericin B (11 of 11 cured) and slightly less effective with amphotericin alone (14 of 19 cured).56 More recently, itraconazole has proven effective for initial therapy of most patients,56 with amphotericin B being reserved for those with extensive involvement and for itraconazole failures. In contrast, fluconazole has demonstrated only modest success in osteoarticular sporotrichosis.57,58 Long-term suppressive therapy with itraconazole may be required for patients with AIDS.56

**ASPERGILLOSIS**

Aspergillus species are ubiquitous, but infection occurs only rarely in immunocompetent individuals. In contrast, invasive infection is an important life-threatening complication in immunocompromised adults and children.59-61 It may spread directly from the lung to adjacent vertebrae, disk spaces, and ribs (more often in children) or through the bloodstream (Figure 112-4).62-64 Rare cases of monoarthritis with adjacent osteomyelitis are also reported. The knee is the most commonly involved joint. The organism may be observed in infected tissue (see Figure 112-4B). The galactomannan EIA has been validated as a surrogate marker for invasive aspergillosis and the (1→3)-β-D-glucan assay may similarly provide support for the diagnosis of invasive fungal infection. Treatment with combined surgical débridement and antifungal therapy is an ongoing challenge.60,63,65 Voriconazole has proven superior to amphotericin for the treatment of invasive aspergillosis and is now recommended for initial therapy.65,66 Liposomal amphotericin B, posaconazole, caspofungin, and micafungin have demonstrated efficacy in salvage situations.67 Combination therapy has not demonstrated clear improvement in outcomes over monotherapy.

**HISTOPLASMOSIS**

*Histoplasma capsulatum* is a soil fungus that causes endemic disease in the midwestern and southeastern United States.33,68 Bone and joint involvement is rare but has been reported in the knee, wrist, and ankle. Immunosuppression
including the use of TNF antagonists predisposes to disseminated histoplasmosis in adults and children, which may be confused clinically with sarcoidosis, tuberculosis, and reactive inflammatory conditions. Diagnosis depends on appropriate use of fungal staining and culture methods, antigen detection, and serologic antibody testing. A case report emphasizes the rare occurrence of fungal prosthetic joint arthritis. The more common osteoarticular involvement with histoplasmosis is a hypersensitivity syndrome accompanying acute pulmonary infection; it is characterized by self-limited polyarthritis, erythema nodosum, and erythema multiforme. Liposomal amphotericin B followed by itraconazole is the preferred treatment for severe infection and itraconazole for less severe cases.

SCEDOSPORIOSIS

Scedosporium species are environmental molds that have recently been identified as fungal pathogens in both immunocompetent and immunocompromised hosts. They may cause focal invasive and disseminated infection after cutaneous inoculation. Scedosporium prolificans has a predilection for bone and cartilage, leading to both septic arthritis and osteomyelitis. Infections are difficult to eradicate with surgery and antifungal agents, and the organism is resistant to amphotericin. Case reports suggest improved infection control with voriconazole or voriconazole combined with terbinafine.

TREATMENT OF FUNGAL INFECTION

Antifungal chemotherapy has improved over the past several decades, first with the introduction of amphotericin B and then with the oral antifungal agents fluconazole, ketoconazole, fluconazole, and itraconazole. More recent advances include the development of less toxic formulations of amphotericin B, liposomal amphotericin B, and amphotericin B lipid complex. Voriconazole and posaconazole, which are broad-spectrum azole antifungals, have demonstrated improved activity against aspergillosis and mucormycosis, respectively. The echinocandin antifungal agents caspofungin, micafungin, and anidulafungin have emerged as alternative therapies for aspergillosis and as the treatments of choice for some Candida infections. For detailed treatment guidelines, several excellent reviews are available. In choosing an appropriate drug (Table 112-2) and course of treatment, the clinician must consider the infecting agent, clinical manifestations of the disease, immune status of the host, antimicrobial resistance, drug side effect profile, and direct and indirect costs of treatment. Treatment has become more complex because of immune system compromise in infected patients being treated for transplant rejection, autoimmune disorders, malignant disease, and AIDS.

Itraconazole has become the first choice for treatment of the endemic mycoses—blastomycosis, histoplasmosis, and sporotrichosis. A loading dose of itraconazole of 200 mg three times a day for 3 days is recommended, followed by 200 mg to 400 mg daily. Absorption of itraconazole is unpredictable, and blood levels of itraconazole should be measured to ensure adequate drug exposure. Absorption of itraconazole requires stomach acid, so concurrent administration of drugs that reduce the acidity of the stomach such as proton pump inhibitors and H₂ blockers should be avoided. At least 6 months of treatment is required, and some patients may need up to a year of therapy. For cryptococcosis, fluconazole is the recommended azole. Amphotericin B is the preferred drug for meningeal and life-threatening infections. Specific treatment protocols and detailed side effect profiles are presented in reviews. Infectious Diseases Society of America guidelines and infection-specific references (see Table 112-2).

FUNGAL INFECTION AS A CONSEQUENCE OF ANTI RHEUMATIC THERAPY

Many of the same fungal infections associated with osteoarticular disease may cause infection in patients treated with antirheumatic therapy, particularly biologic therapy. Animal models of infection point to an important role for TNF in host defense against many of these organisms including Aspergillus, Candida, Cryptococcus, Coccidioides, and Sporothrix. Coccidioidomycosis infection following TNF antagonist therapy has been reported in endemic areas in the southwestern United States. In the largest published series, 12 of 13 cases followed therapy with infliximab and 1 was associated with etanercept. All but two of these cases were believed to represent new infection rather than reactivation. In one medical center included in this series, the relative risk of coccidioidomycosis infection was 5.23 with infliximab therapy compared with other antirheumatic therapy. In all cases, coccidioidomycosis infection developed in the absence of other known risk factors including diabetes, pregnancy, and HIV infection. Coccidioidomycosis infection has also been reported in nonendemic areas, presumably as a consequence of fomite exposure. Disseminated histoplasmosis infections have been reported following TNF antagonist therapy in endemic areas in the Ohio-Mississippi River valley regions of the central United States; as with coccidioidomycosis infections, these have been reported more frequently following infliximab therapy than with the other agents. The reason for the larger number of cases with infliximab is unclear; possibilities include a unique mechanism of action for infliximab, larger numbers of patients treated, patient selection, concomitant immunosuppressive therapy, or some combination of all of these. In most cases, patients

---

Table 112-2 Drug Treatment of Osteoarticular Mycotic Infections

<table>
<thead>
<tr>
<th>Infection</th>
<th>Drug Recommended</th>
<th>Infection-Specific References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coccidioidomycosis</td>
<td>Itraconazole</td>
<td>3, 7, 12, 13</td>
</tr>
<tr>
<td>Blastomycesis</td>
<td>Itraconazole</td>
<td>18, 27, 117</td>
</tr>
<tr>
<td>Cryptococcus</td>
<td>Fluconazole</td>
<td>34, 36, 117</td>
</tr>
<tr>
<td>Candidiasis</td>
<td>Fluconazole</td>
<td>44-46, 117, 118</td>
</tr>
<tr>
<td>Sporotrichosis</td>
<td>Itraconazole</td>
<td>50, 57, 58</td>
</tr>
<tr>
<td>Aspergillosis</td>
<td>Voriconazole</td>
<td>61, 64, 117</td>
</tr>
<tr>
<td>Histoplasmosis</td>
<td>Itraconazole</td>
<td>68, 75</td>
</tr>
<tr>
<td>Scedosporosis</td>
<td>Voriconazole</td>
<td>77, 80, 81</td>
</tr>
</tbody>
</table>
developing disseminated histoplasmosis with TNF antagonist therapy have been taking additional immunosuppressive therapy. Patients typically present with cough, dyspnea, fever, and malaise and can rapidly become quite ill. Histoplasma antigen may be identified in the urine in 92% of cases with disseminated histoplasmosis, and this test may assist rapid diagnosis. Exposure to Histoplasma may result in asymptomatic latent infection; it has been difficult, however, to determine whether symptomatic infections following TNF antagonist therapy represent reactivation or new infection. Both pulmonary and disseminated infections with cryptococcal infections have been reported with TNF antagonist therapy, as have infections with Aspergillus, Candida, Scedosporium, and Sporothrix. Fungal infections have not been reported as a consequence of abatacept or rituximab therapy in rheumatic diseases. Whether this relates to decreased risk with these agents or smaller numbers of patients treated remains to be seen. Pneumocystis jiroveci pneumonia (PCP) is caused by a fungus originally classified as a protozoan (Pneumocystis carinii). PCP, an opportunistic infection seen in association with HIV infection, has been reported with a number of antirheumatic therapies including cyclophosphamide and low-dose methotrexate, often in combination with corticosteroid therapy. Interestingly, PCP as a consequence of low-dose methotrexate therapy has been reported as a particular concern in Japan, where the prevalence of asymptomatic carriage of Pneumocystis has been reported to be as high as 18.8% in the elderly. These same authors have suggested that polymerase chain reactive (PCR) testing may identify carriers at increased risk for PCP during therapy. The use of biologic therapy including rituximab and TNF antagonists has been associated with the development of PCP. Pneumocystis jiroveci pneumonia presents as fever, dry cough, and dyspnea. The clinical presentation in patients receiving immunosuppressive therapy is typically more acute than in HIV-associated cases. Organisms may be identified in sputum from HIV patients but are seen less commonly in rheumatologic patients; presumptive diagnosis may be made on the basis of PCR testing for P. jiroveci DNA. Elevated serum levels of β-D-glucan, a common component of fungal cell walls, may aid in the diagnosis. One series of 21 patients identified older age, pre-existing pulmonary disease, higher corticosteroid doses, and low serum albumin and IgG levels as potential risk factors for PCP in patients treated with infliximab. Chest radiographs may show diffuse infiltrates, while CT scans demonstrate ground-glass opacities (Figures 112-5 and 112-6); radiographic findings may be difficult to distinguish from methotrexate pneumonitis. Treatment includes supplemental oxygen and trimethoprim/sulfamethoxazole (TMP/SMX) or pentamidine isethionate. High-dose corticosteroids are commonly used as adjunctive therapy in the treatment of HIV patients with PCP but are less well studied in patients taking immunosuppressives.

PCP infection is also of concern with nonbiologic immunosuppressive therapy. In particular, the use of cyclophosphamide to treat systemic lupus, vasculitis, and other autoimmune diseases has been associated with a risk of developing PCP, although the overall risk appears to be low. In a recent review of published data on 76,156 cases of systemic lupus erythematosus treated with cyclophosphamide, the risk of PCP was 15.88 per 10,000 patients, or 0.158%.

Potential risk factors include high-dose corticosteroid use, lymphopenia (especially low CD4 counts), renal disease, and overall high disease activity. Vasculitis patients with pulmonary involvement may be at increased risk, and diagnosis may be difficult in both these patients and lupus patients with pulmonary disease. Unfortunately, there are no published guidelines to address the use of PCP prophylaxis in autoimmune diseases or a clear consensus regarding the standard of care in this situation. Recent clinical trials in vasculitis have routinely employed PCP prophylaxis as part of the protocol, suggesting it is becoming the standard of care. Two recent surveys of U.S. rheumatologists, however, found that just 50.4% and 69.5%, respectively, routinely prescribed prophylactic antibiotics. Academic rheumatologists and more recent graduates are more likely to prescribe prophylactic therapy. TMP/SMX appears to be more effective than dapsone or aerosolized
References


The references for this chapter can also be found on www.expertconsult.com.
Rheumatic Manifestations of Human Immunodeficiency Virus Infection

JOHN D. REVEILLE • EMILY W. HUNG

In the years since acquired immunodeficiency syndrome (AIDS) was initially described in 1981, the human immunodeficiency virus (HIV) pandemic has become one of the leading global health crises. According to new data in the Joint United Programme on HIV/AIDS (UNAIDS) 2008 report, the AIDS epidemic seems to be slowing down globally, but new cases are continuing to increase at alarming rates in certain regions, such as southern Africa, eastern Europe, and central and eastern Asia. An estimated 33 million people worldwide are living with HIV (Figure 113-1). Approximately 2.7 million people became newly infected with HIV in 2007, and 2 million people died.

Progress in dealing with the HIV epidemic, including progress in education and in public health awareness, has undoubtedly influenced the decrease in prevalence seen among young people in some countries in recent years. As availability of newer treatment strategies and better access to health care result in increased life expectancy in the next decade, it is expected that HIV infection will be managed increasingly as a chronic illness, and complications such as musculoskeletal and rheumatic conditions associated with HIV infection and its treatment are expected to increase (Tables 113-1 and 113-2).

Among the rheumatologic disorders, clinicians face the challenge of treating potentially disabling inflammatory disorders with immunosuppressive therapy in the face of ongoing viral-induced immunocompromise. Diagnosing infection is especially important in an immunocompromised patient because the probability of an opportunistic infection as a cause for a musculoskeletal complaint increases with advancing stages of the patient’s HIV disease. At early stages (CD4+ count >300/µL), opportunistic infections are unlikely, although bacterial infections (especially tuberculosis) still can occur. There should be a very high threshold for using immunosuppressive drugs in this population.

HIV-ASSOCIATED BONE AND JOINT DISEASE

HIV-Associated Arthralgia

More than 5% of HIV-positive patients may have otherwise unexplained arthralgia. Arthralgias and myalgias also form a part of the constitutional symptoms of HIV seroconversion. Whether arthralgia can be attributed to circulating viral and host immune complexes owing to HIV infection per se or to other infections (e.g., hepatitis C) has not been determined. The pathogenesis is unclear but may involve cytokines or transient bone ischemia. However, patients presenting with arthralgia alone rarely progress to inflammatory joint disease. The most appropriate treatment consists of non-narcotic analgesics and reassurance.

Painful Articular Syndrome

Painful articular syndrome is a self-limited syndrome lasting less than 24 hours, associated with few objective clinical findings, and characterized by severe bone and joint pain. It occurs predominantly in the late stages of HIV infection. Its cause is unknown, and no evidence of synovitis has been found in these patients. The knee is most commonly affected, but the elbow and shoulders also can be involved. Radiographic features are nonspecific; occasionally, periaricular osteopenia is seen. Treatment is symptomatic.

HIV-Associated Arthritis

The first reports of a seronegative arthritis associated with HIV infection appeared in 1988, with frequencies of 12%. HIV-associated arthritis seems to be most common in sub-Saharan Africa, where HIV infection is pandemic. In the Congo, where the seroprevalence of HIV infection is 7% to 8%, AIDS is the leading cause of aseptic arthritis (60% of cases). This is usually an oligoarthritis (Table 113-3),
anti-inflammatory drugs (NSAIDs) and, in more severe cases, low-dose glucocorticoids. Hydroxychloroquine and sulfasalazine also have been used.7

Reactive Arthritis Occurring in HIV Infection

Early reports in the United States suggested that reactive arthritis occurred more commonly in the setting of HIV infection; however, later studies showed that this may be reflective of the sexually active nature of the population at highest risk for HIV infection.8 This contention is not borne out in studies from sub-Saharan Africa, where HLA-B27 is rare, as were reports of spondyloarthritis before the HIV epidemic. With the arrival of AIDS, a dramatic upsurge in the prevalence of reactive arthritis and undifferentiated spondyloarthritis, and less often psoriatic arthritis,6,9 was seen, suggesting a pathogenic role of HIV infection.

The typical presentation is a seronegative lower extremity peripheral arthritis, usually accompanied by enthesitis (dactylitis, Achilles tendinitis, and plantar fasciitis). Mucocutaneous features are common, especially keratoderma blennorrhagicum (Figure 113-2) and circinate balanitis. Extensive psoriasiform skin rashes can occur. The clinical overlap makes it difficult sometimes to distinguish HIV-associated reactive arthritis from psoriatic arthritis.10

<table>
<thead>
<tr>
<th>Table 113-1 Rheumatic Diseases Associated with or Occurring in Patients with Human Immunodeficiency Virus (HIV) Infection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unique to HIV Infection</strong></td>
</tr>
<tr>
<td>Diffuse infiltrative lymphocytosis syndrome</td>
</tr>
<tr>
<td>HIV-associated arthritis</td>
</tr>
<tr>
<td>Zidovudine-associated myopathy</td>
</tr>
<tr>
<td>Painful articular syndrome</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

IRIS, immune reconstitution inflammatory syndrome.

Table 113-2 Distribution of Various Rheumatic Diseases from Various Sites

<table>
<thead>
<tr>
<th>Feature</th>
<th>Cincinnati, Ohio</th>
<th>Houston, Texas</th>
<th>Madrid, Spain</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>1100</td>
<td>4467</td>
<td>556</td>
</tr>
<tr>
<td>HIV-associated arthralgia</td>
<td>NA</td>
<td>0.7%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Myalgia</td>
<td>0.7%</td>
<td>0.6%</td>
<td>4.5%</td>
</tr>
<tr>
<td>PsA/Reactive arthritis</td>
<td>0.5%</td>
<td>0.6%</td>
<td>0.5%</td>
</tr>
<tr>
<td>HIV-associated arthritis</td>
<td>0%</td>
<td>0.5%</td>
<td>0.4%</td>
</tr>
<tr>
<td>DILS/Sjögren’s syndrome</td>
<td>NR</td>
<td>3%-4%</td>
<td>NR</td>
</tr>
</tbody>
</table>

DILS, diffuse infiltrative lymphocytosis syndrome; HIV, human immunodeficiency virus; NA, not applicable; NR, not reported; PsA, psoriatic arthritis.
Urethritis occurs in similar frequency as in HIV-negative reactive arthritis. Axial involvement and uveitis seem to be less common but do occur. Longitudinal studies from Africa have described an aggressive course with a poor prognosis.11,12

HLA-B27 is found in 80% to 90% of patients with HIV-associated reactive arthritis, at least among whites.10 Studies from Africa have found most patients to be HLA-B27 negative, however.11,12 Some studies suggest that the presence of HLA-B27 antigen may slow the progression to AIDS.11,12 In asymptomatic HIV-infected, HLA-B27-positive individuals, cytotoxic T lymphocyte response is dominated by recognition of a gag-encoded p24 protein epitope that is not seen in HIV-positive, HLA-B27-negative individuals.13,14 Other human leukocyte antigen (HLA) class I antigens that have been associated with a better outcome in HIV infection have been implicated in psoriasis and psoriatic arthritis and include HLA-B13 and HLA-B17 (B57, B58).12,13 HLA-B*5703 was protective against HIV progression in a Zambian population but conferred susceptibility to spondyloarthritis.15

**Treatment**

Treatment is similar to that for HIV-negative patients with reactive arthritis. NSAIDs are the mainstay; in particular, indomethacin is recommended, not only for its efficacy, but also for its inhibition of HIV replication that has been observed in vitro, which seems to be unique to this NSAID.16 Patients frequently have an inadequate response to NSAIDs alone. Sulfasalazine has been shown to be effective in some studies at doses of 2 g/day, and one study suggested that it ameliorated HIV infection.17,18 Methotrexate was initially believed to be contraindicated because of its immunosuppressive effect, but with careful monitoring of HIV viral loads, CD4+ counts, and the patient’s clinical status, more recent studies have suggested a place for methotrexate in the treatment of reactive arthritis and psoriatic arthritis occurring in HIV infection.19

Hydroxychloroquine has been reported to be efficacious not only in treating HIV-associated reactive arthritis, but also in reducing HIV replication in vitro and in reducing HIV viral loads in vivo.20 Arthritis and the cutaneous lesions of HIV-associated reactive arthritis and psoriatic arthritis have been found to respond to etretinate (0.5 to 1 mg/kg/day),21 although because of the side effects of this drug, its use should be reserved for patients unresponsive to other treatments. Tumor necrosis factor blockers have been used,22,23 although these agents should be used with extreme caution and only in patients with CD4+ counts greater than 200/µL and HIV viral load less than 60,000 copies/mm3.24,25 One prospective study of eight HIV patients with spondyloarthritis or rheumatoid arthritis found tumor necrosis factor blockers effective and safe for up to 5 years when these precautions were followed at initiation of therapy.25

**Psoriasis and Psoriatic Arthritis**

The extent of skin involvement with psoriasis can be extensive (Figure 113-3) in HIV-positive patients, especially in patients not on antiretroviral treatment. Of note, cutaneous T cell lymphoma can resemble psoriasis and should be considered in the differential diagnosis of psoriasis in HIV-positive individuals.26 A report from Zambia found 27 of 28 African patients with psoriatic arthritis to be HIV positive.27 The arthritis was predominantly polyarticular, lower limb, and progressive. Psoriasis was commonly an extensive guttate-plaque admixture and, in contrast to the articular disease, was nonremittive with the onset of AIDS.28 Anti-retroviral treatment has been shown to be effective in treating HIV-associated psoriasis and its associated arthritis.29 Phototherapy may improve the skin rash but also may enhance viral replication, worsen HIV disease, and increase the risk of skin cancer. Other agents reported to be efficacious include cyclosporine (although renal function must be monitored carefully) and etretinate. Methotrexate also can be used, albeit with caution.30 Tumor necrosis factor blockers can be used in patients with refractory disease, and a
Hypertrophic Pulmonary Osteoarthropathy

Hypertrophic pulmonary osteoarthropathy affects bones, joints, and soft tissues and can develop in HIV-infected patients with *Pneumocystis jiroveci* pneumonia. It is characterized by severe pain in the lower extremity; digital clubbing; nonpitting edema; and periarticular soft tissue involvement of the ankles, knees, and elbows. The skin over the affected areas is glistening, edematous, and warm. Radiography reveals extensive periosteal reaction and subperiosteal proliferative changes in the long bones of the lower extremity. A bone scan shows increased uptake along the cortical surfaces. Treatment of *P. jiroveci* pneumonia usually alleviates this condition.

Osteopenia and Osteoporosis

Osteopenia and osteoporosis occur more than three times as commonly in HIV-infected patients regardless of antiretroviral treatment and can result in pathologic fractures. One meta-analysis found 15% of HIV-positive patients to have osteoporosis, and 52% to have osteopenia. Abnormal bone metabolism was attributed to the HIV infection itself by some authors. Risk factors for the development of osteopenia include use of protease inhibitors, longer duration of HIV infection, high viral load, high lactate levels, low bicarbonate levels, increased alkaline phosphatase levels, and lower body weight before antiretroviral therapy.

Vitamin D deficiency is also common in HIV patients, with reported frequency of 47% for moderate to severe vitamin D deficiency in one cohort. A retrospective review of 211 HIV-positive patients found vitamin D deficiency to be associated with concomitant hepatitis C infection, previous AIDS, and higher CD4+ counts. Clinicians should have a low threshold to screen for vitamin D deficiency and should provide adequate repletion as necessary. Bisphosphonates and, in patients with HIV wasting syndrome, testosterone, have been used to preserve bone density.

HIV-ASSOCIATED MUSCLE DISEASE

Muscle involvement in HIV infection varies from uncomplicated myalgias or asymptomatic creatine kinase elevation to severe, disabling, HIV-associated polymyositis or pyomyositis (Table 113-4). HIV seroconversion also can coincide with myoglobinuria and acute myalgia, suggesting that myotropism for HIV may be present early in the infection.

Myalgia and Fibromyalgia

One-third of HIV-positive outpatients report myalgias, and 11% describe fibromyalgia. Fibromyalgia is associated with longer disease duration and a history of depression. Treatment is similar to that for fibromyalgia in the non-HIV setting.

Noninflammatory Necrotizing Myopathy and HIV-Related Wasting Syndrome

Severe wasting from chronic infection, malignancy, malabsorption, and nutritional deficiency often accounts for...
weakness and disability in patients with AIDS. This wasting leads to loss of lean body and muscle mass. Cachexia and muscle wasting associated with HIV constitute a slim disease in Africa. A noninflammatory necrotizing myopathy of unclear pathogenesis has been described, accounting for 42% of patients diagnosed with myopathy. Even in patients without significant wasting, muscle biopsy specimens have shown diffuse atrophy, mild neurogenic atrophy, or thick filament loss without conspicuous inflammation. Whether this condition is immune mediated, as some have suggested, or whether it is due to metabolic or nutritional factors remains unclear. Corticosteroids have been reported to restore muscle strength and mass.

Nemaline Myopathy

Nemaline myopathy is a rare disorder that has been described in some HIV-positive patients, in addition to its occurrence as a congenital disorder. Nemaline myopathy represents a nonspecific myofibril alteration resulting from Z band disruption. Muscle biopsy specimens disclose prominent, randomly distributed atrophic type 1 fibers with numerous intracytoplasmic rod bodies in the centers of the fibers, corresponding to nemaline rods at electron microscopy. Necrotic fibers and inflammatory infiltrates usually are not found. Some patients have been described to have associated monoclonal gammopathy. Although no inflammation is noted, corticosteroids may be useful. In addition, two cases of successful treatment with intravenous immunoglobulin (IVIG) have been reported.

HIV-Associated Polymyositis

HIV-associated polymyositis most typically manifests early in the course of HIV infection and may be the presenting feature. In one large series of HIV-positive outpatients from a county clinic in Texas, the frequency was 2.2 per 1000. The pathogenesis of HIV-associated polymyositis is unclear—possibly stemming from direct viral invasion (leading to a cytopathic effect and subsequent muscle necrosis), as suggested in one pathologic study, or from an autoimmune response of the HIV host, as suggested by another study. The most common manifestation is a subacute, progressive proximal muscle weakness occurring in the setting of an elevated creatine kinase. Myalgia is not a prominent presenting feature. Skin involvement is unusual, as is involvement of extraocular muscles and facial muscles. On the other hand, only a handful of cases of dermatomyositis in HIV have been reported, usually in the setting of advanced immunodeficiency. 

It has been suggested that creatine kinase levels may be less elevated or even normal in some HIV-associated polymyositis patients. One retrospective report from sub-Saharan Africa found that creatine kinase elevations were fourfold lower in those with HIV-associated polymyositis than in patients with polymyositis without HIV infection. On MRI, T2-weighted studies with or without fat saturation show high signal intensity without rim enhancement, in contrast to pyomyositis, in which rim enhancement is seen. MRI also is helpful in guiding muscle biopsy—the definitive diagnostic test. Electromyographic studies reveal myopathic motor unit potentials with early recruitment and full interference patterns and fibrillation potentials, positive sharp waves, and complex repetitive discharges indicative of an irritative process. Light microscopy of muscle biopsy specimens shows interstitial inflammatory infiltrates of variable intensity accompanied by degenerating and regenerating myofibrils, similar to those seen in polymyositis without HIV (Figure 113-4). Concomitant vasculitis rarely occurs. In specimens from HIV-positive and HIV-negative patients with myositis, the predominant cell populations were CD8+ T cells and macrophages invading or surrounding healthy muscle fibers that express major histocompatibility complex (MHC) class I antigens on their cell surfaces. Endomysial infiltrates in specimens from HIV-positive patients differed from those of patients with polymyositis without HIV infection only by a significant reduction of CD4+ cells.

| Table 113-4 Myopathies Associated with Human Immunodeficiency Virus (HIV) Infection
<table>
<thead>
<tr>
<th>HIV-Associated Myopathies</th>
<th>Myopathies Secondary to Antiretrovirals</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV polymyositis</td>
<td>Zidovudine myopathy</td>
<td></td>
</tr>
<tr>
<td>Inclusion body myositis</td>
<td>Toxic mitochondrial myopathies related to other NRTIs</td>
<td></td>
</tr>
<tr>
<td>Nemaline myopathy</td>
<td>HIV-associated lipodystrophy syndrome</td>
<td></td>
</tr>
<tr>
<td>Diffuse infiltrative lymphocytosis syndrome</td>
<td>Immune reconstitution syndrome related to HAART</td>
<td></td>
</tr>
<tr>
<td>HIV wasting syndrome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vasculitic processes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myasthenia gravis and other myasthenic syndromes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic fatigue and fibromyalgia</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HAART, highly active antiretroviral therapy; NRTIs, nucleoside reverse transcriptase inhibitors.

Figure 113-4 Muscle biopsy specimen from a patient with human immunodeficiency virus–associated polymyositis.
Treatment is similar to that provided for other inflammatory myopathies. Creatine kinase elevation and muscle weakness respond to moderate-dose glucocorticoids. Refractory cases may require immunosuppressive agents, such as methotrexate, azathioprine, or mycophenolate mofetil. Intravenous immunoglobulin has been used with some success. These agents should be used with caution, however, with careful monitoring of the patient’s clinical status, CD4+ counts, and HIV mRNA levels.

Creatine kinase elevation is commonly encountered in outpatients with HIV infection, secondary to HIV per se, behaviors associated with higher risk for HIV infection (e.g., cocaine use), or HIV treatment. In most patients, these elevations are transient and are of little consequence, but they require careful follow-up for any sign of clinical deterioration before electrodiagnostic and biopsy studies are undertaken.

**Inclusion Body Myositis**

Inclusion body myositis has been recognized as a complication of HIV infection. This condition is clinically, histologically, and immunologically identical to sporadic inclusion body myositis. Muscle biopsy specimens suggest two concurrently ongoing processes—an autoimmune process mediated by cytotoxic T cells and a degenerative process manifested by vacuolated muscle fibers and deposits of amyloid-related proteins. Of particular interest has been the finding of elevated mRNA levels and constitutive expression of Toll-like receptor 3, which is known to mediate inflammatory stimuli from pathogens and endogenous danger signals, and to link the innate and adaptive immune systems, in muscle fibers of patients with HIV-associated inclusion body myositis in close proximity with infiltrating mononuclear cells. One review of four cases of HIV-associated inclusion body myositis found that involved CD8+ cells surrounding muscle fibers were virus-specific and may cross-react with antigens on the surface of muscle fibers, suggesting that HIV may trigger a viral-specific inflammatory response that can lead to inclusion body myositis.

**Myopathy Associated with Treatment**

A reversible toxic mitochondrial myopathy occurring in patients who received high doses of zidovudine has been described, which manifests as myalgias, muscle tenderness, and proximal muscle weakness mimicking HIV polymyositis. Reports have also described mitochondrial toxicity presenting as ptosis or ophthalmoplegia in HIV-positive patients on zidovudine and other nucleoside reverse transcriptase inhibitors (NRTIs) such as didanosine. Histologically, it is characterized by the presence of ragged red fibers, a term coined to designate atrophic ragged red fibers with marked myofilament alterations, including thick myofilament loss and cytoplasmic body formation, and minimal inflammatory infiltrates. Symptoms tend to improve as the drug is discontinued, with creatine kinase levels returning to normal within 4 weeks of discontinuing the drug, and muscle strength returning within 8 weeks. In any HIV-infected patient presenting with an elevated creatine kinase, especially when symptoms of myalgia or muscle weakness are present, zidovudine should be discontinued for 4 weeks and the patient re-evaluated before electromyography or muscle biopsies are undertaken.

**Rhabdomyolysis**

Rhabdomyolysis can occur at all stages of HIV infection and may be separated into three groups: (1) HIV-associated rhabdomyolysis, including rhabdomyolysis in primary HIV infection, recurrent rhabdomyolysis, and isolated rhabdomyolysis; (2) drug-induced rhabdomyolysis; and (3) rhabdomyolysis at the end stage of AIDS, associated or not with opportunistic infection of muscle. Drugs implicated in rhabdomyolysis in HIV patients include didanosine, lamivudine, trimethoprim-sulfamethoxazole, ritonavir, indinavir, and raltegravir.

**DIFFUSE INFILTRATIVE LYMPHOCYTOSIS SYNDROME**

Diffuse infiltrative lymphocytosis syndrome (DILS), found exclusively in HIV-positive patients, is characterized by salivary gland enlargement and peripheral CD8+ lymphocytosis often accompanied by sicca symptoms and other extraglandular features. The prevalence of DILS is declining since the introduction of HAART. Using parotid enlargement as a criterion, the prevalence in Houston, Texas, was 4% in the pre-HAART era, declining to 0.8% after the introduction of aggressive HIV therapy. In another study from Greece, using xerophthalmia and xerostomia as the defining criteria (and requiring confirmatory minor salivary gland biopsy specimen and technetium scintigraphy), the prevalence of DILS was 7.8% and decreased dramatically after the introduction of HAART.

The primary immunogenetic association has been with HLA-DRB1 alleles expressing the ILEDE amino acid sequence in the third diversity region—usually HLA-DRB1*1102, DRB1*1301, and DRB1*1302. Delayed progression to AIDS in patients with DILS has been attributed to delay in the evolution of the HIV virus from the less aggressive M-tropic strain to the more rapidly replicating T-tropic strain by a more effective CD8+ lymphocyte response. This response has been attributed in part to the finding of sequence homology of a six-residue epitope shared by HLA-DRB1 alleles associated with DILS with a V3 loop on M-tropic HIV strains. Studies of immunophenotypes of circulating and tissue-infiltrating lymphocytes and salivary gland T cell receptor sequence analysis suggested that DILS represents an MHC-restricted, antigen-driven, oligoclonal selection of CD8+, CD29+ lymphocytes that express selective homing receptors and infiltrate the salivary glands, lungs, and other organs, where they are postulated to suppress HIV replication.

Minor salivary gland biopsy specimens show a focal sialadenitis, similar to that observed in Sjögren’s syndrome, although destruction of the salivary glands tends to be lessened (Figure 113-5). CD8+ lymphocytes constitute most of the inflammatory infiltrate, in contrast to that seen in primary (non–HIV-associated) Sjögren’s syndrome. Lymphoepithelial cysts are seen frequently in the parotid glands of patients with DILS, leading to inspissated salivary secretions that may be painful.
CHAPTER 113 | RHEUMATIC MANIFESTATIONS OF HUMAN IMMUNODEFICIENCY VIRUS INFECTION

1857

1. HIV seropositive by enzyme-linked immunosorbent assay and Western blot analysis.
2. Bilateral salivary gland enlargement or xerostomia persisting for longer than 6 months.
3. Histologic confirmation of salivary or lacrimal gland lymphocytic infiltration in the absence of granulomatous or neoplastic enlargement. Minor salivary gland biopsy specimens are usually positive (see Figure 113-5). Gallium-67 scintigraphy (Figure 113-7) of the salivary glands has been used when lip biopsy was not feasible or equivocal. Tc99m pertechnetate scanning offers little diagnostic help. Scintigraphy is used as a primary diagnostic aid in patients on protease inhibitors because minor salivary gland biopsy specimens are rarely positive in patients on these drugs. CT also has been used to determine the extent of glandular swelling and to evaluate parotid cysts and possible salivary glandular malignancy.

Patients with asymptomatic glandular swelling and mild, if any, sicca symptoms can be observed over time (Table 113-7). Antiretroviral treatment is effective in treating the glandular swelling and sicca symptoms associated with DILS

The characteristic, if not defining, presentation of DILS consists of painless parotid enlargement, often massive (Figure 113-6). This enlargement is accompanied by sicca symptoms in more than 60% of patients. Although parotid and submandibular enlargement is nearly universal in this disorder, certain extraglandular features also are prominent (Table 113-5). DILS and Sjögren’s syndrome share some similarities and differences (Table 113-6).

Diagnostic criteria have been proposed for DILS as follows:

Table 113-5  Extraglandular Features of Diffuse Infiltrative Lymphocytosis Syndrome

<table>
<thead>
<tr>
<th>Pulmonary</th>
<th>Neurologic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymphocytic interstitial pneumonitis*</td>
<td>Cranial nerve VII palsy†</td>
</tr>
<tr>
<td></td>
<td>Aseptic lymphocytic meningitis</td>
</tr>
<tr>
<td></td>
<td>Peripheral neuropathy</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td></td>
</tr>
<tr>
<td>Lymphocytic hepatitis</td>
<td></td>
</tr>
<tr>
<td>Renal</td>
<td></td>
</tr>
<tr>
<td>Renal tubular acidosis</td>
<td></td>
</tr>
<tr>
<td>Interstitial nephritis</td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td></td>
</tr>
<tr>
<td>Peripheral arthritis</td>
<td></td>
</tr>
<tr>
<td>Polymyositis</td>
<td></td>
</tr>
<tr>
<td>Hematologic</td>
<td></td>
</tr>
<tr>
<td>Lymphoma‡</td>
<td></td>
</tr>
</tbody>
</table>

*25% to 50%, but decreasing.
†Due to mechanical compression by inflamed parotid tissue.
‡Poor prognostic indicator.

Table 113-6  Similarities and Differences between Diffuse Infiltrative Lymphocytosis Syndrome (DILS) and Sjögren’s Syndrome

<table>
<thead>
<tr>
<th>Feature</th>
<th>DILS</th>
<th>Sjögren’s Syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parotid swelling</td>
<td>Ubiquitous</td>
<td>Uncommon</td>
</tr>
<tr>
<td>Sicca symptoms</td>
<td>Common</td>
<td>Very common</td>
</tr>
<tr>
<td>Extraglandular symptoms</td>
<td>Rare</td>
<td>Uncommon</td>
</tr>
<tr>
<td>Autoantibodies (antinuclear antibodies, anti-Ro/La)</td>
<td>DRB1<em>1102, DRB1</em>1301, DRB1*1302</td>
<td>Common</td>
</tr>
<tr>
<td>HLA class II association</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 113-5  Minor salivary gland biopsy specimen from a patient with diffuse infiltrative lymphocytosis syndrome. Note the relative preservation of the glandular architecture, even with significant interstitial inflammation.

Figure 113-6  Massive bilateral asymmetric salivary gland enlargement in a patient with diffuse infiltrative lymphocytosis syndrome. This was the presenting feature of human immunodeficiency virus infection in this patient. Computed tomography revealed this to be a solid mass. At follow-up 2 years after this photograph was taken, the gland had not changed in size.
headaches, and neurologic symptoms are common in HIV-positive patients, and triggers of vasculitis range from specific infectious agents and drugs to idiopathic causes. Among infective causes, cytomegalovirus and tuberculosis are probably the most common. Inflammatory vasculitides are less common rheumatologic diseases that occur in less than 1% of HIV patients.

One series found 34 (23%) of 148 “symptomatic” HIV-positive patients to have vasculitis. Of these patients, 11 met American College of Rheumatology criteria for a distinct category of vasculitis, including hypersensitivity vasculitis in 6, polyarteritis nodosa in 4, and Henoch-Schönlein purpura in 1. Another series of 98 Chinese patients found that 20% had vasculitis, including 15 with Behçet’s-like disease, 2 cases of Henoch-Schönlein purpura, 2 cases of digital gangrene, and 1 case of central nervous system vasculitis.

Granulomatosis with polyangiitis (formerly Wegener’s granulomatosis) and pulmonary microscopic polyangiitis can occur in patients with high CD4+ counts and during immune reconstitution. Churg-Strauss vasculitis has also been described. Behçet’s syndrome and relapsing polychondritis occur in HIV infection and respond to HAART. Rapidly progressive focal necrotizing vasculitis of the aorta and large arteries with aneurysm formation and rupture has been described in Africans with HIV infection. Giant cell arteritis likewise has been described in patients with HIV infection with aortic root dilation. Kawasaki disease has been reported in HIV-positive children and adults. Cryoglobulinemic vasculitis with associated lymphocytic interstitial pneumonia occurs with and without hepatitis C co-infection.

Patients with isolated central nervous system angiitis usually present with organic brain syndromes and neurologic deficit. In children and in one case report in an adult, HIV-associated cerebral aneurysmal arteriopathy was described as causing multiple fusiform aneurysms in the circle of Willis. Central nervous system vasculitis may manifest as recurrent strokes. Although imaging studies (MRI, angiography) may be helpful, brain biopsy may be necessary to establish the diagnosis. Necrotizing granulomatous vasculitis not limited to the central nervous system has been reported in patients with low CD4+ counts that

---

**Table 113-7** Treatment of Diffuse Infiltrative Lymphocytosis Syndrome

<table>
<thead>
<tr>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reassurance and education</td>
</tr>
<tr>
<td>Regular dental care</td>
</tr>
<tr>
<td>No specific treatment for asymptomatic individuals</td>
</tr>
<tr>
<td>Effective antiretroviral treatment</td>
</tr>
<tr>
<td>Pilocarpine or cevimeline for sicca symptoms</td>
</tr>
<tr>
<td>Systemic glucocorticoids</td>
</tr>
<tr>
<td>Drainage and instillation of corticosteroids into parotid lymphoepithelial cysts</td>
</tr>
<tr>
<td>Radiation of parotid cysts</td>
</tr>
</tbody>
</table>

---

**Figure 113-7** The “snowman” sign. Gallium-67 scintigraphy of the parotid glands of a patient with diffuse infiltrative lymphocytosis syndrome occurring in the setting of hemophilia.
respond to antiretroviral therapy. More recently, a case report has described leukocytoclastic cerebral vasculitis treated with anti-CD25 antibody.

Diagnosis is based on a high degree of suspicion and on angiography and biopsy of specific organ beds. Similar to immunocompetent patients, perinuclear antineutrophil cytoplasmic antigen (pANCA) and cytoplasmic antineutrophil cytoplasmic antigen (cANCA) may be useful with granulomatosis with polyangiitis or microscopic polyangiitis. However, biopsy with cultures is important to rule out infectious mimics.

Corticosteroids are the mainstay of treatment for HIV-associated vasculitis, although cytotoxic agents such as cyclophosphamide, intravenous immunoglobulin, and plasmapheresis have been used in refractory cases. Painful neuropathy secondary to vasculitis responds well to high-dose glucocorticoids, in contrast to HIV-associated peripheral neuropathy.

**PRIMARY PULMONARY HYPERTENSION**

Pulmonary hypertension is a severe life-limiting disease that affects younger patients. Patients with AIDS and primary pulmonary hypertension present with a higher degree of pulmonary hypertension than non-AIDS patients. The predominant histopathologic finding has been a plexogenic pulmonary arteriopathy, although thromboembolic changes also have been reported. One group found an association with HLA-DRB1*1301 and HLA-DRB1*1302 and with the linked allele HLA-DRB3*0301.

Symptoms include progressive shortness of breath, pedal edema, nonproductive cough, fatigue, syncope or near-syncpe, and chest pain. Pulmonary function tests show mild restrictive patterns with variably reduced diffusing capacities. In a review of 131 cases of pulmonary hypertension associated with HIV infection, the interval between the diagnosis of HIV disease and the diagnosis of pulmonary hypertension was 33 months. The median length of time from diagnosis to death was 6 months. Responses to vasodilator agents—calcium channel blockers, sildenafil, intravenous and inhaled prostanoids, and endothelin vasodilator agents—calcium channel blockers, sildenafil, intravenous and inhaled prostanoids, and endothelin antagonists—and to HAART vary, and some studies show improved mortality.

**HIV-ASSOCIATED MUSCULOSKELETAL INFECTION**

**Pyomyositis**

Pyomyositis is a primary infection of skeletal muscle that does not arise from contiguous infection; it is presumably hematogenous in origin and often is associated with abscess formation. Rarely seen in developed countries, infectious myositis nonetheless is an important complication of HIV infection in areas most endemic for HIV, such as Africa and India. It tends to occur in later stages of the infection, with CD4+ counts less than 200/µL. *Staphylococcus aureus* is the most common pathogen. Other organisms that have been implicated include *Streptococcus pyogenes*, *Cryptococcus neoformans*, *Mycobacterium tuberculosis*, *Mycobacterium avium-intracellulare*, *Nocardia asteroides*, *Salmonella enteritidis*, *Escherichia coli*, *Citrobacter freundii*, *Morganella morganii*, *Pseudomonas aeruginosa*, and group A streptococci.

The clinical course of pyomyositis can be roughly divided into three stages: invasive, suppurative, and late. The first stage, which typically lasts 1 to 3 weeks, is characterized by localized cramp-like pain and induration in conjunction with a low-grade fever. Large muscle groups, particularly those of the lower extremities, are most often affected. The degrees of pain and fever increase in the second stage, which is characterized further by the development of edema and pus in the affected muscle. Untreated, the disease progresses to the third stage; within 3 weeks of onset, sepsis and death can occur. The mortality rate associated with pyomyositis has been estimated to range from 1% to 20%. Ultrasound and MRI with contrast enhancement are effective in localizing infection, although sometimes, tagged white blood cell scans may be needed. Oral and intravenous antibiotics in conjunction with surgical drainage are often required.

**Bacterial Arthritis and Osteomyelitis**

No data suggest that bacterial infection of bones or joints occurs more frequently in patients with HIV infection. *S. aureus* is the most common infectious agent encountered, but parenteral drug use and not HIV infection per se may account for this. Many other organisms have been reported to cause osteomyelitis in HIV-infected patients, including *Mycobacterium tuberculosis*, *Salmonella*, *Nocardia asteroides*, *Streptococcus pneumoniae*, *Neisseria gonorrhoeae*, cytomegalovirus, invasive Aspergillus, *Toxoplasma gondii*, *Tourolipsis glabrata*, *Cryptococcus neoformans*, and *Coccidioides immitis*. Osteomyelitis is associated with mortality rates of greater than 20% in HIV-infected patients. The most frequently involved bones are the wrist, tibia, femoral heads, and thoracic cage, but other rare sites, such as the patella and the mandible, have been reported.

**Musculoskeletal Tuberculosis**

Musculoskeletal involvement, the fourth most common extrapulmonary manifestation of tuberculosis, is found in about 1% to 5% of patients with tuberculosis. It can mimic many skeletal diseases and can manifest in various locations. Less than 50% of reported patients with musculoskeletal tuberculosis have radiographic evidence of pulmonary tuberculosis. *M. tuberculosis* disseminates hematogenously after an acute or reactivated pulmonary infection. Usually, skeletal tuberculosis lesions in immunocompetent patients are solitary, but in AIDS patients, they may have a multicentric distribution in about 30% of cases. The vertebrae are the most common site involved, mostly the lower thoracic or upper lumbar segments. The frequency of tuberculous spondylitis is 50% to 66%, peripheral arthritis 20% to 30%, osteomyelitis 10% to 20%, and tenosynovitis and bursitis about 1% to 3%. Treatment includes four-drug antitubercular therapy and often surgical intervention.

**Atypical Mycobacterial Infection**

Musculoskeletal infection caused by atypical mycobacterial species is unusual in immunocompetent individuals.
Atypical mycobacterial species most commonly implicated in causing septic arthritis or osteomyelitis in HIV include M. avium-intracellulare complex, Mycobacterium kansasi, Mycobacterium haemophilum, Mycobacterium terrae, and Mycobacterium fortuitum. M. haemophilum has been most frequently implicated in skeletal infection, accounting for more than half of cases, and M. kansasi is second, accounting for an additional 25%. These are systemic infections that have involved several joints or skeletal sites. Cutaneous lesions, such as nodules, ulcers, and draining sinus tracts, occur in approximately 50% of patients. These infections tend to occur late in the course of HIV, usually when the CD4 T lymphocyte count is less than 100/µL. M. avium-intracellulare complex osteomyelitis has also been associated with immune reconstitution inflammatory syndrome after initiation of HAART. Along with standard antituberculosis therapy, clarithromycin is effective.

**Fungal Infections**

In addition to bacterial infection, patients with advanced HIV infection (CD4 T lymphocyte count <100/µL) are at high risk for fungal musculoskeletal infections, particularly infections caused by Candida albicans and Sporothrix schenckii. S. schenckii can manifest with oligoarticular or even polyarticular involvement and tenosynovitis, and can be particularly difficult to eradicate, requiring long-term suppressive antifungal therapy. Various disseminated fungal infections, such as histoplasmosis, cryptococcosis, and blastomycosis, occur in HIV and often cause osteomyelitis.

**Parasitic Infections**

Muscle toxoplasmosis is found in profoundly immunodepressed patients, typically presenting with a painful subacute myopathy and concurrent multisystemic toxoplasmosis. Toxoplasma cysts are observed mainly in muscle fibers in muscle biopsy specimens, and identification of cysts in muscle biopsy specimens against T. gondii, including pyrimethamine and sulfadiazine or trisulfapyrimidines.

**RESPONSE OF OTHER RHEUMATIC DISEASES TO HIV INFECTION**

Early reports suggested that rheumatoid arthritis went into remission in the face of HIV infection. Early reports likewise suggested that HIV infection might reduce the activity of bacillary angiomatosis after doxycycline or erythromycin therapy occurs, although bone lesions may need surgical drainage.
systemic lupus erythematosus, particularly at times of low CD4+ T cell counts. With emergence of the immune reconstitution syndrome, most autoimmune diseases appear de novo or recur with institution of HAART and increased CD4+ counts.1,102

HAART-RELATED IMMUNE RECONSTITUTION SYNDROME

Coverage of antiretroviral therapy has increased from 7% in 2003 to 20% in 2005. Immune reconstitution inflammatory syndrome (IRIS) is a paradoxical clinical deterioration that occurs in patients with HIV who receive HAART as a result of improvement in cellular immunity. It was initially described with recrudescence of infections, but now, autoimmune and autoimmune phenomena are being described. A meta-analysis of more than 13,000 HIV patients starting HAART found that 13% developed IRIS.103 The mean time of onset of IRIS from the time of HAART initiation is approximately 9 months.102

Shelburne and co-workers104 put forth four criteria required for the diagnosis of immune reconstitution inflammatory syndrome, as follows:

1. Patient has been diagnosed with AIDS.
2. Treatment with anti-HIV therapy results in increased CD4+ counts and decreased HIV viral load.
3. Infectious and inflammatory symptoms appear during therapy.
4. Symptoms cannot be explained by a new cause.

Improved understanding of the immunology of IRIS has helped elucidate how atypical, hyperaccentuated inflammatory host responses to pre-existing or coexisting infection can occur in HIV patients taking HAART. HIV infection causes a relentless decline in CD4+ memory and naïve cells, an increase in activated T cells in the peripheral blood, and thymic dysfunction. HAART can lead to sustained suppression of HIV and concomitant repopulation of T cell counts in a biphasic mode.105 The first phase represents the release of predominantly memory CD4+ cells and lasts a few weeks to months. The second phase, from approximately 6 months on, represents the main phase of naïve T cell proliferation and is accompanied by changes in T helper cytokine production profiles.106,107 IRIS can occur during both phases of immune recovery, and different infections and autoimmune phenomena occur in phase 1 compared with phase 2.

Organ-specific autoimmune phenomena have been described more often than generalized systemic autoimmune disease and tend to occur later during reconstitution. These phenomena may be a manifestation of naïve T cell release as opposed to memory T cell reconstitution. Graves' thyroiditis occurring about 21 months after initiation of HAART has been described in about 17 cases.108 Terminal ileitis, alopecia universalis, cerebral CD8+ lymphocytosis, and Guillain-Barré syndrome have been reported.109,110 Polymyositis,111 rheumatoid arthritis,102 systemic lupus erythematosus,111 Kawasaki-like febrile illness,112 autoimmune hepatitis, adult Still's disease, and sarcoidosis113 newly developing after initiation of HAART also have been described. These conditions tend to occur earlier during reconstitution compared with organ-specific autoimmunity.

In addition, IRIS has been reported in the setting of tumor necrosis factor (TNF) blocker discontinuation, with cryptococcal pneumonia occurring in an HIV-positive patient after stopping adalimumab.113

If a diagnosis of IRIS is made, HAART is continued and most symptoms resolve with little or no therapy. If inflammatory symptoms involve areas where significant damage secondary to uncontrolled inflammation is likely to occur, such as in the central nervous system or eye, HAART should be stopped, and careful use of corticosteroids should be considered. IRIS is less likely to occur if the CD4+ count is greater than 200/µL when HAART is initiated. Other risk factors for IRIS include being HAART naïve and the presence of a high antigenic burden with opportunistic infection when HAART is begun.114,115 Continued systematic analysis is needed, however, and guidelines need to be established for defining the autoimmunity associated with immune reconstitution. The prognosis for most IRIS cases is favorable because a robust inflammatory response may predict an excellent response to HAART in terms of immune reconstitution and, perhaps, improved survival.

RHEUMATOLOGIC COMPLICATIONS OF HIV TREATMENT

The myopathy associated with nucleoside transcriptase inhibitors such as zidovudine and the osteonecrosis and parotid lipomatosis associated with the use of protease inhibitors have been discussed previously. In addition to these conditions, cases of adhesive capsulitis, Dupuytren's contracture, tenosynovitis, and temporomandibular joint dysfunction have been reported as a consequence of indinavir treatment.116

LABORATORY ABNORMALITIES ASSOCIATED WITH HIV INFECTION

Humoral immunologic abnormalities are frequent in patients with HIV but are rarely associated with severe clinical signs. The most common laboratory abnormality is polyclonal hyperglobulinemia, found in 45% of HIV-positive individuals.117 Rheumatoid factor and antinuclear antibodies, usually in low titer, have been described in 17% of patients with HIV infection in some series,117 although anti-dsDNA antibodies and hypocomplementemia are rare. IgG anticardiolipin antibodies are found in 95% of untreated patients with AIDS, particularly in patients with advanced disease, and overall in 20% to 30% of HIV-positive individuals.118 They are rarely associated with thrombotic events. cANCAs and pANCAs have been described in the serum of HIV-positive individuals, as have anti–glomerular basement membrane antibodies.119 Cryoglobulinemia is decreasing in this population since the introduction of HAART.120 Indeed with HAART, many of these serologic abnormalities tend to decrease or resolve.

CONCLUSION

The impact of the global HIV pandemic continues to grow, and rheumatologists need to be aware of the wide spectrum of rheumatic diseases that occur in HIV-positive patients.
HAART has changed the natural history of HIV infection. It has modified the frequency and expression of some HIV-related clinical syndromes and has been associated directly (toxicity) and indirectly (immune reconstitution) with the development of new ones. With longer survival and newer refinements in treatment, the spectrum of rheumatic disease seen in HIV-positive patients is very much a moving target for rheumatologists and is likely to continue to evolve.

References

1863

RHEUMATIC MANIFESTATIONS OF HUMAN IMMUNODEFICIENCY VIRUS INFECTION


The references for this chapter can also be found on www.expertconsult.com.
Viral Arthritis

STANLEY J. NAIDES

KEY POINTS

<table>
<thead>
<tr>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute-onset, symmetric polyarthritis can be caused by viral infection, especially when accompanied by rash.</td>
<td></td>
</tr>
<tr>
<td>Always take exposure, travel, occupation, and vaccination histories.</td>
<td></td>
</tr>
<tr>
<td>Parvovirus B19 is the most common viral arthritis in the United States.</td>
<td></td>
</tr>
<tr>
<td>In adults with parvovirus B19 infection, rash may be subtle or absent.</td>
<td></td>
</tr>
<tr>
<td>Rubella arthritis occurs in young adults. Rubella vaccination has reduced the overall incidence of rubella infection but has shifted the peak age to young adults.</td>
<td></td>
</tr>
<tr>
<td>Arthralgia, arthritis, or neuropathic pain may occur after rubella vaccination; these conditions are usually self-limited in duration.</td>
<td></td>
</tr>
<tr>
<td>Alphaviruses are mosquito-borne causes of arthritis and rash. Outbreaks occur in endemic areas associated with rising mosquito populations and should be considered in travelers entering the United States.</td>
<td></td>
</tr>
<tr>
<td>Hepatitis B virus infection presents as an arthritis-urticaria syndrome.</td>
<td></td>
</tr>
<tr>
<td>Hepatitis C virus infection causes cryoglobulinemia and vasculitis. Cryoglobulinemic vasculitis often presents as palpable purpura of the lower legs.</td>
<td></td>
</tr>
<tr>
<td>The history of risk behaviors associated with hepatitis C virus infection may be remote.</td>
<td></td>
</tr>
</tbody>
</table>

Viruses are candidate causative agents for various rheumatic diseases in part because arthralgia and arthritis are prominent features of certain viral infections. Understanding how viruses cause arthritis and the nature of virus–host cell interactions may suggest how viruses precipitate, establish, or maintain chronic inflammatory arthritis such as rheumatoid arthritis.

Viral effects in a given host may depend on host factors such as age, gender, genetic background, infection history, and immune response. The ability of a given virus to infect a host may also depend on the viral mode of host entry, tissue tropism, replication strategy, cytopathologic effects, ability to establish persistent infection, viral expression of host-like antigens, and ability to alter host antigens. Viral modification of the regulation of cellular gene expression may contribute to autoimmunity. Infected cells may die by classic cell necrosis, programmed cell death (apoptosis), or autophagy. Initiation of an immune response to virally encoded antigens on the cell surface may target that cell for destruction and alter cell-cell interactions. The antibody response may generate immune complexes that are deposited locally at the site of viral infection or systemically in synovium. Alternatively, cells may survive, but their behavior may be altered by the expression of viral genes. Transactivation of cellular genes by viral gene products may induce the cell cycle or cytokines that elicit or perpetuate an immune response targeting host cells. Molecular mimicry of host autoantigens by viral proteins may break immune tolerance. Viral components may elicit “danger signals” that trigger an immune response.12

PARVOVIRUS B19

Human parvovirus B19 is a member of the family Paroviridae, subfamily Paroviridae, genus Erythroivirus. It consists of the small, single-stranded DNA viruses that autonomously replicate in erythroid precursors (hence the genus name). B19 has no envelope and is approximately 23 nm in diameter. Productive infection occurs in erythroid precursors; infection of nonerythroid tissues occurs but is restricted, which means that if assembly of virions occurs, it is inefficient, or that nonstructural but not capsid structural viral genes are expressed, preventing virion assembly. Parvoviruses are species specific and are not known to readily cross species barriers. The common canine parvovirus does not infect humans.

Epidemiology

B19 infection is common and occurs worldwide. B19 typically is transmitted by respiratory secretions but may also be transmitted via pooled blood products. Outbreaks commonly occur in late winter and spring, when close contact is most common, although epidemics may also occur in summer and fall. Most B19 infections, especially in children, remain asymptomatic or are diagnosed as nonspecific viral illnesses. Outbreaks tend to occur in 3- to 5-year cycles, representing the time required for a new cohort of susceptible children to enter school. Up to 60% of adults have serologic evidence of past B19 infection.14 Susceptible adults in occupations with multiple exposures to children, such as schoolteachers and pediatric nurses, are at greatest risk (up to 50%) of acquiring infection during outbreaks.15 Sporadic cases do occur during nonepidemic periods. The diagnosis should be entertained even in the absence of surveillance data suggesting an outbreak.

Pathogenesis

The onset of joint symptoms and rash is associated temporally with the appearance of serum anti-B19
immunoglobulin (Ig)M antibody, suggesting a role for circulating immune complexes during the acute phase of the illness. 6 Although little evidence of circulating virus has been found in patients who have chronic joint symptoms, B19 DNA may be found in the bone marrow and synovium of patients with chronic B19 arthropathy. Persistence in chronic B19 arthropathy may be facilitated by failure to develop IgG antibodies to the N-terminal region of the minor capsid protein VP1, known to encode neutralizing epitopes. 7 The presence of antibody to the B19 nonstructural protein NS1 in some cases of chronic B19 arthropathy probably reflects the immune response to NS1 on the surface of B19 virions or NS1 spilled during cell death. 7 NS1 protein itself, however, may play a pathogenic role in perpetuating chronic B19 arthropathy through its interaction with cellular genes. 8 NS1 protein upregulates in vitro transcription from the interleukin (IL)-6 promoter and from human immunodeficiency virus (HIV) long terminal repeats in the presence of tat and an intact tar element. 9,10 A high prevalence of B19 DNA and proteins in synovium from rheumatoid arthritis patients was reported in association with enhanced synovial production of IL-6 and tumor necrosis factor. 8 These findings remain controversial. 11 B19 may induce apoptosis through NS1, which is known to be toxic to cells. 12,13 Production of NS1 in nonpermissive synoviocytes could theoretically induce autoimmunity by disrupting normal patterns of cell interactions and intercellular regulation.

Diagnosis

Clinical Features

The incubation period from B19 infection to symptom onset is 7 to 18 days. B19 causes transient aplastic crisis in the setting of chronic hemolytic anemia. 6 In otherwise healthy children, B19 causes erythema infectiosum, or fifth disease, characterized by bright red “slapped cheeks” and a macular or maculopapular eruption on the torso and extremities. Up to 70% of infected children may be asymptomatic; others may have mild flu-like symptoms, including fever, headache, sore throat, cough, anorexia, vomiting, diarrhea, and arthralgia. In adults, the rash tends to be subtler, and the slapped-cheek rash is usually absent. Uncommon dermatologic manifestations include vesicular or hemorrhagic vesiculopustular eruptions, purpura with or without thrombocytopenia, Henoch-Schönlein purpura, and a “socks and gloves” acral erythema. B19 infection may be associated with paresthesias in the fingers and, rarely, with numbness of the toes. Progressive arm weakness has been associated with mild nerve conduction slowing and decreased motor and sensory potential amplitudes. B19 may cross the placenta to infect the fetus, which may develop hydrops fetalis on the basis of B19-induced anemia or viral cardiomyopathy. Less commonly, B19 may cause pancytopenia, isolated anemia, thrombocytopenia, leukopenia, myocarditis, neuropa thy, or hepatitis. 14 Reports suggest that B19 may be associated with vasculitis, including giant cell arteritis. 15,16

Patients with congenital or acquired immunodeficiencies, including prior chemotherapy or acquired immunodeficiency syndrome (AIDS) due to HIV infection, may develop persistent B19 infection with chronic or recurrent anemia, thrombocytopenia, or leukopenia. B19 infection is the leading cause of pure red cell aplasia in patients with AIDS. 17,18

In a study of an erythema infectiosum outbreak in Port Angeles, Washington, in which subjects were identified on the basis of rash, the incidence of arthralgia and joint swelling increased with age (Table 114-1). 19 In adults, a severe flu-like illness consisting of fever, chills, malaise, and myalgias may precede or accompany a sudden-onset, moderately severe, symmetric polyarthritis in a rheumatoid-like distribution. The arthritis is characterized by prominent involvement of the finger proximal interphalangeal, metacarpophalangeal, wrist, knee, and ankle joints. Within 24 to 48 hours of onset, all affected joints become involved. Axial skeleton involvement is uncommon. Joint symptoms are usually self-limited.

After the initial infection, objective joint swelling, heat, and erythema, when present, tend to resolve over several weeks. A minority of patients have prolonged symptoms that fall into one of two patterns. Approximately two-thirds have continuous morning stiffness and arthralgias with intermittent flares. The remaining patients are symptom free between flares. Chronic B19 arthropathy may last months to years. Pain remains a prominent feature during flares; patients commonly report morning stiffness. Approximately 12% of patients presenting with “early synovitis” have B19 infection; most are women. 6

Table 114-1  Prevalence of Joint Symptoms in Fifth Disease by Age: Port Angeles, Washington, 1961-1962

<table>
<thead>
<tr>
<th>Symptom</th>
<th>0-9 yr</th>
<th>10-19 yr</th>
<th>&gt;20 yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td>5.1</td>
<td>11.5</td>
<td>77.2</td>
</tr>
<tr>
<td>Swelling</td>
<td>2.8</td>
<td>5.3</td>
<td>59.6</td>
</tr>
</tbody>
</table>


Laboratory Tests

Viremia lasts 5 to 6 days and is associated with an absence of reticulocytosis and, in otherwise normal individuals, a minimal decrease in the concentrations of hemoglobin, neutrophils, and lymphocytes. Flu-like symptoms may occur during viremia. An IgM antibody response follows the initial viremia in 4 to 6 days and is associated with clearing of viremia and cessation of nasal shedding of virus.

The antibody response is associated with the second phase of clinical illness, characterized by rash and joint symptoms. Onset of the anti-B19 IgG antibody response occurs almost concurrently with the IgM response. The two clinical phases of illness often overlap. Low to moderate titers of rheumatoid factor and anti-DNA, antilymphocyte, antinuclear, and antiphospholipid antibodies may be present initially. 20,24

During viremia, immune electron microscopy may detect virions in serum. However, this method is not readily available to clinicians. B19 DNA may be detected during viremia. However, because adult patients usually present...
after the onset of joint symptoms, the most useful diagnostic test is anti-B19 IgM serology. Radioimmunoassays and enzyme-linked immunosorbent assays have been used to detect B19 antigen and specific antibody to B19 capsid. The anti-B19 IgM antibody response is usually positive for 2 months after the acute illness and may wane shortly thereafter. In some patients, anti-B19 IgM may be detected for 6 months or longer. A positive anti-B19 IgG antibody test in the absence of anti-B19 IgM usually is not diagnostically helpful because of the high seroprevalence of anti-B19 IgG in the adult population. Reports of B19 DNA in normal synovium suggest that testing for B19 DNA in these tissues is of little clinical utility in the absence of anti-B19 IgM.27

Differential Diagnosis

Many patients with B19 arthropathy meet the American Rheumatism Association criteria for a diagnosis of rheumatoid arthritis: morning stiffness lasting longer than an hour; symmetric involvement; involvement of at least three joints; and involvement of the finger proximal interphalangeal, metacarpophalangeal, and wrist joints. Rheumatoid factor may be present at low to moderate titers. Absence of both rheumatoid nodules and joint destruction differentiate B19 arthropathy from classic erosive rheumatoid arthritis.

Occasionally, B19 infection may present with features of systemic lupus erythematosus (SLE). Whether this represents a clinical mimic or indicates that B19 plays a role in initiating or precipitating SLE in these patients remains to be determined.6

Rubella in adults may present with rash and symmetric polyarthralgia or polyarthritis that is clinically indistinguishable from B19 infection. A history of prenatal rubella testing, prior rubella vaccination, or rubella exposure may aid the clinician in choosing the appropriate diagnostic serologies.

Treatment and Outcome

No specific treatment or vaccine has been identified for B19 infection. Therefore, treatment is symptomatic and consists of nonsteroidal anti-inflammatory drugs. Intravenous immunoglobulin has been successful in the treatment of bone marrow suppression and B19 persistence in immunocompromised patients, but initial studies suggest that this is not applicable to patients with chronic arthropathy. Long-term prognosis is good. Although subjective arthralgias and morning stiffness may be prolonged, joint destruction is not a feature of chronic B19 arthropathy. The role of B19 as a cofactor in the development of classic erosive rheumatoid arthritis has not been confirmed.

TOGAVIRUSES

The family Togaviridae includes the Rubivirus and Alphavirus genera.

Rubella Virus

Rubella virus is the sole member of the genus Rubivirus. It consists of enveloped, single-stranded RNA viruses. The rubella virion is spherical and measures 50 to 70 nm in diameter, with a 30-nm dense core. Envelope glycoproteins form 5- to 6-nm spike-like projections that contain hemagglutination activity.28
Two rheumatologic syndromes may complicate natural infection or vaccination in children. In the catcher’s crouch syndrome, a lumbar radiculoneuropathy causes popliteal fossa pain on arising in the morning. Exacerbation of the pain by knee extension encourages the assumption of a baseball catcher’s crouch position. The pain gradually subsides through the day but recurs the next morning. In the arm syndrome, brachial neuropathy causes arm and hand pain and dysesthesias that are worse at night. Both syndromes may occur beginning 1 to 2 months after infection or vaccination, with the initial episode lasting up to 2 months. Episodes recur for up to 1 year but eventually resolve without long-term sequelae.33

Laboratory Tests. Although rubella may be cultured from tissues and body fluids, including throat swabs, detecting antirubella IgM antibody usually establishes the diagnosis of acute rubella infection. Diagnosis by anti-IgG antibody seroconversion requires paired acute and convalescent sera. IgM and IgG are usually present at the onset of joint symptoms. IgM antibody levels peak 8 to 21 days after symptom onset and wane by 5 weeks. Antirubella IgG rises rapidly over a period of 1 to 3 weeks and is long lived. A single positive IgG serum sample or a set of untitered IgG-positive screens documents only immunity.29

Differential Diagnosis. Rubella arthritis needs to be differentiated from other viral arthritides and from inflammatory arthritides, including rheumatoid arthritis. It may be confused with parvovirus B19 infection.

Treatment and Outcome

Nonsteroidal anti-inflammatory drugs are useful to control symptoms. Some investigators have suggested the use of low to moderate doses of steroids to control symptoms and viremia.40 Long-term prognosis is good.

Alphaviruses

Members of the genus Alphavirus are enveloped, single-stranded RNA viruses transmitted by mosquitoes.38 Several cause acute febrile arthropathy, and their names reflect local appreciation of their clinical impact. For example, chikungunya means “that which twists or bends up” (Tanzania). The related o’nyong-nyong virus means “joint breaker” in the Acholi (Uganda) dialect. Igbo-ora is “the disease that breaks your wings.”

Epidemiology

Chikungunya, o’nyong-nyong, and igbo-ora viruses form a serologically related group. Chikungunya virus was isolated during an epidemic of febrile arthritis in Tanzania between 1952 and 1953. Similar epidemics probably occurred in Africa, Asia, India, Indonesia, and possibly the southern United States as early as 1779.35 Mosquitoes responsible for transmission to humans define its geographic distribution (Table 114-2). A feared consequence of global warming is spread of the geographic range of infected mosquitoes.36,39

Chikungunya fever occurs endemic and in epidemics.40 Outbreaks have been described in the Indian Ocean islands, Malaysia, and Hong Kong.41-44 An outbreak occurred in Italy in 2007. A large-scale outbreak of the serologically related o’nyong-nyong virus occurred in the Acholi province of northwestern Uganda in February 1959; this outbreak spread through Uganda and the surrounding region at a rate of 2 to 3 kilometers daily, affecting more than 2 million people within 2 years.45 After the initial o’nyong-nyong epidemic, clinical disease was not detected again until it re-emerged in the Acholi region in 1996.46 Despite the absence of outbreaks in the intervening years, serologic surveys have demonstrated that o’nyong-nyong virus is endemic.47 Weber’s line is a hypothetical demarcation separating the Australian and Asiatic geographic zones. Antibodies to chikungunya virus are found west of Weber’s line and to the east of it.48

Weber’s line is a hypothetical demarcation separating the Australian and Asiatic geographic zones. Antibodies to chikungunya virus are found west of Weber’s line, and Ross River virus antibodies are found only east of it. Ross River virus causes epidemics of fever and rash in Australia, New Zealand, and the western Pacific islands.49 In the Fiji Islands from 1979 to 1980, Ross River virus caused febrile polyarthritis in more than 40,000 individuals.49 In Australia, endemic cases and epidemics occur in tropical and

<table>
<thead>
<tr>
<th>Virus</th>
<th>Reservoir</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chikungunya virus</td>
<td>Aedes species</td>
<td>Africa, Asia</td>
</tr>
<tr>
<td></td>
<td>Mansonia africana</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles funestus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles gambiae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes aegypti</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes species</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles species</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Culex species</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Culex annulirostris</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mansonia uniformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes polynesiensis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Baboons, monkeys, bat species</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scotophilus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>O’nyong-nyong virus</td>
<td>Aedes species</td>
<td>Asia</td>
</tr>
<tr>
<td></td>
<td>Mansonia africana</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles funestus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles gambiae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes vigilax</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes camptorhynchus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Culex annulirostris</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mansonia uniformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes polynesiensis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rodents, marsupials, domestic animals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Igbo-ora virus</td>
<td>Aedes species</td>
<td>Africa</td>
</tr>
<tr>
<td></td>
<td>Mansonia africana</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles funestus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles gambiae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes vigilax</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes camptorhynchus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Culex annulirostris</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mansonia uniformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes polynesiensis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rodents, marsupials, domestic animals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Ross River virus</td>
<td>Aedes species</td>
<td>Pacific Islands</td>
</tr>
<tr>
<td></td>
<td>Mansonia africana</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles funestus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles gambiae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes vigilax</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes camptorhynchus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Culex annulirostris</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mansonia uniformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes polynesiensis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rodents, marsupials, domestic animals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Barmah forest virus</td>
<td>Aedes species</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Mansonia africana</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles funestus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles gambiae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes vigilax</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes camptorhynchus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Culex annulirostris</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mansonia uniformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes polynesiensis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rodents, marsupials, domestic animals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Sindbis virus</td>
<td>Aedes species</td>
<td>Sweden, Finland, Karelian</td>
</tr>
<tr>
<td></td>
<td>Mansonia africana</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles funestus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles gambiae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes vigilax</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes camptorhynchus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Culex annulirostris</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mansonia uniformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes polynesiensis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rodents, marsupials, domestic animals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Mayaro virus</td>
<td>Haemagogus janthinomys</td>
<td>Bolivia, Brazil, Peru</td>
</tr>
<tr>
<td></td>
<td>Mansonia africana</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles funestus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles gambiae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes vigilax</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes camptorhynchus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Culex annulirostris</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mansonia uniformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aedes polynesiensis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rodents, marsupials, domestic animals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
</tbody>
</table>
temperate regions annually. Most cases occur in Queensland and New South Wales territories, where high rainfall and subsequent increases in mosquito populations usually precede epidemic periods. Infection rates in Australia range from 0.2% to 3.5% per year. Male and female infection rates are similar, but a female predominance has been noted in presenting cases. Most infected adults are symptomatic; the case rate for children is lower. Barmah Forest virus, another alphavirus with an increasing incidence in Australia, may manifest in a fashion similar to Ross River virus.51,56

Individuals involved in outdoor activities or occupations in forested areas in Sweden, Finland, and the neighboring Karelian isthmus of Russia are at greatest risk for infection with Sindbis virus; in those regions, it is known as Okelbo disease, Pogosta disease, and Karelian fever, respectively. Birds are the intermediate host. It has also been reported in central Africa, Zimbabwe, South Africa, and Australia in sporadic cases or small outbreaks.35

Mayaro virus, first recognized in Trinidad in 1954, is endemic in the tropical rain forests of Bolivia, Brazil, and Peru. Cases have been imported into the United States in individuals traveling from endemic areas.58

Diagnosis

Clinical Features. Chikungunya fever presents with an explosive onset of high fever and severe arthralgia after a 1- to 12-day incubation period. The fever lasts 1 to 7 days. Typically, a macular or maculopapular, sometimes pruritic, rash on the torso, extremities, and occasionally the face, palms, and soles occurs on day 2 to 5 of illness as the patient defervesces. The rash may last 1 to 5 days and may recur with fever. Isolated petechiae and mucosal bleeding may occur. In some patients, involved skin desquamates.99,60

Chemosis is prominent. Headache, photophobia, retroorbital pain, pharyngitis, anorexia, nausea, vomiting, and abdominal pain may be present. Diffuse myalgia and back and shoulder pain are common. Migratory polyarthralgia, stiffness, and swelling affect predominantly the small joints of the hands, wrists, feet, and ankles. Large joints are less severely affected. Previously injured joints may be disproportionately affected. Large effusions are uncommon. Symptoms in children tend to be milder. Low-titer rheumatoid factor may be found in those with long-standing symptoms.

O’nyong-nyong fever is clinically similar to chikungunya fever.61,62 In 1984, igbo-ora caused an epidemic of fever, myalgias, arthralgias, and rash in four Ivory Coast villages. Sequelae of isolates from the 1996 outbreak of o’nyong-nyong fever suggested that igbo-ora virus is a variant of o’nyong-nyong virus.46

Ross River virus polyarthralgia is severe, incapacitating, and often migratory and asymmetric.63 Symptoms follow a 7- to 11-day incubation period. Finger interphalangeal and metacarpophalangeal joints, wrists, knees, ankles, shoulders, elbows, and toes are often involved. Polyarticular swelling and tenosynovitis are common. Arthralgias are worse in the morning and after inactivity. Rash is macular, papular, or maculopapular and may be pruritic. Vesicles, papules, or petechiae are typically seen on the trunk and extremities. The palms, soles, and face may be involved. Rash typically appears 1 to 2 days before joint symptoms, but it may occur anywhere from 11 days before to 15 days after the onset of arthralgias, and it resolves by fading to a brownish discoloration or by desquamation. Half of patients have no fever, and those who do may have only modest fever lasting 1 to 3 days. Nausea, headache, and myalgia are common. Respiratory symptoms, mild photophobia, and lymphadenopathy may occur. Up to a third of patients have paresthesias and palm or sole pain. Carpal tunnel syndrome may be seen. Arthritis is less common and less prominent in Barmah Forest virus infection than in Ross River virus infection, but the rash is more common and florid.64,65

Rash and arthralgia are the presenting symptoms in Sindbis virus infection, although one may precede the other by a few days. Constitutional symptoms are usually mild and include low-grade fever, headache, fatigue, malaise, nausea, vomiting, pharyngitis, and paresthesias. A macular rash typically begins on the torso and then spreads to arms and legs, palms, soles, and occasionally the head. Macules evolve to form papules that tend to vesiculate. Vesiculation is prominent on pressure points, including the palms and soles. As the rash fades, a brownish discoloration is left. Vesicles on the palms and soles may become hemorrhagic. The rash may recur during convalescence.66

A Mayaro virus outbreak in Belterra, Brazil, in 1988 was characterized by sudden onset of fever, headache, dizziness, chills, and arthralgias in the wrists, fingers, ankles, and toes. The clinical attack rate was 80%. Joint swelling, unilateral inguinal lymphadenopathy, and leukopenia may be present. A maculopapular rash on the trunk and extremities lasts about 3 days.47

Laboratory Tests. The diagnosis of alphavirus infection requires laboratory confirmation. Any febrile patient residing in or returning from an endemic area should undergo a laboratory investigation. Chikungunya virus may be isolated from serum on days 2 through 4 of illness.60 Neutralizing antibody, hemagglutination inhibition activity, and complement fixation tests may be used to detect antibodies. Chikungunya virus–specific IgM antibodies may be found for 6 months or longer.60 O’nyong-nyong virus may be isolated by intracerebral injection into suckling mice, in which it produces alopecia, rash, and running. Hemagglutination inhibition or complement fixation tests identify o’nyong-nyong virus.20,71 Because chikungunya and o’nyong-nyong viruses are closely related serologically, mouse antisera raised to chikungunya virus or o’nyong-nyong virus react equally well with o’nyong-nyong virus, but o’nyong-nyong antisera do not react well with chikungunya virus. Molecular detection methods have improved diagnostic specificity.23,75 Specific reverse transcriptase polymerase chain reaction–based assays have been developed for viral RNA detection.73,76

In chikungunya fever, synovial fluid shows decreased viscosity, poor mucin clot, and 2000 to 5000 white blood cells/mm³. Ross River virus has been isolated only from antibody-negative sera. In Australian epidemics before 1979, patients were antibody positive at the time of presentation. In contrast, patients during the Pacific island epidemics of 1979 to 1980 remained viremic and seronegative for up to 1 week after the onset of symptoms. Synovial fluid cell counts range from 1500 to 13,800 cells/mm³, predominantly monocytes and vacuolated macrophages.77 Barmah Forest virus infection is confirmed by rising titers of specific
infected patients have persistent arthralgias for months. Arthritis may persist for as long as 5 to 6 years. A smaller number reported in a few patients for up to 3 years. Sodes gradually resolve, but joint symptoms have been at least 6 months in up to half of patients. Joint symptoms may recur. Residual polyarthralgia may be present. Joint symptoms may be able to resume their daily activities within 4 weeks, although tend to improve joint symptoms. Half of all patients are ruled out.

Pathogenesis

Little is known about the pathogenesis of chikungunya fever or arthritis. Involved skin shows erythrocyte extravasation from superficial capillaries and perivascular cuffing. The virus adsorbs to human platelets, causing aggregation, suggesting a mechanism for bleeding. Synovitis probably results from direct viral infection of synovium. In one patient with chronic arthropathy, the synovium appeared atrophic on arthroscopy and was histologically normal. The mechanisms of o’nyong-nyong virus pathogenesis are unknown. However, the virus was isolated from peripheral blood mononuclear cells in a patient in Chad.

Ross River virus antigen may be detected early in monocytes and macrophages by immunofluorescence, but intact virus is not identifiable by electron microscopy or cell culture. Erythematous and purpuric rashes show mild dermal perivascular mononuclear cell infiltrates, mostly T lymphocytes. Purpuric areas also show erythrocyte extravasation. Viral antigen may be detected in epithelial cells in erythematous and purpuric skin lesions and in perivascular zones in erythematous lesions.

Sindbis virus has been isolated from a skin vesicle in the absence of viremia. Skin lesions show perivascular edema, hemorrhage, lymphocytic infiltrates, and areas of necrosis. Anti–Sindbis virus IgM may persist for years, raising the possibility that Sindbis virus arthritis is associated with viral persistence.

Treatment and Outcome

Management is supportive. Nonsteroidal anti-inflammatory agents are useful, but aspirin should be avoided in view of the tendency for alphavirus rashes to develop a hemorrhagic component. Chloroquine has been used in chikungunya fever when nonsteroidal anti-inflammatory agents failed.

During the acute attack, range-of-motion exercises may decrease stiffness. In general, management of alphavirus infection is symptomatic; patients recover without sequelae. After acute chikungunya fever, symptoms may persist for months before resolution. Approximately 10% of patients still have joint symptoms 1 year after infection.

A few patients may develop chronic arthralgia. Case reports suggest that a few patients with chronic arthropathy develop destructive joint lesions, but a second process cannot be ruled out.

For persons with Ross River virus arthritis, mild exercise tends to improve joint symptoms. Half of all patients are able to resume their daily activities within weeks, although residual polyarthralgia may be present. Joint symptoms may recur.

Arthralgia, myalgia, and lethargy may continue for at least 6 months in up to half of patients. Relapsing episodes gradually resolve, but joint symptoms have been reported in a few patients for up to 3 years.

Nonerosive chronic arthropathy is common after Sindbis virus infection, with up to one-third of patients having arthropathy 2 years or longer after onset. A smaller number have symptoms for as long as 5 to 6 years. Mayaro virus–infected patients have persistent arthralgias for months.

HEPATITIS B VIRUS

Hepatitis B virus (HBV), a member of the family Hepadnaviridae, genus Orthohepadnavirus, is an enveloped, double-stranded, icosahedral DNA virus measuring 42 nm in diameter.

Epidemiology

HBV occurs worldwide and is transmitted by parenteral and sexual routes. Prevalence is highest in Asia, the Middle East, and sub-Saharan Africa. In China, the prevalence is as high as 10%, compared with 0.01% in the United States. In endemic regions, infection occurs at an early age, frequently perinatally. Early HBV infection is usually asymptomatic. Rates of HBV carriage and specific antibody positivity decline with age. In the West, most infections are acquired during adulthood through sexual or needle exposure, leading to acute hepatitis. Of those with hepatitis, 5% to 10% develop persistent infection. In endemic regions, HBV is a common cause of chronic liver disease and a leading cause of hepatocellular carcinoma.

Clinical Features

The time from infection to clinical hepatitis is usually 45 to 120 days. A preicteric prodromal period lasts several days to a month and may be associated with fever, myalgia, malaise, anorexia, nausea, and vomiting. Joint involvement is usually sudden in onset and often severe, with symmetric and simultaneous involvement of several joints. Alternatively, arthritis may be migratory or additive.

The joints of the hand and knee are most often affected, but wrists, ankles, elbows, shoulders, and other large joints may be involved as well. Fusiform swelling occurs in the small joints of the hand. Morning stiffness is common. Arthritis and urticaria may precede jaundice by days to weeks and may persist for several weeks, but they usually subside soon after the onset of clinical jaundice. Arthritis is usually limited to the preicteric prodrome. Those who develop chronic active hepatitis or chronic HBV viremia may have recurrent polyarthralgia or polyarthritis.

Polymyositis nodosa may be associated with chronic hepatitis B viremia.

Diagnosis

Urticaria in the presence of polyarthritis should suggest the possibility of HBV infection. Acute hepatitis may be asymptomatic, but elevated bilirubin and transaminases are usually present when arthritis appears. At the onset of arthritis, peak levels of serum hepatitis B surface antigen (HBsAg) are detectable. Virions, viral DNA, polymerase, and hepatitis B antigen may be detectable in serum. Anti–hepatitis B core antigen IgM antibodies indicate acute HBV infection rather than past or chronic infection.

Pathogenesis

Significant viremia occurs early in infection. Soluble immune complexes with circulating HBsAg form as anti-HBsAg antibodies are produced. An immune complex–mediated arthritis usually results, with immune complex
Liver biopsy for staging of liver disease is usually indicated in the absence of positive serologic findings. Methods in the absence of positive serologic findings. Detection of HCV RNA by polymerase chain reaction amplification allows confirmation of HCV viremia, viral load, and genotypic confirmation. A recombinant antigen strip immunoblot assay is widely used for serodiagnosis. Serologic tests use an array of antigens in an enzyme immunoassay. Diagnosis is confirmed by polymerase chain reaction amplification methods in the absence of positive serologic findings. Liver biopsy for staging of liver disease is usually indicated in patients who have serum anti-HCV antibody or RNA, even in the setting of normal liver enzymes, because liver enzymes do not reflect liver histology. A number of algorithms based on blood measures of liver involvement have been proposed to aid in staging.

Pathogenesis

HCV infection persists despite antibody response to viral epitopes. Increased CD4+CD25+ regulatory T lymphocytes may blunt the immune response to HCV. A high rate of mutation in the envelope protein is responsible for the emergence of neutralization-escape mutants and quasispecies. HCV may contain an IgG Fc binding region on its surface; humoral immune response to HCV would, by epitope spreading, also target bound immunoglobulin Fc structures. Chronic HCV infection leads to cirrhosis, end-stage liver failure, and hepatocellular carcinoma after a period of up to 20 years, but the frequency of these sequelae is debated, and the mechanisms by which they occur are unknown.

Treatment

Initially, interferon-alpha2b was used to suppress viral titers and ameliorate HCV liver disease in about half of patients and may have benefited HCV-associated cryoglobulinemia. Relapse after completion of the initial course of therapy was common. The current use of pegylated interferons that increase drug half-life and decrease clearance, and the addition of ribavirin, have improved outcomes. Controversy continues regarding whether interferon therapy precipitates autoimmune diseases such as autoimmune thyroiditis.

Those with cryoglobulinemia who fail interferon therapy require immunosuppressive therapy when vasculitis is present.

HUMAN T-LYMPHOTROPIC VIRUS TYPE 1

Human T-lymphotropic virus type 1 (HTLV-1), a retrovirus, is endemic in southern Japan, where it has been associated with oligoarthritis and a nodular rash (Figure 114-1). Anti-HTLV serology is positive. Type C viral particles are found in skin nodules. Synovial tissue is infiltrated by leukemic T lymphocytes with lobulated nuclei.

OTHER VIRUSES

Joint involvement occasionally occurs in numerous other commonly encountered viral syndromes. Children with varicella rarely develop brief monoarticular or pauciarticular arthritis. Mumps in adults is occasionally associated with small or large joint synovitis preceding or following the onset of parotitis by up to 4 weeks. Mumps arthritis may last several weeks. Infection with adenovirus and with coxsackieviruses A9, B2, B3, B4, and B6 has been associated with recurrent episodes of polyarthritis, pleuritis, myalgia, rash, pharyngitis, myocarditis, and leukocytosis. Epstein-Barr virus—induced mononucleosis is frequently accompanied by polyarthralgia, but monoarticular knee arthritis sometimes occurs. A few cases of polyarthritides, fever, and myalgia due to echovirus 9 infection have been reported.
Arthritis associated with herpes simplex virus or cytomegalovirus infection is rare, but a severe cytomegalovirus polyarthritis has been described in several immunocompromised bone marrow transplant recipients. Herpes hominis occasionally causes arthritis of the knee in wrestlers, a condition referred to as herpes gladiatorum. Knee arthritis has been reported as a rare complication after vaccinia inoculation.

Selected References

59. Moore CG:
63. Moore CG:
63. Moore CG:
70. Moore CG:
72. Moore CG:
74. Moore CG:
76. Moore CG:


Full references for this chapter can be found on www.expertconsult.com.


133. Burns LJ, Gingrich RD: Cytomegalovirus infection presenting as polyarticular arthritis following autologous BMT, Bone Marrow Transplant 11:77, 1993.
Acute rheumatic fever (ARF) is a delayed, nonsuppurative sequel of a pharyngeal infection with group A streptococci. Although a dramatic decline in the severity and the mortality of the disease has been noted, reports have described its resurgence in the United States.

Adequate treatment of documented streptococcal pharyngitis markedly reduces the incidence of subsequent ARF. Appropriate antimicrobial prophylaxis prevents recurrence of disease in known patients with ARF.

The clinical presentation of ARF varies. The lack of a single pathognomonic feature led to the development of the revised Jones criteria, which should be used to establish a diagnosis.

The terms migrating and migratory are often used to describe the polyarthritis of ARF, but these designations do not signify that the inflammation disappears in one joint when it appears in another. Rather, the various localizations usually overlap in time, and the onset, as opposed to the full course of arthritis, "migrates" from joint to joint.

Many investigators have suggested that poststreptococcal migratory arthritis (in adults and children) in the absence of carditis might be an entity distinct from ARF. Although these features may be seen (admittedly rarely), migratory arthritis without evidence of other major Jones criteria but supported by two minor manifestations still should be considered ARF, especially in children.

Antibiotic prophylaxis with penicillin should be started immediately after resolution of the acute episode. The optimal regimen consists of oral penicillin VK, 250,000 U twice a day, or parenteral penicillin G, 1.2 million U intramuscularly every 4 weeks.

Pathogenetic features are often used to describe the arthritis, "migrates" from joint to joint. The terms migrating and migratory are often used to describe the polyarthritis of ARF, but these designations do not signify that the inflammation disappears in one joint when it appears in another. Rather, the various localizations usually overlap in time, and the onset, as opposed to the full course of arthritis, "migrates" from joint to joint.

Many investigators have suggested that poststreptococcal migratory arthritis (in adults and children) in the absence of carditis might be an entity distinct from ARF. Although these features may be seen (admittedly rarely), migratory arthritis without evidence of other major Jones criteria but supported by two minor manifestations still should be considered ARF, especially in children.

Antibiotic prophylaxis with penicillin should be started immediately after resolution of the acute episode. The optimal regimen consists of oral penicillin VK, 250,000 U twice a day, or parenteral penicillin G, 1.2 million U intramuscularly every 4 weeks.

Acute rheumatic fever (ARF) is a delayed, nonsuppurative sequel of a pharyngeal infection with group A streptococci. After the initial streptococcal pharyngitis, a latent period of 2 to 3 weeks occurs. The onset of disease usually is characterized by an acute febrile illness, which may manifest in one of three classic ways: (1) The patient may present with migratory arthritis predominantly involving the large joints of the body; (2) concomitant clinical and laboratory signs of carditis and valvulitis may be noted; or (3) involvement of the central nervous system may manifest as Sydenham's chorea. The clinical episodes are self-limiting, but damage to the valves may be chronic and progressive, resulting in cardiac decompensation and death.

Although a dramatic decline in severity and mortality of the disease has been observed since the turn of the 20th century, reports in recent years have described its resurgence in the United States and in many military installations throughout the world—a reminder that the disease remains a public health problem even in developed countries. In addition, the disease continues essentially unabated in many developing countries. Estimates suggest that 10 to 20 million new cases will be reported per year in countries where two-thirds of the world population lives. For all of these reasons, it is important to keep rheumatic fever in the differential diagnosis of acute febrile illnesses, as well as acute inflammatory arthritis, in both children and adults.

**Epidemiology**

The incidence of ARF began to decline long before the introduction of antibiotics into clinical practice, decreasing from 250 to 100 patients per 100,000 population from 1862 to 1962 in Denmark. The introduction of antibiotics in 1950 rapidly accelerated this decline, until by 1980, the incidence ranged from 0.23 to 1.88 patients per 100,000, with disease occurring primarily in children and teenagers. A notable exception has been in the native Hawaiian and Maori populations (both of Polynesian ancestry), among whom the incidence continues to be 13.4 per 100,000 hospitalized children per year.

Only a few M serotypes (types 5, 14, 18, and 24) have been identified with outbreaks of ARF, suggesting that certain strains of group A streptococci may be more “rheumatogenic” than others. In Trinidad, types 41 and 11 have been the most common strains isolated from the oropharynx of patients with ARF. In our own series, conducted over a 20-year period (Table 115-1), many different M serotypes were isolated, including six strains that could not be typed. Kaplan and colleagues isolated several M types from patients seen during an outbreak of ARF in Utah, and these strains were mucoid and nonmucoid in character. Whether or not certain strains are more “rheumatogenic” than others remains unresolved. What is true, however, is that a streptococcal strain capable of causing well-documented pharyngitis is generally capable of causing ARF, although some notable exceptions have been recorded.

**Pathogenesis**

Although little evidence suggests the direct involvement of group A streptococci in the affected tissues of ARF patients, a large body of epidemiologic and immunologic evidence indirectly implicates group A streptococci in initiation of the disease process: (1) It is well known that outbreaks of ARF closely follow epidemics of streptococcal sore throat or scarlet fever; (2) adequate treatment of documented
streptococcal pharyngitis markedly reduces the incidence of subsequent ARF; (3) appropriate antimicrobial prophylaxis prevents recurrence of disease in known patients with ARF; and (4) if one tests the sera of ARF patients for three antistreptococcal antibodies (streptolysin O, hyaluronidase, and streptokinase), most ARF patients (whether or not they recall an antecedent streptococcal sore throat) are found to have elevated antibody titers to these antigens.

A note of caution is necessary concerning documentation (clinical or microbiologic) of an antecedent streptococcal infection. The frequency of isolation of group A streptococci from the oropharynx is extremely low, even in populations with limited access to antibiotics. An age-related discrepancy in the clinical documentation of an antecedent sore throat has been noted. In older children and young adults, the recollection of a streptococcal sore throat approaches 70%; in younger children, this rate approaches only 20%. This is important to have a high index of suspicion of ARF in children or young adults presenting with signs of arthritis or carditis or both, even in the absence of a clinically documented sore throat.

Another intriguing, and as yet unexplained, observation has been the invariable association of ARF only with streptococcal pharyngitis. Although many outbreaks of impetigo have occurred, ARF almost never occurs after infection with these strains. In Trinidad, where impetigo and ARF are common infections, the strains colonizing the skin are different from those associated with ARF, and this does not influence the incidence of ARF. The explanation for these observations remains obscure.

Group A streptococci fall into two main classes based on differences in the C-repeat regions of the M protein. One class is associated with streptococcal pharyngeal infection, and the other (with some exceptions) is commonly associated with impetigo. The particular strain of streptococci may be crucial in initiating the disease process. The pharyngeal site of infection with its large repository of lymphoid tissue also may be important in initiation of the abnormal humoral response by host to antigens cross-reactive with target organs. Finally, although impetigo strains do colonize the pharynx, they do not seem to elicit as strong an immunologic response to the M protein moiety as do the pharyngeal strains. This may prove to be an important factor, especially in light of known cross-reactions between various streptococcal structures and mammalian proteins.

### GROUP A STREPTOCOCCI

Figure 115-1 shows a schematic cross-section of group A streptococci. The capsule is composed of equimolar concentrations of N-acetyl glucosamine and glucuronic acid and is structurally identical to hyaluronic acid of mammalian tissues. Although numerous attempts to produce antibodies to this capsule have been unsuccessful, Fillet and colleagues were able to show high antibody titers to hyaluronic acid using techniques designed to detect nonprecipitating antibodies in the sera of immunized animals. Similar antibodies have been noted in humans. Data establishing the importance of this capsule in human infection have been almost nonexistent, although Stollerman commented on the presence of a large mucoid capsule as one of the more important characteristics of certain “rheumatogenic” strains.

With respect to the M protein moiety, investigations by Lancefield and others spanning almost 70 years have established that the M protein molecule (at least 80 distinct serologic types) is perhaps the most important virulence factor in group A streptococcal infection of humans. The protein is a helical, coiled-coil structure that bears a striking structural homology to the cardiac cytoskeletal proteins, tropomyosin and myosin, and to many other coiled-coil structures, including keratin, DNA, lamin, and vimentin. When the amino acid sequence of many M proteins was delineated, it was possible to localize specifically the cross-reactive areas of the molecules. Studies of Dale and Beachey showed that the segment of the M protein involved in the opsonic reaction cross-reacted with human sarcolemma antigens. Sargent and co-workers more precisely localized this cross-reaction to the M protein amino acid residues 164 through 197.

Evidence implicating these cross-reactions in the pathogenesis of ARF remains scant. Antibodies to myosin have been failures in the treatment of rheumatic fever. Some investigators have noted an increased frequency of antibody formation to cardiac myosin in patients with ARF, and others have shown a direct correlation between the frequency of such antibodies and the severity of disease. Nevertheless, these observations remain controversial.

### Table 115-1 Positive Throat Cultures for Group A β-Hemolytic Streptococci among Rockefeller University Hospital Rheumatic Fever Patients (n = 87)

<table>
<thead>
<tr>
<th>M Type</th>
<th>RHD</th>
<th>No RHD</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nontypable</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL</td>
<td>9</td>
<td>14</td>
<td>23</td>
</tr>
</tbody>
</table>

No RHD, patients without rheumatic heart disease; RHD, patients with rheumatic heart disease.

![Figure 115-1 Schematic representation of various structures of group A streptococci. Note the wide variety of cross-reactions between antigens and mammalian tissues.](image-url)
The relationship of cell wall mucopeptides to the pathogenesis of ARF remains obscure. Elevated levels of antimucopeptide antibody have been detected not only in the sera of patients with ARF, but also in the sera of patients with rheumatoid arthritis and juvenile rheumatoid arthritis; however, its pathogenetic relationship to clinical disease has been difficult to establish. No evidence indicates that cell wall antigens are present in the Aschoff lesion or in the myocardial tissue obtained from patients with ARF. Perhaps the most significant cross-reactions lie in the streptococcal membrane structure. We have shown that immunization with membrane material elicited antibodies that were bound to heart sections in a pattern similar to that observed with acute RF sera (Figure 115-2).

Kingston and Glynn were the first to show that animals immunized with streptococcal antigens developed antibodies in their sera that stained astrocytes. Husby and associates showed that sera from ARF patients with chorea exhibited antibodies that were specific for caudate cells. Absorption of the sera with streptococcal membrane antigens eliminated reactivity with caudate cells. Numerous other cross-reactions between streptococcal membranes and other organs have been reported (e.g., renal basement membranes, basement membrane proteoglycans, skin [particularly keratin]). In the context of this chapter, space does not permit an exhaustive discussion of these cross-reactions, and the reader is referred to other studies for a more detailed discussion. Whether or not these cross-reactions (especially the cross-reactions seen with basement membranes and skin) play a role in the disease awaits further study.

GENETICS

The concept that ARF might be the result of a host genetic predisposition has intrigued investigators for more than a century. It has been variously suggested that the disease gene is transmitted in an autosomal dominant fashion or been detected in the sera of ARF patients, but they also are present in a high percentage of the sera obtained from individuals who had a streptococcal infection but did not subsequently develop ARF. The significance of this observation is unclear because myosin is an internal protein of cardiac muscle cells and is not easily exposed to M protein cross-reacting antibodies. The group-specific carbohydrate of the streptococcus is a polysaccharide chain consisting of repeating units of rhamnose capped by N-acetyl glucosamine molecules. The N-acetyl glucosamine is immunodominant and gives rise to the serologic group specificity of group A streptococci.

Goldstein and associates first described the cross-reaction between group A carbohydrate and valvular glycoproteins, and this reactivity was related to the N-acetyl glucosamine moiety present in both structures. Goldstein and Caravano noted that rheumatic fever (RF) sera reacted to the heart valve glycoprotein. Fillet (unpublished data) observed strong reactivity of RF sera with purified proteoglycan material. These cross-reactions could involve the sugar moiety present in the proteoglycan portion of the glycoprotein and the carbohydrate.

It generally has been assumed that group A anticarbohydrate antibodies do not play a role in phagocytosis of group A streptococci. Salvadori and co-workers showed, however, that human sera containing high titers of anti–group A carbohydrate antibody promoted opsonization and phagocytosis of many different M protein–specific strains, and the opsonophagocytic antibodies were directed to the N-acetyl glucosamine moiety of the group A carbohydrate. The mucopeptide portion of the cell wall is the “backbone” of the organism and is quite rigid in structure. It is composed of repeating units of muramic acid and N-acetyl glucosamine, cross-linked by peptide bridges. It is particularly difficult to degrade and induces a wide variety of lesions when injected into various species, including arthritis in rats and myocardial granulomas in mice resembling (but not identical to) RF Aschoff lesions.

The group-specific carbohydrate of the streptococcus is a polysaccharide chain consisting of repeating units of rhamnose capped by N-acetyl glucosamine molecules. The N-acetyl glucosamine is immunodominant and gives rise to the serologic group specificity of group A streptococci. Goldstein and associates first described the cross-reaction between group A carbohydrate and valvular glycoproteins, and this reactivity was related to the N-acetyl glucosamine moiety present in both structures. Goldstein and Caravano noted that rheumatic fever (RF) sera reacted to the heart valve glycoprotein. Fillet (unpublished data) observed strong reactivity of RF sera with purified proteoglycan material. These cross-reactions could involve the sugar moiety present in the proteoglycan portion of the glycoprotein and the carbohydrate.

It generally has been assumed that group A anticarbohydrate antibodies do not play a role in phagocytosis of group A streptococci. Salvadori and co-workers showed, however, that human sera containing high titers of anti–group A carbohydrate antibody promoted opsonization and phagocytosis of many different M protein–specific strains, and the opsonophagocytic antibodies were directed to the N-acetyl glucosamine moiety of the group A carbohydrate. The mucopeptide portion of the cell wall is the “backbone” of the organism and is quite rigid in structure. It is composed of repeating units of muramic acid and N-acetyl glucosamine, cross-linked by peptide bridges. It is particularly difficult to degrade and induces a wide variety of lesions when injected into various species, including arthritis in rats and myocardial granulomas in mice resembling (but not identical to) RF Aschoff lesions.

The relationship of cell wall mucopeptides to the pathogenesis of ARF remains obscure. Elevated levels of antimucopeptide antibody have been detected not only in the sera of patients with ARF, but also in the sera of patients with rheumatoid arthritis and juvenile rheumatoid arthritis; however, its pathogenetic relationship to clinical disease has been difficult to establish. No evidence indicates that cell wall antigens are present in the Aschoff lesion or in the myocardial tissue obtained from patients with ARF. Perhaps the most significant cross-reactions lie in the streptococcal membrane structure. We have shown that immunization with membrane material elicited antibodies that were bound to heart sections in a pattern similar to that observed with acute RF sera (Figure 115-2).

Kingston and Glynn were the first to show that animals immunized with streptococcal antigens developed antibodies in their sera that stained astrocytes. Husby and associates showed that sera from ARF patients with chorea exhibited antibodies that were specific for caudate cells. Absorption of the sera with streptococcal membrane antigens eliminated reactivity with caudate cells. Numerous other cross-reactions between streptococcal membranes and other organs have been reported (e.g., renal basement membranes, basement membrane proteoglycans, skin [particularly keratin]). In the context of this chapter, space does not permit an exhaustive discussion of these cross-reactions, and the reader is referred to other studies for a more detailed discussion. Whether or not these cross-reactions (especially the cross-reactions seen with basement membranes and skin) play a role in the disease awaits further study.

GENETICS

The concept that ARF might be the result of a host genetic predisposition has intrigued investigators for more than a century. It has been variously suggested that the disease gene is transmitted in an autosomal dominant fashion or been detected in the sera of ARF patients, but they also are present in a high percentage of the sera obtained from individuals who had a streptococcal infection but did not subsequently develop ARF. The significance of this observation is unclear because myosin is an internal protein of cardiac muscle cells and is not easily exposed to M protein cross-reacting antibodies. The group-specific carbohydrate of the streptococcus is a polysaccharide chain consisting of repeating units of rhamnose capped by N-acetyl glucosamine molecules. The N-acetyl glucosamine is immunodominant and gives rise to the serologic group specificity of group A streptococci.

Goldstein and associates first described the cross-reaction between group A carbohydrate and valvular glycoproteins, and this reactivity was related to the N-acetyl glucosamine moiety present in both structures. Goldstein and Caravano noted that rheumatic fever (RF) sera reacted to the heart valve glycoprotein. Fillet (unpublished data) observed strong reactivity of RF sera with purified proteoglycan material. These cross-reactions could involve the sugar moiety present in the proteoglycan portion of the glycoprotein and the carbohydrate.

It generally has been assumed that group A anticarbohydrate antibodies do not play a role in phagocytosis of group A streptococci. Salvadori and co-workers showed, however, that human sera containing high titers of anti–group A carbohydrate antibody promoted opsonization and phagocytosis of many different M protein–specific strains, and the opsonophagocytic antibodies were directed to the N-acetyl glucosamine moiety of the group A carbohydrate. The mucopeptide portion of the cell wall is the “backbone” of the organism and is quite rigid in structure. It is composed of repeating units of muramic acid and N-acetyl glucosamine, cross-linked by peptide bridges. It is particularly difficult to degrade and induces a wide variety of lesions when injected into various species, including arthritis in rats and myocardial granulomas in mice resembling (but not identical to) RF Aschoff lesions.

The relationship of cell wall mucopeptides to the pathogenesis of ARF remains obscure. Elevated levels of antimucopeptide antibody have been detected not only in the sera of patients with ARF, but also in the sera of patients with rheumatoid arthritis and juvenile rheumatoid arthritis; however, its pathogenetic relationship to clinical disease has been difficult to establish. No evidence indicates that cell wall antigens are present in the Aschoff lesion or in the myocardial tissue obtained from patients with ARF. Perhaps the most significant cross-reactions lie in the streptococcal membrane structure. We have shown that immunization with membrane material elicited antibodies that were bound to heart sections in a pattern similar to that observed with acute RF sera (Figure 115-2).

Kingston and Glynn were the first to show that animals immunized with streptococcal antigens developed antibodies in their sera that stained astrocytes. Husby and associates showed that sera from ARF patients with chorea exhibited antibodies that were specific for caudate cells. Absorption of the sera with streptococcal membrane antigens eliminated reactivity with caudate cells. Numerous other cross-reactions between streptococcal membranes and other organs have been reported (e.g., renal basement membranes, basement membrane proteoglycans, skin [particularly keratin]). In the context of this chapter, space does not permit an exhaustive discussion of these cross-reactions, and the reader is referred to other studies for a more detailed discussion. Whether or not these cross-reactions (especially the cross-reactions seen with basement membranes and skin) play a role in the disease awaits further study.

GENETICS

The concept that ARF might be the result of a host genetic predisposition has intrigued investigators for more than a century. It has been variously suggested that the disease gene is transmitted in an autosomal dominant fashion or
in an autosomal recessive fashion with limited penetrance, or that it is possibly related to the genes conferring blood group secretor status. Renewed interest in the genetics of ARF occurred with recognition that gene products of the human major histocompatibility complex (MHC) were associated with certain clinical disease states. Using an alloserum from a multiparous donor, an increased frequency of a B cell alloantigen was reported in several genetically distinct and ethnically diverse populations of ARF patients and was not MHC related. More recently, a monoclonal antibody (D8/17) was prepared by immunizing mice with B cells from an ARF patient. A B cell antigen identified by this antibody was found to be expressed on increased numbers of B cells in 100% of rheumatic patients of diverse ethnic origins, and in only 10% of normal individuals. The antigen defined by this monoclonal antibody showed no association with or linkage to any of the known MHC haplotypes, and it did not seem to be related to B cell activation antigens. Studies with D8/17 have been expanded to a larger number of patients with RF (see Table 115-1) of diverse ethnic origins with essentially the same results. As discussed subsequently, the presence or absence of elevated levels of D8/17+ B cells in cases of questionable RF has been helpful in establishing or ruling out the diagnosis.

These studies contrast with other reports in which an increased frequency of HLA-DR4 and HLA-DR2 has been seen in white and black patients with rheumatic heart disease (RHD). Other studies have implicated HLA-DR1 and HLA-DRW6 as susceptibility factors in South African black patients with RHD. Guilherme and associates have reported an increased frequency of HLA-DR7 and HLA-DW53 in RF patients in Brazil. These seemingly conflicting results concerning HLA antigens and RF susceptibility prompt speculation that these reported associations might be of class II genes close to (or in linkage disequilibrium with), but not identical to, the putative RF susceptibility gene. Alternatively, and more likely, susceptibility to ARF is polygenic, and the D8/17 antigen might be associated with only one of the genes (i.e., genes of the MHC complex encoding for DR antigens) conferring susceptibility. Although the explanation remains to be determined, the presence of the D8/17 antigen does seem to identify a population at special risk of contracting ARF (Table 115-2).

**ETILOGIC CONSIDERATIONS**

Although a large body of immunologic and epidemiologic evidence has implicated group A streptococci in the induction of the disease process, the precise pathologic mechanisms involved remain obscure. At least three main theories have been proposed. The first theory is concerned with the question of whether persistence of the organism is important. Despite several controversial reports, no investigators have been able to show consistently and reproducibly live organisms in RF cardiac tissues or valves.

The second theory revolves around the question of whether deposition of toxic products is required. Although an attractive hypothesis, little or no experimental evidence has been obtained to support this concept. Halbert and colleagues have suggested that streptolysin O (an extracellular product of group A streptococci) is cardiotoxic and might be carried to the site by circulating complexes containing streptolysin O and antibody. Despite an extensive search for these products, no such complexes in situ have been identified, however. Renewed interest in these extracellular toxins has emerged more recently with the observation by Schlievert and co-workers that certain streptococcal pyrogenic toxins (A and C) may act as superantigens. These antigens may stimulate large numbers of T cells through their unique bridging interaction with T cell receptors of specific Vβ types and class II MHC molecules. This interaction is distinct from conventional antigen presentation in the context of the MHC complex. When activated, these cells elaborate tumor necrosis factor, interferon-γ, and numerous interleukin moieties, contributing to the initiation of pathologic damage. It has been suggested that in certain disease states, such as rheumatoid arthritis, autoreactive cells of specific Vβ lineage may “home” to the target organ.

Although an attractive hypothesis, no data concerning the role of these superantigens in ARF have yet been forthcoming. Perhaps the best evidence to date favors a third theory of an abnormal host immune response (humoral and cellular) in a genetically susceptible individual to the streptococcal antigens cross-reactive with mammalian tissues. Evidence supporting this theory may be divided into three broad categories:

1. Employing a wide variety of methods, numerous investigators have documented the presence of heart-reactive antibodies in ARF sera. The prevalence of these antibodies has ranged from 33% to 85% in various series. Although these antibodies are seen in other individuals (notably individuals with uncomplicated streptococcal infections that do not progress to RF and patients with poststreptococcal glomerulonephritis), the titers are always lower than those seen in RF and decrease with time during the convalescent

---

**Table 115-2** Frequency of the D8/17 Marker in Patients with Rheumatic Fever, Patients with Other Diseases, and Controls in Various Geographic Populations

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>% Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rheumatic Fever Patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>43/45</td>
<td>93</td>
</tr>
<tr>
<td>New Mexico</td>
<td>30/31</td>
<td>97</td>
</tr>
<tr>
<td>Utah*</td>
<td>18/18</td>
<td>100</td>
</tr>
<tr>
<td>Russia (Georgia)</td>
<td>27/30</td>
<td>90</td>
</tr>
<tr>
<td>Russia (Moscow)</td>
<td>50/52</td>
<td>96</td>
</tr>
<tr>
<td>Mexico</td>
<td>35/39</td>
<td>89</td>
</tr>
<tr>
<td>Chile</td>
<td>45/50</td>
<td>90</td>
</tr>
<tr>
<td><strong>Normals</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russia</td>
<td>4/78</td>
<td>5</td>
</tr>
<tr>
<td>New York</td>
<td>6/68</td>
<td>8</td>
</tr>
<tr>
<td>Chile</td>
<td>8/50</td>
<td>16</td>
</tr>
<tr>
<td>Mexico</td>
<td>6/72</td>
<td>8</td>
</tr>
<tr>
<td><strong>Other Diseases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>2/42</td>
<td>4</td>
</tr>
<tr>
<td>Ischemic heart disease</td>
<td>0/10</td>
<td>0</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>1/25</td>
<td>4</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>1/12</td>
<td>9</td>
</tr>
</tbody>
</table>

*Acute patients.
period (Table 115-3). An important point in terms of diagnosis and prognosis has been the observation by Zabriskie and associates\(^2\) that these heart-reactive antibody titers decline over time. By the end of 3 years, these titers are essentially undetectable in patients who had only a single attack (Figure 115-3). This pattern is consistent with the well-known clinical observation that recurrences of RF most often occur within the first 2 to 3 years after the initial attack and become rarer 5 years after an initial episode.

As illustrated in Figure 115-4, this pattern of titers also has prognostic value. During the 2- to 5-year period after the initial attack, a patient’s titers decreased to undetectable levels. With a known break in prophylaxis starting in year 6, at least two streptococcal infections occurred, as evidenced by an increase in antistreptolysin O (ASO) titers during that period. The concomitant increase in heart-reactive antibody titers was notable. The final infection was followed by a clinical recurrence of classic rheumatic carditis complete with isolation of the organism, elevated heart-reactive antibodies, and acute phase reactants 11 months after the initial attack.

2. Sera from patients with ARF also contain increased levels of antibodies to myosin and tropomyosin compared with sera from patients with pharyngeal streptococcal infections that do not progress to ARF. These myosin affinity purified antibodies also cross-react with M protein moieties, suggesting that this molecule could be the antigenic stimulus for the production of myosin antibodies in these sera.\(^2,33\)

3. Finally, as indicated earlier, autoimmune antibodies are a prominent finding in chorea, another major clinical manifestation of ARF, and these antibodies are directed against the cells of the caudate nucleus. The titer of this antibody corresponds with clinical disease activity.\(^34\) Although not autoimmune in nature, the presence of elevated levels of immune complexes in ARF has been well documented in the sera and in the joints of ARF patients.\(^54\) Elevated levels of immune complexes, which may be as high as the levels seen in classic poststreptococcal glomerulonephritis, may be responsible for the immune complex vasculitis seen in ARF tissues and may provide the initial impetus for vascular damage, followed by the secondary penetration of autoreactive antibodies. Support for this concept is the close clinical similarity of RF arthritis to experimentally induced serum sickness in animals or the arthritis seen secondary to drug hypersensitivity.

Deposition of host immunoglobulin and complement also is seen in the cardiac tissues of ARF patients, suggesting autoimmune deposition of immunoglobulins in or near the Aschoff lesions. At a cellular level, ample evidence has been found for the presence of lymphocytes and macrophages at the site of pathologic damage in the heart in patients with ARF.\(^55\) The cells are predominantly CD4\(^+\) helper lymphocytes during acute stages of the disease (4:1). The ratio of CD4\(^+\) to CD8\(^+\) lymphocytes (2:1) more closely approximates the normal ratio in chronic valvular specimens. Most of these cells express DR antigens. A potentially important finding has been the observation that macrophage-like fibroblasts present in the diseased valves express DR antigens\(^36\) and might be the antigen-presenting cells for the CD4\(^+\) lymphocytes. Increased cellular reactivity to streptococcal antigens also has been noted in the peripheral blood mononuclear cell preparations of ARF patients compared with these cells isolated from nephritis patients.\(^57\)

This abnormal reactivity peaks at 6 months after the attack, but may persist for 2 years after the initial episode. The reactivity was specific only for the strains associated with ARF, suggesting an abnormal humoral and cellular response to streptococcal antigens unique to RF-associated streptococci. Support for the potential pathologic importance of these T cells is strengthened further by the observation that lymphocytes obtained from experimental animals

### Table 115-3 Heart-Reactive Antibody Titers in Sera of Patients with Acute Rheumatic Fever Compared with Uncomplicated Streptococcal Infections

<table>
<thead>
<tr>
<th>Clinical Disorder</th>
<th>No. Patients</th>
<th>Serum Dilutions</th>
<th>Average ASO Titer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute rheumatic fever (grade I)</td>
<td>34</td>
<td>4+ 2+ +*</td>
<td>700</td>
</tr>
<tr>
<td>Uncomplicated streptococcal infection (grade II)</td>
<td>40</td>
<td>1+ 0 0</td>
<td>561</td>
</tr>
<tr>
<td>APSGN</td>
<td>20</td>
<td>+/- 0 0</td>
<td>520</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>10</td>
<td>0 0 0</td>
<td>ND</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>10</td>
<td>0 0 0</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Serum samples obtained at onset of rheumatic fever and at a comparable time in the group with uncomplicated scarlet fever.
†Serum samples obtained during active disease.

APSGN, acute poststreptococcal glomerulonephritis; ASO, antistreptolysin O; ND, not determined.

![Figure 115-3](image-url) Serial heart-reactive antibody titers in 40 patients with documented acute rheumatic fever. Note the slow decline of these titers over the first 2 years after the initial episode and the absence of these antibodies 5 years after the initial attack.
sensitized to cell membranes, but not to cell walls, are specifically cytotoxic for syngeneic embryonic cardiac myofibers in vitro.58 In humans, normal mononuclear cells primed in vitro by M protein molecules from an RF-associated strain also are cytotoxic for myofibers, but specificity solely for cardiac cells was lacking in human studies.59 Similar studies have not been performed yet using lymphocytes from active ARF patients (Table 115-4).

**CLINICAL FEATURES**

The clinical presentation of ARF varies, and the lack of a single pathognomonic feature has resulted in the development of revised Jones criteria (Table 115-5),60 which are used to establish a diagnosis. These criteria were established only as guidelines for the diagnosis and were never intended to be “etched in stone.” Depending on the age, geographic location, and ethnic population, emphasis on one criterion for the diagnosis of ARF may be more important than others. Manifestations of RF that are not clearly expressed pose a dilemma because of the importance of identifying a first rheumatic attack definitively to establish the need for prophylaxis of recurrences (see later). Some of the isolated manifestations, particularly polyarthritis, may be difficult or impossible to distinguish from other diseases, especially at their onset. The diagnosis can be made, however, when “pure” chorea is the sole manifestation

---

**Table 115-4** Composition of Mononuclear Cellular Infiltrates in Acute and Chronic Active Rheumatic Valvulitis

<table>
<thead>
<tr>
<th>Patient</th>
<th>Type of Valve</th>
<th>Type of Valvulitis</th>
<th>HLA-DR</th>
<th>CD14+</th>
<th>CD20-</th>
<th>CD3-</th>
<th>CD4+</th>
<th>CD8+</th>
<th>CD4/CD8 Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Valvulitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Mitral</td>
<td>Acute</td>
<td>58.9</td>
<td>42.6</td>
<td>5.1</td>
<td>49.5</td>
<td>75.6</td>
<td>23.9</td>
<td>3.1</td>
</tr>
<tr>
<td>2</td>
<td>Mitral</td>
<td>Acute</td>
<td>49.8</td>
<td>43.1</td>
<td>6.9</td>
<td>43.1</td>
<td>58.7</td>
<td>34.3</td>
<td>1.9</td>
</tr>
<tr>
<td>3</td>
<td>Mitral</td>
<td>Acute</td>
<td>52.7</td>
<td>51</td>
<td>3.9</td>
<td>38.1</td>
<td>65.9</td>
<td>26.5</td>
<td>2.3</td>
</tr>
<tr>
<td>4</td>
<td>Aortic</td>
<td>Acute</td>
<td>63.9</td>
<td>42</td>
<td>5.5</td>
<td>52.4</td>
<td>75.4</td>
<td>18.9</td>
<td>4</td>
</tr>
<tr>
<td>Chronic Valvulitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Mitral</td>
<td>Chronic active</td>
<td>49.4</td>
<td>47.4</td>
<td>7.4</td>
<td>44.3</td>
<td>53.7</td>
<td>38.8</td>
<td>1.4</td>
</tr>
<tr>
<td>5</td>
<td>Mitral</td>
<td>Chronic active</td>
<td>48.8</td>
<td>39.1</td>
<td>1.4</td>
<td>53.9</td>
<td>45.2</td>
<td>51.5</td>
<td>0.9</td>
</tr>
<tr>
<td>6</td>
<td>Aortic</td>
<td>Chronic active</td>
<td>67.8</td>
<td>35</td>
<td>4</td>
<td>36.8</td>
<td>47.5</td>
<td>49.1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Mitral</td>
<td>Chronic active</td>
<td>41.8</td>
<td>23.4</td>
<td>8</td>
<td>65.9</td>
<td>57.3</td>
<td>33.3</td>
<td>1.7</td>
</tr>
<tr>
<td>8</td>
<td>Mitral</td>
<td>Chronic active</td>
<td>69.6</td>
<td>48.7</td>
<td>8.2</td>
<td>65.9</td>
<td>57.3</td>
<td>33.3</td>
<td>1.7</td>
</tr>
<tr>
<td>9</td>
<td>Mitral</td>
<td>Chronic active</td>
<td>55.4</td>
<td>24.2</td>
<td>8.1</td>
<td>59.8</td>
<td>64.9</td>
<td>24.7</td>
<td>2.6</td>
</tr>
<tr>
<td>10</td>
<td>Mitral</td>
<td>Chronic active</td>
<td>80.4</td>
<td>34.1</td>
<td>13.4</td>
<td>44.4</td>
<td>44.8</td>
<td>50.9</td>
<td>0.9</td>
</tr>
<tr>
<td>11</td>
<td>Mitral</td>
<td>Chronic active</td>
<td>46.1</td>
<td>29.6</td>
<td>0.8</td>
<td>65.6</td>
<td>61.6</td>
<td>33.3</td>
<td>1.8</td>
</tr>
</tbody>
</table>

* Determined in the frozen valve samples studied.
+ CD14+ monocytes/macrophages.
- B cells.
+ Pan T cells.
+ Helper T cells.
+ Suppressor cells.
Criteria

Two major or one major and

61

2004, World Health Organization.

consultation, From

Disease (RHD)

Rheumatic Fever (RF) and Rheumatic Heart

Organization Criteria for the Diagnosis of

Table 115-6

rent ARF in patients with established RHD and chronic

updated the Jones criteria to allow for the diagnosis of recur-

other cause. More recently, the World Health Organization

update: guidelines for diagnosis of rheu-


because of the rarity with which this syndrome is due to any

other cause. More recently, the World Health Organization

updated the Jones criteria to allow for the diagnosis of recur-

rent ARF in patients with established RHD and chronic

RHD (Table 115-6).

Table 115-6  Summary of 2002 World Health

Organization Criteria for the Diagnosis of

Rheumatic Fever (RF) and Rheumatic Heart

Disease (RHD)

<table>
<thead>
<tr>
<th>Diagnostic Categories</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary episode of RF</td>
<td>Two major or one major and two minor manifestations plus evidence of a preceding group A streptococcal infection</td>
</tr>
<tr>
<td>Recurrent attack of RF in patients without established RHD</td>
<td>Two major or one major and two minor manifestations plus evidence of a preceding group A streptococcal infection</td>
</tr>
<tr>
<td>Recurrent attack of RF in patients with established RHD</td>
<td>Two minor manifestations plus evidence of a preceding group A streptococcal infection</td>
</tr>
<tr>
<td>Rheumatic chorea, insidious onset of rheumatic carditis</td>
<td>One major manifestation or evidence of a preceding group A streptococcal infection</td>
</tr>
<tr>
<td>Chronic valve lesions of RHD (i.e., patients presenting for the first time with pure mitral stenosis, mixed mitral valve disease, and aortic valve disease)</td>
<td>No other criteria required for diagnosis of RHD</td>
</tr>
</tbody>
</table>

*The diagnosis is made if the patient has two major criteria OR one major and two minor criteria if the history includes a preceding streptococcal infection.


Arthritis

In classic, untreated cases, the arthritis of ARF affects several joints in quick succession, each for a short time. The legs usually are affected first, and the arms are affected later. The terms migrating and migratory are often used to describe the polyarthritis of ARF, but these designations are not meant to signify that the inflammation disappears in one joint when it appears in another. Rather, the various localizations usually overlap in time, and onset, as opposed to the full course of arthritis, “migrates” from joint to joint.

Joint involvement is more common, and also is more severe, in teenagers and young adults than in children. Arthritis is usually the earliest symptomatic manifestation of the disease, although asymptomatic carditis may precede it. Rheumatic polyarthritis may be excruciatingly painful but is almost always transient. The pain is usually more prominent than the objective signs of inflammation. When the disease is allowed to express itself fully, unmodified by anti-inflammatory treatment, more than half of patients studied show a true polyarthritis, with inflammation in 6 to 16 joints. Classically, each joint is maximally inflamed for only a few days, or a week at the most; the inflammation decreases, perhaps lingering for another week or so, and then disappears completely. Radiographs at this point may show a slight effusion but most likely are unremarkable.

In routine practice, many patients with arthritis or arthralgias are treated empirically with salicylates or other nonsteroidal anti-inflammatory drugs; arthritis subsides quickly in the joints already affected and does not migrate to new joints. Therapy may deprive the diagnostician of a useful sign. In a large series of patients with ARF and associated arthritis, most of whom had been treated, involvement of only a single large joint was common (25%). One or both knees were affected in 76%, and one or both ankles were affected in 50%. Elbows, wrists, hips, or small joints of the feet were involved in 12% to 15% of patients, and shoulders or small joints of the hand were affected in 7% to 8%. Rarely affected joints included the lumbosacral (2%), cervical (1%), sternooclavicular (0.5%), and temporomandibular (0.5%). Involvement of the small joints of the hands or feet alone occurred in only 1% of patients. Analysis of the synovial fluid in well-documented cases of ARF with arthritis generally reveals a sterile, inflammatory fluid. The complement components C1q, C3, and C4 may be decreased, indicating their consumption by immune complexes in the joint fluid.

Poststreptococcal Reactive Arthritis

Numerous investigators have suggested that poststreptococcal migratory arthritis (in adults and children) in the absence of carditis might be a distinct entity from ARF for the following reasons: (1) The latent period between the antecedent streptococcal infection and the onset of poststreptococcal reactive arthritis is shorter (1 to 2 weeks) than the 3 to 4 weeks usually seen in classic ARF; (2) the response of the poststreptococcal reactive arthritis to aspirin and other nonsteroidal medications is poor compared with the dramatic response seen in classic ARF; (3) evidence of carditis is not usually seen in these patients, and the severity of the arthritis is marked; and (4) extra-articular
manifestations (e.g., tenosynovitis, renal abnormalities) are often seen in these patients.

Although these features may be seen (admittedly rarely), migratory arthritis without evidence of other major Jones criteria, if supported by two minor manifestations (see Table 115-5), still must be considered ARF, especially in children. Variations in the response to aspirin in these children often are not documented with serum salicylate levels, and an unusual clinical course is insufficient to exclude the diagnosis of ARF. Appropriate prophylactic measures should be taken. Support for this concept may be found in the work of Crea and Mortimer. In their series of patients with ARF, 50% of the children who presented solely with signs of migratory arthritis went on to develop significant valvular damage. RF also occurs in adults. Although migratory arthritis is a common presenting symptom, an outbreak in San Diego Naval Training Camp revealed a 30% incidence of valvular damage in these patients. The importance of clearly defining this reactive arthritis as an RF variant has obvious implications for secondary prophylactic treatment. As suggested by some investigators, poststreptococcal reactive arthritis is a benign condition without the need for prophylaxis. Because most patients fulfill the Jones criteria (one major, two minor), they should be considered as having RF and, in our opinion, should be treated accordingly.

Carditis

Cardiac valvular and muscle damage can manifest in a variety of signs or symptoms. These manifestations include organic heart murmurs, cardiomegaly, congestive heart failure, and pericarditis. Mild to moderate chest discomfort, pleuritic chest pain, and a pericardial friction rub are indications of pericarditis. On clinical examination, the patient can have new or changing organic murmurs, most commonly mitral regurgitant murmurs and occasionally aortic regurgitant murmurs and systolic ejection murmurs, caused by acute valvular inflammation and deformity. Rarely, a Carey Coombs mid-diastolic murmur caused by rapid flow over the mitral valve is heard. If the valvular damage is severe, and concurrent cardiac dysfunction is present, congestive heart failure may occur. Congestive heart failure is the most life-threatening clinical syndrome of ARF and must be treated aggressively and early with a combination of anti-inflammatory drugs, diuretics, and, occasionally, steroids to decrease cardiac inflammation acutely.

Electrocardiogram abnormalities may include all degrees of heart block, including atrioventricular dissociation, but first-degree heart block is not associated with a poor prognosis. Second-degree or third-degree heart block occasionally can be symptomatic. If heart block is associated with congestive heart failure, temporary pacemaker placement may be required. The most common manifestation of carditis is cardiomegaly, as seen on radiograph. Among patients at the Rockefeller University Hospital who were diagnosed with ARF between 1950 and 1970 with an average of 20 years of follow-up, 90% had evidence of carditis at diagnosis (Table 115-7). In Bland and Jones's classic review of 1000 patients with ARF, only 65% of patients were diagnosed with carditis. However, when Doppler sonography was employed in the clinical evaluation of patients during the Utah outbreak, 91% of patients had carditis, indicating that, with more sensitive measurements of cardiac dysfunction, almost all ARF patients have signs of acute carditis.

Rheumatic Heart Disease

RHD is the most severe sequel of ARF. Usually occurring 10 to 20 years after the original attack, it is the major cause of acquired valvular disease in the world. The mitral valve is mainly involved, and aortic valve involvement occurs less often. Mitral stenosis is a classic RHD finding and can manifest as a combination of mitral insufficiency and stenosis, secondary to severe calcification of the mitral valve. When symptoms of left atrial enlargement are present, mitral valve replacement may become necessary.

In various studies, the incidence of RHD in patients with a history of ARF has varied. In Bland and Jones's classic follow-up study of patients with ARF, after 20 years, one-third of patients had no murmur, another one-third had died, and the remaining one-third were alive with RHD. Most of the patients who died had RHD. Although the classic dogma is that patients with RHD invariably have had more than one attack of ARF, analysis of our patients at the Rockefeller University Hospital disproves this notion. The population studied consisted of 87 patients who had only one documented attack of ARF with no evidence (clinical or laboratory) of recurrence during a 20-year follow-up under close supervision. More than 80% had carditis at admission, and approximately 50% now have organic murmurs (see Table 115-7). Valvular damage manifesting as organic murmurs later in life is still likely to occur in 5% of patients, particularly if they presented with evidence of carditis at initial diagnosis. All of the patients in our population who ended up with RHD had presented with carditis at diagnosis.

<table>
<thead>
<tr>
<th>Table 115-7</th>
<th>Signs and Symptoms of Acute Rheumatic Fever: Rockefeller University Hospital, 1950 to 1970</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RHD (n = 40) (%)</td>
</tr>
<tr>
<td>Carditis</td>
<td>100</td>
</tr>
<tr>
<td>Arthritis</td>
<td>67.5</td>
</tr>
<tr>
<td>Epistaxis</td>
<td>0</td>
</tr>
<tr>
<td>Chorea</td>
<td>5</td>
</tr>
<tr>
<td>Pericarditis</td>
<td>2.5</td>
</tr>
<tr>
<td>Subcutaneous nodules</td>
<td>7.5</td>
</tr>
<tr>
<td>Erythema marginatum</td>
<td>0</td>
</tr>
</tbody>
</table>

RHD, rheumatic heart disease.
**Chorea**

Sydenham’s chorea (chorea minor, or St. Vitus’ dance) is a neurologic disorder consisting of abrupt, purposeless, nonrhythmic involuntary movements, muscular weakness, and emotional disturbances. Involuntary movements disappear during sleep but may occur at rest and may interfere with voluntary activity. Initially, it may be possible to suppress these movements, which may affect all voluntary muscles, with the hands and face usually the most obviously affected. Grimaces and inappropriate smiles are common. Handwriting usually becomes clumsy and provides a convenient way of following the patient’s course. Speech is often slurred.

Movements are commonly more marked on one side and occasionally are completely unilateral (hemichorea). Muscular weakness is best revealed by asking the patient to squeeze the examiner’s hands: The pressure of the patient’s grip increases and decreases continuously and capriciously—a phenomenon known as relaxing grip, or the milking sign. Emotional changes manifest as outbursts of inappropriate behavior, including crying and restlessness. In rare cases, psychological manifestations may be severe, possibly resulting in transient psychosis. The neurologic examination fails to reveal sensory losses or pyramidal tract involvement. Diffuse hypotonia may be present.

Chorea may follow streptococcal infection after a latent period, which is longer, on average, than the latent period of other rheumatic manifestations. Some patients with chorea have no other symptoms, but other patients develop chorea weeks or months after arthritis. In both cases, examination of the heart may reveal murmurs.

It has been known for years that the early symptoms of chorea often manifest as emotional or behavioral changes in the patient, and only later do the choreiform motor symptoms appear. It also has been noted that many chorea patients years after choreiform symptoms had subsided would present with behavioral disorders, such as tics or obsessive-compulsive disorders. These earlier observations, combined with the known presence of antibrain antibodies in the sera of Sydenham’s chorea patients, raised the question of whether a prior streptococcal infection (or infection with other microbes) might induce antibodies cross-reactive with brain antigen involved in neural pathways associated with behavior. Two more recent articles suggest a strong association of the D8/17 B cell marker (described earlier) with children with obsessive-compulsive disorder (see Table 115-6). Although Swedo and co-workers selected patients on the basis of a strong history of prior streptococcal infection, Murphy and colleagues noted a strong association of the marker in patients with obsessive-compulsive disorder without a history of streptococcal infection. These preliminary studies suggest that streptococci and probably other microbes may induce antibodies that functionally disrupt the basal ganglia pathways, leading not only to classic chorea, but also to behavioral disorders without evidence of classic chorea.

**Subcutaneous Nodules**

The subcutaneous nodules of ARF are firm and painless. The overlying skin is not inflamed and usually can be moved over the nodules. The diameter of these round lesions varies from a few millimeters to 1 or 2 cm. They are located over bony surfaces or prominences or near tendons. Their number varies from a single nodule to a few dozen and averages three or four; when numerous, they are usually symmetric. Nodules are rarely present for longer than 1 month. They are smaller and more short-lived than the nodules of rheumatoid arthritis. Although in both diseases, the elbows are most frequently involved, rheumatic nodules are more common on the olecranon, whereas nodules of rheumatoid arthritis are usually found 3 or 4 cm distal to it. Rheumatic subcutaneous nodules generally appear only after the first few weeks of illness, usually only in patients with carditis.

**Erythema Marginatum**

Erythema marginatum is an evanescent, nonpruritic skin rash, pink or faintly red, that affects usually the trunk and sometimes the proximal parts or the limbs, but not the face. This lesion extends centrifugally, while the skin in the center returns gradually to normal—hence the name erythema marginatum. The outer edge of the lesion is sharp, whereas the inner edge is diffuse. Because the margin of the lesion is usually continuous, making a ring, it is also termed erythema annulare. Individual lesions may appear and disappear in a matter of hours, usually to return. A hot bath or shower may make them more evident or may reveal them for the first time. Erythema marginatum usually occurs in the early phase of the disease. It often persists or recurs, even when all other manifestations of disease have disappeared. Occasionally, the lesions appear for the first time or, more likely, are noticed for the first time late in the course of the illness or even during convalescence. This disorder usually occurs only in patients with carditis.

**Minor Manifestations**

**Fever**

Temperature is increased in almost all ARF attacks and ranges from 38.4° C to 40° C. Usually, fever decreases in approximately 1 week without antipyretic treatment and may become low grade for another 1 or 2 weeks. Fever rarely lasts for longer than 3 to 4 weeks.

**Abdominal Pain**

The abdominal pain of RF resembles that of other conditions associated with acute microvascular mesenteric inflammation and is nonspecific. It usually occurs at or near the onset of the RF attack, so that other manifestations may not yet be present to clarify the diagnosis. In many cases, abdominal pain may mimic acute appendicitis.

**Epistaxis**

In the past, epistaxis occurred most prominently and severely in patients with severe and protracted rheumatic carditis. Early clinical studies reported a frequency of 48%, but it probably occurs even less frequently now (see Table 115-6). Although epistaxis has been correlated in the past
with the severity of rheumatic inflammation, it is difficult to assess retrospectively the possible thrombathenic effect of large doses of salicylates, administered for prolonged periods in protracted attacks.

**Rheumatic Pneumonia**

Pneumonia may appear during the course of severe rheumatic carditis. This inflammatory process is difficult or impossible to distinguish from pulmonary edema or the alveolitis associated with respiratory distress syndromes owing to a variety of pathophysiologic states.

**LABORATORY FINDINGS**

The diagnosis of ARF cannot readily be established by laboratory tests. Nevertheless, such tests may be helpful in two ways: first, in showing that an antecedent streptococcal infection has occurred, and second, in documenting the presence or persistence of an inflammatory process. Serial chest radiographs may be helpful in following the course of carditis, and an electrocardiogram may reflect the inflammatory process on the conduction system. Throat cultures are usually negative by the time ARF appears, but an attempt should be made to isolate the organism. It is our practice to take three throat cultures during the first 24 hours, before administration of antibiotics. Streptococcal antibodies are more useful because (1) they reach a peak titer at about the time of onset of ARF; (2) they indicate true infection, rather than transient carriage; and (3) when several tests for different antibodies are performed, any significant recent streptococcal infection can be detected.

To show a rising titer, it is useful to take a serum specimen when the patient is first seen and to take another 2 weeks later for comparison. The specific antibody tests that have been used most frequently to diagnose streptococcal infection are those directed against extracellular products, including ASO, anti-DNase B, antiyaluronidase (anti-diphosphoryridine nucleotide [anti-DNase]), and anti-streptokinase. ASO has been the most widely used test and is generally available in U.S. hospitals. ASO titers vary with age, season, and geography. They reach peak levels in elementary school-aged children; titers of 200 to 300 Todd units/mL are common in healthy children. After streptococcal pharyngitis, the antibody response peaks at about 4 to 5 weeks, which is usually during the second or third week of ARF (depending on how early it is detected). Antibody titers decrease rapidly over the next several months, and after 6 months, they decline more slowly.

Because only 80% of cases of documented ARF exhibit an increase in the ASO titer, it is recommended that other antistreptococcal antibody tests be done in the absence of a positive ASO titer. These include anti-DNase B, antiyaluronidase, and anti-streptozyme (a combination of various streptococcal antigens). Streptococcal antibodies, when increased, support but do not prove the diagnosis of ARF, and they are not a measure of rheumatic activity. Even in the absence of intercurrent streptococcal infection, titers decline during the rheumatic attack despite the persistence or severity of rheumatic activity.

**Acute Phase Reactants**

Acute phase reactants are elevated during ARF, just as they are during other inflammatory conditions. C-reactive protein and erythrocyte sedimentation rates are almost invariably elevated during the active rheumatic process, if they are not suppressed by antirheumatic drugs. These values may be normal, however, during episodes of pure chorea or persistent erythema marginatum. Particularly when treatment has been discontinued or is being tapered off, C-reactive protein and erythrocyte sedimentation rates are useful in monitoring “rebounds” of rheumatic inflammation, which indicate that the rheumatic process is still active. If C-reactive protein or erythrocyte sedimentation rate remains normal a few weeks after antirheumatic therapy is discontinued, the attack may be considered ended unless chorea appears. Usually, no exacerbation of systemic inflammation occurs, and chorea is present as an isolated manifestation.

**Anemia**

A mild, normochromic, normocytic anemia of chronic infection or inflammation may be seen during ARF. Suppressing the inflammation usually improves the anemia; hematocrit therapy usually is not indicated.

**Other Supporting Findings**

As noted in Figures 115-3 and 115-4 and Table 115-2, two other tests have been helpful in our experience in confirming the diagnosis of ARF, especially when the diagnosis is in doubt. First, one can detect elevated titers of heart-reactive antibodies directed against sarcolemmal antigens in most ARF patients. Elevated levels of these antibodies are not seen in uncomplicated streptococcal infection or acute poststreptococcal glomerulonephritis. With the use of enzyme-linked immunosorbent assay, antibodies directed against cytoskeletal constituents such as myosin and tropomyosin also are seen to be elevated in ARF patients; this observation might be helpful in determining whether or not cross-reactive antibodies unique to ARF exist. Second, use of the D8/17 monoclonal antibody mentioned earlier also has proved helpful in the differential diagnosis of ARF from other disorders. In our hands, all RF patients express abnormal levels of D8/17+ B cells, especially during the acute attack. In cases in which the diagnosis of ARF has been doubtful, the presence of elevated levels of D8/17+ B cells has proved very helpful in establishing the correct diagnosis.

**CLINICAL COURSE AND TREATMENT**

The mainstay of treatment for ARF has always been anti-inflammatory agents, most commonly aspirin. Dramatic improvement in symptoms usually is seen after initiation of therapy. Usually, 80 to 100 mg/kg/day in children and 4 to 8 g/day in adults is required for an effect to be seen. Aspirin levels can be measured; 20 to 30 mg/dL is the therapeutic range. Duration of anti-inflammatory therapy can vary, but treatment needs to be maintained until all symptoms are
absent and laboratory values are normal. If severe carditis also is present (as indicated by significant cardiomegaly, congestive heart failure, or third-degree heart block), steroid therapy can be instituted. The usual dosage is 2 mg/kg/day of oral prednisone during the first 1 to 2 weeks. Depending on clinical and laboratory improvement, the dosage is tapered over the next 2 weeks, and during the last week, aspirin may be added in the above recommended dose, sufficient to achieve 20 to 30 mg/dL.

As noted by Cillers, in a clinical review, studies have shown no difference in the risk of cardiac disease at 1 year in groups treated with aspirin or corticosteroids. Similarly, although nonsteroidal anti-inflammatory drugs also have been used to treat acute inflammation, none have been the subject of randomized controlled trials. Whether or not signs of pharyngitis are present at the time of diagnosis, antibiotic therapy with penicillin should be started and maintained for at least 10 days, given in doses recommended for the eradication of streptococcal pharyngitis. Additionally, all family contacts should be cultured and treated for streptococcal infection if positive. If compliance is an issue, depot penicillins (i.e., benzathine penicillin G, 600,000 U in children and 1.2 million U in adults) should be given. Recurrence of ARF is most common within 2 years of the original attack but can occur at any time. The risk of recurrence decreases with age. Recurrence rates have been decreasing, from 20% in past years to 2% to 4% in more recent outbreaks. This decrease might be due to better surveillance and treatment.

**PROPHYLAXIS**

Antibiotic prophylaxis with penicillin should be started immediately after resolution of the acute episode. The optimal regimen consists of oral penicillin V potassium, 250,000 U twice a day, or parenteral penicillin G, 1.2 million U intramuscularly every 4 weeks. One study suggests, however, that injections every 3 weeks are more effective than every-4-week injections in preventing ARF recurrence. If the patient is allergic to penicillin, erythromycin, 250 mg/day, can be substituted. If the patient is allergic to penicillin, a narrow-spectrum oral cephalosporin (e.g., cefadroxil, cephalaxin) or an oral macrolide (e.g., erythromycin 250 mg per day) can be substituted. Because some penicillin-allergic persons (up to 10%) are also allergic to cephalosporins, these agents should not be used in patients with immediate (anaphylactic-type) hypersensitivity to penicillin.

The endpoint of prophylaxis is unclear; most authors believe it should continue at least until the patient is a young adult, which is usually 10 years from an acute attack with no recurrence. In our opinion, individuals with documented evidence of RHD should be on continuous prophylaxis indefinitely because our experience has been that ARF recurrences can occur even in the fifth or sixth decade. A potential problem for ARF recurrence involves young children in the household, who could transmit new group A streptococcal infection to RF-susceptible individuals. The alternative to long-term prophylaxis in an individual with ARF would be the introduction of streptococcal vaccines designed not only to prevent recurrent infection in susceptible individuals with previous ARF, but also to prevent streptococcal disease in general.

**STREPTOCOCCAL VACCINES**

Difficulties associated with developing a streptococcal vaccine have been related mainly to the numerous reports that streptococcal antigens are known to cross-react with mammalian tissues. Despite these caveats, more recent work indicates progress in this area. Perhaps the most advanced has been the work of Dale and colleagues, in which investigators synthesized short peptides (20 to 30 amino acids) of many different M proteins, linked them together, and showed that they can develop type-specific antibodies that also are opsonic. Little toxicity or cross-reactivity to human tissues has been noted with the antigen or the antibodies.

A second approach revolves around the C-repeat region of the M protein moiety, which is common to all group A streptococci. Bessen and Fischetti used a commensural organism commonly found in the oral mucosa of humans in which by genetic engineering they inserted the C-repeat of the M protein, which is preferentially displayed on the surface of the organism. This induces immunoglobulin (Ig) A antibodies, preventing oral colonization of mice by live group A streptococci. Bronze and colleagues have confirmed these results using a different M-type organism. Golbus and Golbus used similar methods, except that they inserted additional amino acids, making their antibodies opsonic.

Based on the observation by Ellis and colleagues that human sera rarely, if ever, contain more than one type-specific M protein antibody, Salvadori and co-workers examined other possible streptococcal antigens that might explain the broad-based immunity to streptococcal infection that occurs with increasing age. Their studies indicate that the streptococcal group A carbohydrate (GRA-CHO) might be a good immunogen for the following reasons. Antibodies to GRA-CHO are present in human sera, increase with age, and are opsonic for several distinct M-type strains. Active and passive immunization with GRA-CHO in mice provided protection against a live lethal challenge in mice. No cross-reactive antibodies have been detected.

Two other candidates also are under consideration. Gerber and associates described a surface antigen present on group A streptococci called C5a peptidase. This enzyme specifically cleaves the human serum chemotxin C5a at the polymorphonuclear binding site. These observations led to experiments in which intranasal inoculation with C5a peptidase resulted in the appearance of antibodies that might explain the broad-based immunity to streptococcal infection that occurs with increasing age. Their studies indicate that the streptococcal group A carbohydrate (GRA-CHO) might be a good immunogen for the following reasons. Antibodies to GRA-CHO are present in human sera, increase with age, and are opsonic for several distinct M-type strains. Active and passive immunization with GRA-CHO in mice provided protection against a live lethal challenge in mice. No cross-reactive antibodies have been detected.

Based on the observation by Ellis and colleagues that human sera rarely, if ever, contain more than one type-specific M protein antibody, Salvadori and co-workers examined other possible streptococcal antigens that might explain the broad-based immunity to streptococcal infection that occurs with increasing age. Their studies indicate that the streptococcal group A carbohydrate (GRA-CHO) might be a good immunogen for the following reasons. Antibodies to GRA-CHO are present in human sera, increase with age, and are opsonic for several distinct M-type strains. Active and passive immunization with GRA-CHO in mice provided protection against a live lethal challenge in mice. No cross-reactive antibodies have been detected.
strain from the circulation and tissues much more rapidly than the wild-type strain.

CONCLUSION

Despite its disappearance in many areas of the world, ARF continues to be a serious problem in the geographic areas inhabited by two-thirds of the population. In developed countries with full access to medical care, better nutrition, and housing, resurgence of the disease emphasizes the need for continued vigilance of physicians and other health officials in diagnosing and treating ARF. Whether this resurgence represents a change in the virulence of the organism or failure to recognize the importance and adequate treatment of an antecedent streptococcal infection remains an area of intense debate and requires careful and controlled epidemiologic surveillance.

The importance of early diagnosis and therapy cannot be overemphasized. Joint manifestations may be transient and self-limited; however, cardiac sequelae may be chronic and life threatening. Nevertheless, ARF remains one of the few autoimmune disorders known to occur as a result of infection with a specific organism. The confirmed observation of increased frequency of a B cell alloantigen in several populations of rheumatic patients suggests that it might be possible to identify at birth individuals who are susceptible to ARF. If so, from a public health standpoint, (1) these individuals would be prime candidates for immunization with any streptococcal vaccine that might be developed in the future; (2) careful monitoring of streptococcal disease in the susceptible population could lead to early and effective antibiotic strategies, resulting in disease prevention; and (3) in individuals previously infected, who later present with subtle or nonspecific manifestations of the disease, the presence or absence of the marker could be valuable in arriving at a diagnosis.

Continued study of ARF as a paradigm for microbial-host interactions has important implications for the study of autoimmune diseases in general, and rheumatic diseases in particular. Further insight into this intriguing host-parasite relationship may shed additional light on diseases in which infection is presumed but has not yet been identified.

References

55. Lue HC, Mil-Wham W, Hsieh KH, et al: Rheumatic fever recur-
Amyloid diseases are defined by the biochemical nature of the protein in the deposited fibril. The proteins are diverse and are unrelated in primary amino acid sequence, resulting in classification of amyloid diseases according to whether they are systemic or localized, acquired or inherited; their recognized clinical patterns also determine how they are classified (Table 116-1).

Each amyloid disease has a shorthand nomenclature, expressed as A for amyloidosis and an abbreviation for the biochemical nature of the fibrils; for example, AL is amyloid of immunoglobulin light chain origin.

The discussion in this chapter is limited to the systemic amyloidoses because these are the diseases that can involve the joints and are potentially confused with autoimmune rheumatologic disorders.

The acquired systemic amyloidoses are AL (immunoglobulin light chain, or primary), AA (reactive, secondary), and Aβ2M (β2-microglobulin, dialysis-associated) types. The AL type is most common, although epidemiologic data are limited. One study based on National Center for Health Statistics data estimated the incidence as 4.5 per 100,000.

AL amyloidosis usually manifests after age 40 years and is associated with rapid progression, multisystem involvement, and short survival. AA amyloidosis is rare, occurring in less than 1% of patients with chronic inflammatory diseases in the United States and Europe, but it is more common in Turkey and the Middle East, where it occurs in association with familial Mediterranean fever.

Aβ2M amyloidosis is a chronic rheumatologic complication that occurs in a few patients on long-term dialysis and is related to a high concentration of β2-microglobulin.
The inherited amyloidoses are rare autosomal dominant diseases in which a variant plasma protein forms amyloid deposits beginning in midlife. The most common form is caused by variant transthyretins (TTRs), of which more than 100 types are known to be associated with amyloidosis. One variant, with a substitution of isoleucine for valine at position 122 (V122I), has a carrier frequency that may be 4% in the black population and is associated with late-onset cardiac amyloidosis. Among a large cohort of African-Americans over age 65, the frequency of congestive heart failure and mortality was higher among those carrying the V122I gene than in age, gender, and ethnically matched controls. In an amyloidosis referral population, African-Americans over age 60 who presented with amyloid cardiomyopathy were twice as likely to have TTR amyloidosis (ATTR) due to V122I as to have the more common AL type of amyloidosis. Even wild-type TTRs can form fibrils, leading to senile systemic amyloidosis, which predominantly affects the heart, in older patients.

### PATHOLOGY AND PATHOGENESIS OF AMYLOID FIBRIL FORMATION

#### Pathologic Features

Amyloid deposits are widespread in AL amyloidosis and can be present in the extracellular spaces and blood vessels of all organs. Deposits in AA amyloidosis usually develop in the kidneys, liver, and spleen, although widespread deposits can be found late in the course of the disease. In Alβ,M amyloidosis, deposits tend to occur in synovial membrane, cartilage, and bone, but visceral organs are sometimes affected. In ATTR amyloidosis, the nervous system, heart, and thyroid are frequently affected organs, and only small deposits are found elsewhere.

All amyloid deposits stain with Congo red dye and exhibit a unique green birefringence by polarized light microscopy, although the deposits also can be recognized on routine hematoxylin and eosin–stained sections. By electron microscopy, amyloid fibrils are 8 to 10 nm wide and of varying lengths, with a 2.5- to 3.5-nm filamentous subunit arranged along the long axis of the fibril in a slow twist. Typing of amyloid deposits can be done with conventional immunohistochemical staining. False-positive results can occur, however, owing to the presence of nonamyloid serum proteins, and immunoelectron microscopy can provide more definitive immunologic identification of the protein in the fibril, or the composition of extracted fibrils can be identified by mass spectrometric proteomic analysis.

#### Pathogenesis of Amyloid Fibril Formation

The exact mechanism of fibril formation is unknown and may differ among the various types of amyloid. Studies suggest a common underlying mechanism, however, in which a partially unfolded protein intermediate forms multimers and then higher-order polymers. Factors that contribute to fibrillogenesis include variant or unstable protein structure, extensive β-conformation of the precursor protein, proteolytic processing of the precursor protein, association with components of the serum or extracellular matrix (e.g., serum amyloid P [SAP] component, amyloid enhancing factor, apolipoprotein E, glycosaminoglycans), and physical properties such as pH of the tissue site.

AL amyloidosis is a plasma cell dyscrasia with an excess of clonal plasma cells in the bone marrow; it can occur in isolation or in combination with multiple myeloma. Similar cytogenetic changes have been identified in these plasma cell diseases, suggesting that they may have a common molecular pathogenesis. By two-dimensional gel electrophoresis and mass spectrometry, it can be seen that the amyloid fibril deposits are composed of intact 23-kD monoclonal immunoglobulin light chains and C-terminal truncated fragments. Although all κ and λ light chain subtypes have been identified in amyloid fibrils, κ subtypes predominate, and the λ 6 subtype seems to have unique structural properties that predispose it to fibril formation, often in the kidney. AL amyloidosis is usually a rapidly progressive disease with amyloid deposits in multiple tissue sites.
The AA type of amyloidosis is a complication of severe, long-standing inflammation, as occurs in chronic inflammatory disorders or infections. AA amyloid fibrils usually are composed of an 8-kD, 76 amino acid amino-terminal portion of the 12-kD precursor, serum amyloid A (SAA). SAA is a polymorphic protein encoded by a family of SAA genes, which are acute phase apoproteins synthesized in the liver and transported by a high-density lipoprotein, HDL3, in the plasma. An underlying inflammatory disease of several years’ duration causing an elevated SAA usually precedes fibril formation, although infection can produce AA deposition more quickly. AA fibril formation can be accelerated by an amyloid enhancing factor present in high concentration in the spleen (which may be early SAA aggregates or deposits), by basement membrane heparan sulfate proteoglycan, or by seeding with AA or heterologous fibrils.

Factors related to β2-microglobulin fibril formation are under investigation. The high prevalence of Aβ2M disease in patients undergoing long-term dialysis argues against an amyloidogenic variant β2-microglobulin molecule. The permeability of dialysis membranes may be a factor because the molecular weight of β2-microglobulin is 11.8 kD—above the porosity of standard membranes. It has been hypothesized that dialysis membranes may be bioincompatible and may induce proinflammatory mediators that stimulate β2-microglobulin and contribute to fibril formation.

In ATTR (also called familial amyloidotic polyneuropathy) and all other forms of familial amyloidosis, inherited mutations or polymorphisms in the genes encoding large serum proteins produce amyloid-prone variants. The process of fibrillogenesis has been best studied for TTR, in which variant TTR molecules seem to be prone to dissociation from stable tetramers and to unfolding, leading to misfolding, polymerization, and fibril formation. The role of aging is intriguing because patients with variant proteins do not have clinically apparent disease until midlife or later, despite the lifelong presence of the abnormal protein. Further evidence of an age-related trigger is that senile cardiac amyloidosis, caused by the deposition of fibrils derived from normal TTR, is exclusively a disease of elderly individuals.

**DIAGNOSIS**

A tissue biopsy specimen showing amyloid fibrils is necessary for the diagnosis of amyloidosis (Figure 116-1). The least invasive biopsy is the abdominal fat aspirate, which is positive in 80% to 90% of patients with AL or ATTR amyloidosis and in 60% to 70% of patients with AA amyloidosis. It is easy to perform after local injection of anesthetic and has a low rate of infectious or hemorrhagic complications (Figure 116-2). If the aspirate is negative, but clinical suspicion for disease persists, a more invasive tissue biopsy should be done. Although a biopsy specimen of a clinically involved organ is recommended, almost any tissue biopsy specimen is likely to be positive if the patient has systemic amyloidosis: In a series of 100 patients with AL amyloidosis, 85% of 249 tissue biopsy specimens were positive, including
microscopy should be done by a laboratory familiar with the techniques and able to perform appropriate controls. Mass spectrometry–based microsequencing of small amounts of protein extracted from fibril deposits ultimately may be the most reliable way to identify the components of the fibrils.\(^ {16,17}\) AA amyloidosis is suspected in patients with renal amyloidosis and a chronic inflammatory condition or infection. AL and ATTR amyloidosis must be ruled out. AA amyloidosis must be confirmed by immunohistochemical staining for AA protein.

Familial amyloidosis must be excluded in every patient who does not have a plasma cell dyscrasia or the AA type of amyloidosis. Although the disease has a dominant inheritance, family history may not be apparent when the disease occurs later in life; also, some cases occur through new mutations. Variant TTR proteins usually can be detected by isoelectric focusing (Figure 116-5).\(^ {36}\) Abnormal isoelectric focusing should prompt genetic testing to determine the precise TTR mutation. Genetic testing should be employed when screening tests fail to identify the fibril protein. With the use of polymerase chain reaction–based sequencing, abnormal fibrinogens and apolipoproteins and variant TTRs can be detected.\(^ {37}\)

A novel renal amyloid protein, leukocyte chemotactic factor 2 (LECT2), has been found in glomeruli, renal vessels, and interstitium in patients with isolated renal amyloidosis and should be considered in patients who are negative for the more common types of amyloidosis.\(^ {38}\)

**CLINICAL FEATURES AND TREATMENT OF THE SYSTEMIC AMYLOIDOSES**

**AL Amyloidosis**

AL amyloidosis usually occurs in middle-aged or older individuals but also can occur in the third or fourth decade of life. It has a wide spectrum of organ system involvement, and presenting features reflect the organs most prominently affected.\(^ {39,40}\) Initial symptoms of fatigue and weight loss are frequent, but the diagnosis is rarely made until symptoms referable to a specific organ appear.

The kidneys are commonly affected; renal amyloidosis is manifested by proteinuria, sometimes massive, with edema and hypoalbuminemia. Mild renal dysfunction is frequent, but rapidly progressing renal failure is rare. Cardiac...
have reviewed the soft tissue and joint manifestations of AL amyloidosis in a series of almost 200 patients. Symptoms and signs mimicking rheumatologic diseases, with arthropathy, subcutaneous tissue deposits, muscle pseudohypertrophy, adenopathy, carpal tunnel syndrome, submandibular gland enlargement, and macroglossia, occur in more than 40% of patients with AL amyloidosis.

Timely diagnosis of AL amyloidosis is crucial. Patients with any of these clinical syndromes should be screened for amyloid deposition and for a plasma cell disorder as previously described. Note that the sensitivity of serum or urine protein electrophoresis (SPEP and UPEP) without immunofixation is poor; these tests alone are not useful for excluding or diagnosing AL amyloidosis.

Extensive multisystem involvement typifies AL amyloidosis, and median survival with no treatment is usually only about 1 year from diagnosis. Current therapies target clonal bone marrow plasma cells using chemotherapy approaches employed for multiple myeloma (Table 116-2). Oral melphalan and prednisone constituted the first regimen tried for AL amyloidosis; it was minimally effective and is rarely used today because recent studies show that substitution of high-dose dexamethasone for prednisone seems to increase response rates significantly.

High-dose intravenous melphalan followed by autologous stem cell transplantation is highly effective. Of more than 500 patients treated on such protocols at Boston Medical Center, approximately 40% of evaluable patients had achieved a hematologic complete response when assessed at 1 year post treatment; most experienced significant improvement or stabilization of organ function. Median survival in treated patients exceeds 4.5 years and
is durable. Other referral centers have found similar results. However, patients with amyloidosis and organ impairment have high rates of treatment-related morbidity and mortality with aggressive treatment. Factors that contribute to mortality include amyloid cardiomyopathy, poor performance or nutritional status, reduced pulmonary function, and amyloid-associated bleeding disorders. The only randomized multicenter study comparing oral melphalan chemotherapy with intravenous (IV) melphalan chemotherapy with stem cell support was plagued by high treatment-related mortality and failed to show benefit for the high-dose regimen. At experienced centers, this treatment is still considered first-line in low-risk patients. Age alone or renal failure does not exclude patients from such treatment. Cardiac biomarkers have been used to stratify risk for high-dose therapy.

For patients with significantly impaired cardiac function or arrhythmias resulting from amyloid involvement of the myocardium, median survival is only about 6 months without treatment, and stem cell mobilization and high-dose chemotherapy are associated with great morbidity. In such patients, orthotopic cardiac transplantation followed by intravenous melphalan and stem cell rescue to prevent fibrillogenesis in the transplanted heart or other organs can be effective.

New agents that are efficacious in reducing the plasma cell burden in multiple myeloma are being tested for AL amyloidosis. These include immunomodulators such as lenalidomide that affect the bone marrow microenvironment, as well as proteasome inhibitors to which plasma cells are particularly sensitive, such as bortezomib. These agents generally have been tested for use in patients for whom alkylator chemotherapy has failed; current clinical trials are examining these novel agents in combination with alkylators, as induction before transplant, and for maintenance. Weekly rather than bi-weekly dosing of bortezomib appears to reduce the incidence of peripheral neuropathy and may be safer in AL amyloidosis patients. Innovative approaches target the amyloid fibrils themselves or accessory binding proteins. The anthracycline derivative 4′-iodo-4′-deoxoxyorubicin (IDOX) was noted to cause resorption of amyloid deposits in model systems, but clinical trials have failed to show any clinical benefit. The agent R-1-[6-([R-2-carboxy-pyrrolidin-1-yl]-6-oxo-hexanoyl]pyrrolidine-2-carboxylic acid (CPHPC) binds serum amyloid P protein and accelerates clearance from the circulation and from amyloid fibrils in animal models, but it has not been proven to provide any clinical benefit on its own. In animal models, the addition of component antibody against SAP appears to further accelerate clearance, and clinical trials with this combination are being planned. Supportive treatment is recommended for patients with all types of amyloidosis (Table 116-3). At times, supportive treatment may be lifesaving (e.g., heart or kidney transplantation, renal dialysis, cardiac pacemaker, nutritional support). Digitalis, calcium channel blockers, and β-blockers are relatively contraindicated because toxicity has been observed at therapeutic levels.

AA Amyloidosis

AA amyloidosis can occur at any age. The primary clinical manifestation is proteinuria or renal insufficiency or both. A study from Finland found AA amyloidosis to be the most common cause of nephrotic syndrome in patients with rheumatoid arthritis. Hepatomegaly, splenomegaly, and autoimmune neuropathy frequently occur as the disease progresses; cardiomyopathy occurs rarely. With chronic inflammatory diseases, amyloid progression is slow, and survival is often longer than 10 years, particularly with treatment for end-stage renal disease. In contrast, untreated infections, such as osteomyelitis, tuberculosis, or leprosy, can produce a more rapidly progressive amyloid syndrome, which remits with effective medical or surgical treatment of the infection.

The major therapy in AA amyloidosis is treatment of the underlying inflammatory or infectious disease. Treatment that suppresses or eliminates the inflammation or infection also decreases the SAA protein. For familial Mediterranean fever, colchicine, 1.2 to 1.8 mg/day, is the appropriate treatment. Colchicine has not been helpful for AA amyloidosis of other causes or for other amyloidoses. A multicenter randomized trial using a new antiamyloid drug, eprosidade, has been completed; the drug was found to significantly delay worsening of renal function in patients with AA amyloidosis. A second multicenter trial requested by the Food and Drug Administration (FDA) is now underway. Eprosidade interferes with the interaction of AA amyloid protein and glycosaminoglycans in tissues and prevents fibril formation and deposition.

AβM Amyloidosis

Several distinct rheumatologic conditions are observed in AβM amyloidosis, including carpal tunnel syndrome, persistent joint effusions, spondyloarthropathy, and cystic bone...
lesions. Carpal tunnel syndrome is usually the first symptom of disease. Persistent joint effusions accompanied by mild discomfort occur in 50% of patients on dialysis for longer than 12 years. Involvement is bilateral, and large joints (shoulders, knees, wrists, and hips) are more frequently affected. The synovial fluid is noninflammatory, and $\beta_2$-microglobulin amyloid deposits can be found if the sediment is examined with Congo red staining. Spondyloarthropathy with destructive changes in the intervertebral disks and paravertebral erosions have occurred in association with $\beta_2$-microglobulin amyloid deposits. Cystic bone lesions sometimes leading to pathologic fractures have been described in the femoral head, acetabulum, humerus, tibial plateau, vertebral bodies, and carpal bones. Although less common, visceral $\beta_2$-microglobulin amyloid deposits occasionally occur in the gastrointestinal tract, heart, tendons, and subcutaneous tissues of the buttocks.

Treatment for $\text{A}\beta_2\text{M}$ amyloidosis is difficult to provide because the 11-kD $\beta_2$-microglobulin molecule is too large to pass through a dialysis membrane. Consistent with a postulated role of copper in initiating $\text{A}\beta_2\text{M}$ fibrillogenesis, copper-free dialysis membranes seem to reduce the incidence of disease. Patients on continuous ambulatory peritoneal dialysis usually have lower plasma levels of $\beta_2$-microglobulin than patients on hemodialysis and may not develop amyloid deposits as quickly. Symptoms of arthropathy are common, and prevalence may approach 100% of individuals on dialysis for longer than 15 years. Patients who have received kidney transplants after developing $\text{A}\beta_2\text{M}$ report improvement in symptoms.

### Table 116-3  Supportive Treatment for All Types of Amyloidosis

<table>
<thead>
<tr>
<th>Organ System</th>
<th>Symptom</th>
<th>Treatment Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac</td>
<td>Congestive failure</td>
<td>Salt restriction of 1-2 g/day</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diuretics: furosemide, spironolactone, metolazone</td>
</tr>
<tr>
<td></td>
<td>Arrhythmia</td>
<td>Pacemaker</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Automatic implantable cardiac defibrillator</td>
</tr>
<tr>
<td>Renal</td>
<td>Nephrotic syndrome</td>
<td>Salt restriction of 1-2 g/day</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Elastic stockings, leg elevation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maintaining dietary protein</td>
</tr>
<tr>
<td></td>
<td>Renal failure</td>
<td>Angiotensin-converting enzyme inhibitor, if blood pressure tolerates</td>
</tr>
<tr>
<td>Autonomic nervous</td>
<td>Orthostatic hypotension</td>
<td>Dialysis (long-term ambulatory peritoneal dialysis or hemodialysis)</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>Diarrhea</td>
<td>Increased dietary salt or added fludrocortisone, depending on edema</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Elastic stockings</td>
</tr>
<tr>
<td></td>
<td>Gastric atony or ileus</td>
<td>Small frequent feedings (6/day) low in fat</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral nutritional supplements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jejunostomy tube feeding</td>
</tr>
<tr>
<td>Periarterial</td>
<td>Macroglossia</td>
<td>Parenteral nutrition</td>
</tr>
<tr>
<td>Hematologic</td>
<td>Intracutaneous bleeding</td>
<td>Soft solid diet</td>
</tr>
<tr>
<td></td>
<td>Factor X deficiency</td>
<td>Partial glossectomy (rarely effective)</td>
</tr>
</tbody>
</table>

### ATTR Amyloidosis

The clinical features of ATTR amyloidosis overlap AL amyloidosi such that the diseases cannot be reliably distinguished on clinical grounds alone. A family history makes ATTR more likely, but many patients seem to present sporadically with new mutations. Within each family, disease begins at nearly the same age, and symptoms usually include neuropathy or cardiomyopathy or both. Peripheral neuropathy begins as a lower extremity sensory and motor neuropathy and progresses to the upper extremities. Autonomic neuropathy is manifested by gastrointestinal symptoms of diarrhea with weight loss and orthostatic hypotension. Cardiomyopathy and conduction system defects are similar to those caused by AL amyloidosis, although in ATTR, heart failure is less common, and the prognosis is better. Vitreous opacities caused by amyloid deposits are pathognomonic of ATTR amyloidosis.

The TTR variant, V122I, is a common allele in African-Americans and is associated with cardiomyopathy. In a large referral population, 25% of African-American patients with amyloidosis had this TTR variant. ATTR due to V122I is likely underdiagnosed because of lack of physician awareness and the difficulty of distinguishing amyloid and hypertensive cardiomyopathy without an endomyocardial biopsy.

Without intervention, survival after ATTR disease onset is 5 to 15 years. Orthotopic liver transplantation, which removes the major source of variant TTR production and replaces it with normal TTR, is the major treatment for ATTR amyloidosis. Liver transplantation arrests disease...
progression, and some improvement in autonomic and peripheral neuropathy may occur. Cardiomyopathy does not improve and in some patients seems to worsen after liver transplantation. A beneficial long-term outcome depends on transplantation early in the disease course. Two international multicenter randomized placebo-controlled clinical trials have been conducted to test the efficacy of the nonsteroidal anti-inflammatory drug, diflunisal, and its analog tafamidis, for the treatment of ATTR amyloidosis (www.bu.edu/amyloid/doctors/trials/html) based on laboratory studies showing that these agents stabilize TTR and prevent unfolding and aggregation. Antisense and small interfering RNA approaches are also being developed.

SUMMARY

Timely and accurate diagnosis of amyloidosis is essential because effective treatments for some forms of amyloidosis are available or are undergoing clinical trials. The first step is recognition of a clinical syndrome consistent with amyloidosis; this is followed by an appropriate biopsy or fat aspirate to identify tissue fibrils. The next priority should be to determine whether an associated plasma cell disorder can be identified, because AL is the most rapidly progressive type of systemic amyloidosis, and therapy should be initiated before heart, kidney, or liver failure occurs. Definitive identification of the amyloid precursor protein is essential for appropriate therapy, and amyloid referral centers can provide specialized diagnostic techniques and access to clinical trials. An understanding of the biophysical properties of amyloid proteins and of the mechanisms of protein misfolding and tissue damage will enable the further development of more specific and less toxic anti-amyloid therapeutics.

Acknowledgments

This chapter was supported by grants from the National Institutes of Health (HL 68705), the Gerry, Gruss, and Wildflower Foundations; the Health (HL 68705), the Gerry, Gruss, and Wildflower Foundations; the

References


The references for this chapter can also be found on www.expert.consult.com.
Sarcoidosis is a heterogeneous multisystem inflammatory disease of unknown cause characterized by the development and accumulation of noncaseating granulomata in any organ system.

Sarcoidosis occurs worldwide and affects people of all racial and ethnic backgrounds.

The pathogenesis of sarcoidosis likely involves the interplay of different cells, cytokines, and other inflammatory mediators.

Rheumatologic manifestations are common in sarcoidosis and are often overlooked or misdiagnosed.

No treatments have been FDA-approved for sarcoidosis or for any of its manifestations, including sarcoid arthritis.

New treatment modalities, including tumor necrosis factor (TNF) inhibitors, appear to be promising for the treatment of extrathoracic sarcoidosis based on case series, although randomized controlled trials have shown limited benefit in cases of pulmonary involvement.

**KEY POINTS**

Sarcoidosis is a heterogeneous multisystem inflammatory disease of unknown cause characterized by the development and accumulation of noncaseating granulomata in any organ system.

Sarcoidosis occurs worldwide and affects people of all racial and ethnic backgrounds.

The pathogenesis of sarcoidosis likely involves the interplay of different cells, cytokines, and other inflammatory mediators.

Rheumatologic manifestations are common in sarcoidosis and are often overlooked or misdiagnosed.

No treatments have been FDA-approved for sarcoidosis or for any of its manifestations, including sarcoid arthritis.

New treatment modalities, including tumor necrosis factor (TNF) inhibitors, appear to be promising for the treatment of extrathoracic sarcoidosis based on case series, although randomized controlled trials have shown limited benefit in cases of pulmonary involvement.

**Sarcoidosis**

NADERA J. SWEISS • ROBERT P. BAUGHMAN

Sarcoidosis is an orphan, systemic, clinically heterogeneous disorder. Its cause has yet to be identified, but environmental, genetic, and infectious causes have been suggested. The hallmark of sarcoidosis is the development and accumulation of noncaseating granulomas in any organ system. Organ system involvement, which is unpredictable and varies between patients, is the major determinant of morbidity and mortality in sarcoidosis. Although any organ system can be involved, the lungs are affected in most cases. Lymphatic, skin, and ocular findings are also common. Given the variability of sarcoidosis manifestations, diagnosing this disorder is often difficult. Patients may be asymptomatic or may present with a range of nonspecific symptoms, but specific symptoms such as cough, dyspnea, burning of eyes, or rash may suggest the diagnosis. When extrapulmonary symptoms develop, they sometimes result in rheumatologic manifestations including but not limited to arthritis, skin lesions, arthralgias, and neuropathy.

**Epidemiology**

Sarcoidosis is a global disease. Because of its clinical heterogeneity and its variable diagnostic criteria in different countries, the worldwide prevalence and incidence of sarcoidosis have been difficult to calculate. In Northern Europe, up to 40 cases per 100,000 people have been reported. A study from Eastern Europe found only 3.68 cases of sarcoidosis per 100,000 people. The incidence of sarcoidosis in Japan is also low, with one study estimating it at 3.7 cases per 100,000 individuals. Within countries, incidence rates may vary between races. In the United States, the annual incidence of sarcoidosis is more than three times higher in black individuals (35.5 per 100,000) than in white people (10.9 per 100,000). Furthermore, the disease course of sarcoidosis is more progressive and may be more fatal in black Americans. Despite numerous epidemiologic studies of sarcoidosis, many clinicians and researchers believe that estimates of prevalence and incidence are lower than actual rates of the disease owing to inaccurate diagnoses or asymptomatic cases that are never diagnosed.

Although sarcoidosis affects men and women of all ages and from diverse ethnic backgrounds, some disparities in how it affects these groups have been noted. The fact that it affects slightly more women than men has been confirmed in studies from around the world; estimates indicate that 57% of patients with sarcoidosis are women. Relative to men, women with sarcoidosis have a greater number of ocular and neurologic manifestations. People of any age may acquire the disease, but the median age of onset is around 40. A second peak of incidence has been reported around age 65, especially in women.

**IMMUNOPATHOGENESIS**

The pathogenesis of sarcoidosis likely involves the interplay of many different cells, cytokines, and other inflammatory mediators. Granuloma development is the characteristic pathologic feature of tissue involvement in sarcoidosis. Physiologically, granulomas act as shields, protecting tissues from pathogens, thereby pre-empting inflammatory reactions. Their formation is the end product of a coordinated effort involving T cell activation, antigen-presenting cell (APC) activation, and cell signaling. Granulomas consist of a core of mononuclear phagocytes, such as epithelioid cells, multinucleated giant cells, and macrophages, which are encased by lymphocytes, including B cells, CD4+ T cells, and CD8+ T cells.

**Innate Immunity**

T cell activation is required for granulomas to form. Early in the course of sarcoidosis, an unknown antigen or multiple antigens activate T cells and macrophages, thereby triggering downstream signaling from both cell types. Locally activated CD4+ T helper cells differentiate into T helper type 1 (Th1)-like cells, causing subsequent elevations in Th1-associated inflammatory mediators, such as interleukin (IL)-2, interferon (IFN)-α, IFN-γ, monocyte chemotactic...
protein-1 (MCP-1), macrophage inflammatory protein-1 (MIP-1), and granulocyte-macrophage colony-stimulating factor (GM-CSF). CD4+ T helper cells also interact with APCs to initiate the development of and preserve granulomas. Thus, T cells undergo oligoclonal proliferation in areas of immune system activity. At this point in the disease process, lymphocyte levels are typically increased, and the ratio of CD4/CD8 cells becomes elevated in the lungs and other affected organs.

**Acquired Immunity**

In sarcoidosis, uptake of antigens is the most likely trigger for activation of macrophages, which are then able to produce IL-12, IL-15, IL-18, and tumor necrosis factor (TNF). Macrophage-derived cytokines contribute to the external signaling milieu that selectively pressures toward Th1 differentiation of CD4+ cells. Ultimately, a feedback loop from downstream-produced cytokine cascade induces macrophages to differentiate into epithelioid cells, which gain secretory capacity, lose phagocytic capacity, and fuse to form multinucleated giant cells. These epithelioid cells form the cellular basis of granulomas. Also contributing to granuloma formation are Th2 cells. These cells synthesize fibronectin and CC motif ligand 18 (CCL18). Release of these mediators results in a positive feedback loop of CCL18-activated and macrophage-mediated collagen formation. Although granulomas spontaneously resolve without causing damage in most cases, this cycle leads to fibrosis in up to 25% of patients with sarcoidosis. As patients’ fibrosis becomes more extensive, their prognosis worsens. Although the mechanisms responsible for granuloma fibrosis have not been fully characterized, patterns of cytokines change, and Th2 cells may pressure toward an increased ratio of CD8 to CD4 cells. Less than 5% of patients die from sarcoidosis, but fibrosis leading to respiratory failure is a contributing factor in many sarcoidosis deaths.

In summary, the essential immunologic events in granuloma formation can be summarized as follows: (1) antigen exposure, (2) antigen processing and presentation by macrophages, resulting in T cell immunity against the antigen, (3) T-effector cell production, (4) macrophage activation, and (5) granuloma formation.

**ETIOLOGY**

Sarcoidosis has yet to be attributed to a single factor. The heterogeneity of the disease suggests that multiple causative agents may be responsible for the variable disease manifestations of sarcoidosis. Given the immunopathogenic mechanisms that underlie sarcoidosis, such a disease trigger might be a T cell antigen that stimulates the cascade of events leading to granuloma formation. Because the lungs are affected in more than 90% of cases of sarcoidosis, it is likely that an environmental agent (including an infectious agent, potentially exposed via the pulmonary route) might contribute to onset of sarcoidosis.

Both inorganic and organic environmental factors with antigenic capabilities have been implicated in the pathogenesis of sarcoidosis. Early studies on the causes of sarcoidosis suggested a link between sarcoidosis and agents associated with a rural lifestyle, such as the lumber industry and burning wood. These data have been extended in the ACCESS study, which found that agricultural debris and wood burning are associated particularly with pulmonary sarcoidosis but not with systemic sarcoidosis. In a different analysis of the ACCESS trial, radiation, insecticides, mildew, and mold were environmental factors associated with systemic sarcoidosis phenotype. These findings may indicate that each of these unique sarcoidosis subtypes has its own causes.

Numerous methods have been used to look for an infectious agent as a cause of sarcoidosis. When in situ hybridization was used, *Mycobacterium tuberculosis* catalase-peroxidase protein (mKatG) was found in nearly 40% of tissue samples from patients with sarcoidosis. Recombinant mKatG protein was then used to measure mKatG antibodies in patients with sarcoidosis, which were present in 50% of patients studied. Others have found evidence of an immunologic reaction to additional mycobacterial antigens. *Propionibacterium acnes* has been found more frequently in granulomas from sarcoidosis patients but can also be found in individuals without sarcoidosis, so its primary role in pathogenesis remains unclear. These studies suggest that sarcoidosis may represent overexposure of the patient to a commonly encountered microorganism associated with dysregulated resultant immune responses. Because several types of bacteria have been associated with sarcoidosis, some clinicians have attempted to use antibiotics to manage the disease. Although skin sarcoidosis has, on occasion, responded to antibiotics, their usefulness in other forms of sarcoidosis appears minimal.

**GENETICS**

Compelling studies of familial clustering and incidence of sarcoidosis among different racial groups indicate that sarcoidosis susceptibility is influenced by the interplay of genetic factors. In the ACCESS study, first-degree relatives were reported to have a fivefold increase in risk of developing the disease. Associations have been found between risk of sarcoidosis and class I and II human leukocyte antigen (HLA) gene products, which have essential roles in antigen presentation. It is considered likely that a susceptibility locus for sarcoidosis exists within the HLA gene region, as is the case with other autoimmune diseases and cancers such as Hodgkin’s lymphoma. An intriguing analysis of data from the ACCESS study identified associations between genetic factors (HLA alleles), environmental factors, and sarcoidosis phenotypes. In considering together several of the factors postulated to cause sarcoidosis, a compelling argument can be made for a genetic factor that predisposes individuals to the disease and a subsequent environmental exposure that triggers onset of sarcoidosis. Specifically, Rosman and colleagues found that HLA-DRB1*1101 and insecticide exposure at work are significantly associated with cardiac sarcoidosis and hypercalcemia. A similar relationship between HLA-DRB1*1101, mold and musty odors, and pulmonary sarcoidosis was described. HLA polymorphisms have also been linked to Löfgren’s syndrome, an acute form of sarcoidosis characterized by bilateral hilar lymphadenopathy (BHL), erythema nodosum (EN), fever, and periarthritis ankle inflammation or
Ongoing genome-wide association studies (GWASs) are seeking to identify additional genes that may be linked to sarcoidosis onset and susceptibility. To date, these studies have identified several novel gene candidates. In family clusters with sarcoidosis, GWASs have identified areas of interest in a German population. A similar analysis of African-American familial sarcoidosis did not have identical findings but did find the same linkage at chromosomes 1p and 9q.

**DIAGNOSING SARCOIDOSIS**

The diagnosis of musculoskeletal disease related to sarcoidosis must be examined on the basis of two different presentations. The first is a patient with known sarcoidosis who presents with musculoskeletal complaints. The other is a patient with musculoskeletal disease in whom sarcoidosis is a possibility. Although the evaluations may be similar, crucial differences in approach are required. One can never be sure of the diagnosis of sarcoidosis. The American Thoracic Society (ATS), the European Respiratory Society (ERS), and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) have developed diagnostic criteria. The patient with appropriate clinical presentation and multiple organ involvement who has granulomas identified in one or more organs, and who has no other cause for the granulomatous reaction, is considered to have sarcoidosis. Figure 117-1 presents some features consistent with the disease. Diagnosis relies

---

Figure 117-1 Various manifestations of sarcoidosis. A, Bilateral hilar adenopathy and right paratracheal lymph node enlargement demonstrated on a posterior anterior chest roentgenogram, Scadding stage 1. B, Facial lesion consistent with lupus pernio. C, Hand changes consistent with sarcoidosis in the fingers. D, Noncontrast magnetic resonance image of the pelvis demonstrating bone marrow replacement by granulomatous tissue. E, Cystic changes (arrows) within the bones of the fingers of a patient with sarcoidosis. F, Gadolinium enhancement of lesions of the spine seen on magnetic resonance image of a sarcoidosis patient. (B, Reproduced with permission from the patient.)
heavily on the finding of granulomas in the biopsy. However, many other conditions can lead to a granulomatous reaction, as is summarized in Table 117-1. Although sarcoidosis granulomas tend to be noncaseating and non-necrotizing, a significant number of sarcoidosis patients exhibit evidence of some necrosis in part of their granulomatous response. Certain features such as erythema nodosum, BHL, gallium scan showing increased activity in the parotids and lacrimal glands, an elevated angiotensin-converting enzyme (ACE) level, and bronchoalveolar lavage with increased lymphocytes with a CD4/CD8 ratio greater than 3.5 are strongly suggestive of sarcoidosis, even in a patient in whom a biopsy has not been performed.

The heterogeneity of the disease leads to varied clinical presentations, which may result in delayed diagnosis. One study of 189 patients with sarcoidosis found that only 15.3% of patients receive a diagnosis during their first visit to a physician. Moreover, sarcoidosis can affect any organ, and this may cause a patient to be referred to a specialist who may not commonly manage sarcoidosis. Complicating further the diagnosis of sarcoidosis is its similarity of presentation to or concurrent existence with a variety of other conditions, most notably autoimmune disorders. Thus, sarcoidosis sometimes is diagnosed after other potential diagnoses have been excluded.

For the patient who presents with musculoskeletal disease and possible sarcoidosis, a vigilant approach is mandatory. A comprehensive multisystem evaluation as is commonly performed in rheumatology practice should prove informative. Table 117-2 lists some of the common manifestations that support or refute the diagnosis. For example, hilar adenopathy on a posterior chest roentgenogram is seen in about two-thirds of patients. Scadding proposed a staging system of the chest roentgenogram: stage 1 disease consists of hilar adenopathy alone (see Figure 117-1A); stage 2 disease is hilar adenopathy plus infiltrates; stage 3 is infiltrates alone; and stage 4 is fibrosis. Although this scoring is commonly used, differences may be noted in staging of the roentgenogram, even among sarcoidosis experts. Patients may have evidence of pathology on chest computed tomography (CT) scan. Table 117-2 lists some less common features, which, when present, support the diagnosis. These include erythema nodosum, lupus pernio (see Figure 117-1B), cranial seventh nerve paralysis, and hypercalcemia.

For patients with sarcoidosis, several tests (summarized in Table 117-3) have been proposed to serve as a minimal evaluation. These tests reflect the fact that sarcoidosis is a multiple-organ disease, and one needs to look for evidence of specific organ involvement. Figure 117-1 demonstrates some of the features associated with sarcoidosis. Magnetic resonance imaging (MRI) with gadolinium may identify involvement of sarcoidosis in the brain or in the heart. MRI findings may also be useful in bone sarcoidosis or sarcoid arthritis.

Several lines of evidence have indicated that 18F-fluorodeoxyglucose positron-emission tomography (18F-FDG/PET) may be useful in extrapulmonary sarcoidosis, but it usually is not recommended as a first diagnostic tool. 18F-FDG/PET can be used to determine which organs are affected, particularly when no evidence of lung involvement is found. This includes cardiac sarcoidosis. Furthermore, this test can point investigators to appropriate organs

| Table 117-1 Major Causes of Granulomatous Reaction besides Sarcoidosis |
|-------------------|-------------------|-------------------|
| Infection         | Environmental     | Miscellaneous     |
| Mycobacterium     | Berylliosis       | ANCA-associated   |
| tuberculosis      | Hard metal        | vasculitis        |
| Fungal            | Zirconium         | Necroting sarco    |
| Mycoplasma        | Tattoo            | sarcoma           |
| Pneumocystis      | Hypersensitivity  | granuloma         |
| jiroveci          | pneumonitis       | Lymphoma          |
| Brucellosis       | Drugs (e.g.,      | Cancer            |
|                  | methotrexate)     | Granulomatous     |
|                  |                   | lesions of unknown|
|                  |                   | significance       |
|                  |                   | Crohn’s disease   |
|                  |                   | Lymphocytic       |
|                  |                   | interstitial      |
|                  |                   | pneumonia         |
|                  |                   | Behcet’s disease  |
|                  |                   | Rheumatoid nodules|

ANCA, antineutrophil cytoplasm antibody.

| Table 117-2 Features Characteristic of Sarcoidosis |
|---------------------------|---------------------------|---------------------------|
| More Likely Sarcoidosis   | Less Likely Sarcoidosis   |
| Chest roentgenogram       | Bilateral hilar adenopathy|
|                           | Upper lobe disease        |
|                           | Subpleural reticulonodular|
| Computed tomography of    | Mediastinal adenopathy    |
| chest                      | Peribronchial thickening  |
|                           | Traction bronchiectasis   |
| Skin lesions               | of upper lobe             |
|                           | Erythema nodosum          |
|                           | Lupus pernio              |
|                           | Maculopapular lesions     |
| Ocular disease             | Uveitis                   |
|                           | Optic neuritis            |
| Neurologic disease         | Cranial seventh nerve     |
| Renal disease              | paralisis                 |
| Laboratory data            | Elevated serum calcium    |
|                           | Elevated alkaline         |
|                           | phosphatase               |
|                           | Positive antineutrophil   |
|                           | cytoplasm antibody        |
### Evaluation of Patient with Sarcoidosis

#### Initial Evaluation Suggested for All Patients
- History, including occupational and environmental exposures
- Physical examination
- Posterior-anterior chest roentgenogram
- Spirometry
- Complete blood count
- Liver function studies and serum calcium
- Routine ophthalmic examination

#### Evaluation Considered in Selected Patients
- Computed tomography scan of chest
- Holter monitor and/or electrocardiogram
- Urine analysis
- Diffusion capacity of lung of carbon monoxide
- X-ray of involved joints and/or ultrasound/magnetic resonance imaging

#### Follow-up Every 6-12 Months
- Any abnormality noted on initial evaluation
- Chest roentgenogram
- Liver function studies and serum calcium
- Spirometry


---

Sarcoid Arthritis

Between 15% and 25% of patients with sarcoidosis have arthritis, which is the most frequently observed rheumatologic symptom of sarcoidosis. In patients with sarcoidosis, arthritis can be acute or chronic; the acute form is most common. Chronic arthritis is typically associated with multisystem sarcoidosis.

**Acute Sarcoid Arthritis**

Febrile arthropathy is the most frequently observed joint involvement in sarcoidosis. Although any joint may be involved, the condition usually is symmetric, involving ankles, knees, wrists, and elbows, and exists with BHL and erythema nodosum. Patients with acute sarcoid arthritis exhibit pain and stiffness, and their joints may be swollen or tender (Supplemental Figure 117-1 on www.expertconsult.com). Acute sarcoid arthritis sometimes resolves within weeks of onset, but occasionally symptoms last for several months. Once it resolves, this condition usually does not recur.

In a study of patients with acute sarcoid arthritis, Visser and colleagues prospectively evaluated patients and published criteria to assist in diagnosis. Of 579 participants, 55 (9%) patients were eventually diagnosed with sarcoid arthritis. Diagnoses were made after it was established that patients had a combination of arthritis and BHL, as determined by chest radiography. From the findings of their study, investigators established criteria with 93% sensitivity and 99% specificity to guide physicians in differentiating between sarcoid and other causes of arthritis. Of the following criteria, patients must have three out of four characteristics to establish a diagnosis of sarcoid arthritis: younger than 40 years of age, EN, symmetric ankle arthritis, and symptoms lasting less than 2 months.

In addition to Visser’s criteria, most symptoms of sarcoid arthritis vary among patients. Patients with sarcoid arthritis frequently have elevated erythrocyte sedimentation rates, but other symptoms, such as fever, are observed only in up to 66% of patients. Of note, a small minority of patients with acute sarcoid arthritis may develop abnormal alterations in their bones that can be seen on radiographs of the hands and feet. In some patients, sarcoid dactylitis may be the presenting symptom (see Figure 117-1C and E), characterized by swollen soft tissues surrounding affected fingers, erythematous skin, or tenderness. Individuals with dactylitis may also have nail abnormalities, such as dystrophy. Therefore, it is important to rule out other conditions that affect the nails, such as psoriatic arthritis, when making sarcoid arthritis diagnoses.

Other bone areas that may be affected by sarcoidosis include nasal bones, pelvic girdle structures, ribs, and skull. Patients with lupus pernio may be particularly at risk for developing abnormalities in their nasal bones. Because of similarities between a patient with sarcoidosis who has nasal bone involvement and a patient with granulomatosis with polyangiitis (GPA) (formerly Wegener's granulomatosis), it is important to distinguish between these conditions. Serology may be useful because most patients with GPA will have a positive antineutrophil cytoplasm antibody (ANCA), but sarcoidosis does not lead to a positive ANCA. On the other hand, about 60% of sarcoidosis patients will have an elevated ACE level at presentation; this is not seen in the presence of vasculitis. Lesions in the pelvic bones (see Figure 117-1D) or in the spine (see Figure 117-1F) may, at first, lead one to suspect cancer metastases. Therefore, a comprehensive assessment is warranted in patients who present with changes in their skeletal bones, even if they exhibit other symptoms reminiscent of sarcoidosis.
Table 117-4  Features Comparing Sarcoidosis with Other Rheumatologic Diseases

<table>
<thead>
<tr>
<th>Clinical Findings</th>
<th>Sarcoidosis</th>
<th>SLE</th>
<th>RA</th>
<th>PsA</th>
<th>Systemic Vasculitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lungs, eyes, and skin most common; any other organ (heart, brain, skin, bone, liver, etc.) may also be affected</td>
<td>Joints, kidneys (nephritis), mucous membranes, circulatory system, nervous system, lymph nodes, spleen</td>
<td>Joints (most common), lungs, blood vessels</td>
<td>Joints (oligoarthritis, polyarthritis), skin (psoriatic lesions), nails (dystrophy, dactylitis)</td>
<td>Blood vessels (wall thickening), skin (purpura, infarct ulcers), nervous system (headache, meningitis, seizures), joints (arthritis), kidneys (hypertension), heart, GI system</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Radiographic Findings</th>
<th>Bilateral and mediastinal hilar adenopathy, reticulonodular opacities</th>
<th>Bilateral, diffuse air space opacity; diaphragm elevation (“shrinking lung syndrome”)</th>
<th>Joint space narrowing, joint erosions, inflammation</th>
<th>Joint space narrowing, erosions, ossification near joints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noncaseating granulomas</td>
<td>Inflammation, blood vessel abnormalities such as vasculitis, immune complex deposits</td>
<td>Swollen synovium; presence of fibroblast-like and macrophage-like synovocytes, macrophages, T and B cells</td>
<td>Increased synovial vascularity, dilated blood vessels, neutrophil infiltration</td>
<td>Granulomatous lesions; multiple bilateral nodules; patchy areas of consolidation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pathologic Findings</th>
<th>Elevated ACE; hypercalcemia; presence of ANAs, ANCs, and anti-dsDNA</th>
<th>Anti-dsDNA, anti-Smith, ANAs, RNP, antithrombin, anti-Ro/SSA, anti-La/SSB, anti-topoisomerase, aPL</th>
<th>Rheumatoid factor, ACPA, aPL</th>
<th>ANCA staining pattern on neutrophils and monocytes; patchy infiltrates, vessel wall granulomas, fibrous tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laboratory Findings</td>
<td>Skin (tightening, thickening, induration), joints, GI system, lung, heart, kidney</td>
<td>Eyes, mouth, and mucous membrane dryness; joints (arthritis), skin, kidneys, nervous system, lymph nodes may be affected</td>
<td>Mild joint space narrowing</td>
<td>Arthritis, conjunctivitis; uveitis; usually following genitourinary or gastrointestinal infection</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MCTD</th>
<th>Raynaud’s phenomenon; joints, muscles, lungs, heart, kidneys, and nervous system may be affected</th>
<th>Skin (tightening, thickening, induration), joints, GI system, lung, heart, kidney</th>
<th>Blood vessels (arterial/venous thrombosis, thrombocytopenia, fatal loss)</th>
<th>Vertebral arthritis, back pain, stiffness, synovitis, peripheral arthritis, pulmonary symptoms, iritis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Diffuse periarticular osteopenia; swelling of soft tissue; joint erosions, joint space narrowing, tuft resorption, and soft tissue atrophy</td>
<td>Pulmonary fibrosis, diffuse reticulonodular pattern</td>
<td>Patchy infiltrates</td>
<td>Vertebral inflammation; sacroiliitis; bone erosions, syndesmophytes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Antiphospholipid Syndrome</th>
<th>Structural glomerular changes</th>
<th>B cell infiltration</th>
<th>Capsular fibrosis; ossification</th>
<th>Proliferation at tendon insertion; sacroilitis; syndesmophytes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANA (anti-topoisomerase), aPL, anti- C, anti-prothrombin, anti-protein S, anti-annexin</td>
<td></td>
<td></td>
<td>Bacterial infection (Shigella, Salmonella, Campylobacter)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sjögren’s Syndrome</th>
<th>Ankylosing Spondylitis</th>
<th>Reactive Arthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eyes, mouth, and mucous membrane dryness; joints (arthritis), skin, kidneys, nervous system, lymph nodes may be affected</td>
<td>Vertebral arthritis, back pain, stiffness, synovitis, peripheral arthritis, pulmonary symptoms, iritis</td>
<td>Arthritis, conjunctivitis; uveitis; usually following genitourinary or gastrointestinal infection</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ankylosing Spondylitis</th>
<th>Leakage of inflammatory markers (CRP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild joint space narrowing</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Autoantigen modifications, B and T cell activation</th>
<th>Structural glomerular changes</th>
<th>B cell infiltration</th>
<th>Capsular fibrosis; ossification</th>
<th>Proliferation at tendon insertion; sacroilitis; syndesmophytes</th>
</tr>
</thead>
</table>

| | Excess matrix deposition, fibrosis; endothelial cell dysfunction and death; destruction of small vessels | Structural glomerular changes | | |
| | Lupus anticoagulant, aPL, anti–protein C, anti–protein S, anti–annexin | ANA, anti-Ro/SSA, rheumatoid factor, HLA-B27 | Elevated ESR, leukocytosis, HLA-B27 | ANCA, HLA-B27; elevated ESR, CRP |

ACE, angiotensin-converting enzyme; ACPA, anticitrullinated protein antibody; ANA, antinuclear antibody; ANCA, antineutrophil cytoplasm antibody; aPL, antiphospholipid antibody; CRP, C-reactive protein; dsDNA, double-stranded DNA; ESR, erythrocyte sedimentation rate; GI, gastrointestinal; MCTD, mixed connective tissue disease; PsA, prostate-specific antigen; RA, rheumatoid arthritis; RNP, ribonucleotide protein; SLE, systemic lupus erythematosus; TNF, tumor necrosis factor.
Figure 117-2 Algorithm describing an approach to treating sarcoidosis arthritis. *, Low dose steroids: <10 to 20 mg prednisone daily. †, Suboptimal response to MTX: intolerance to drug, lack of satisfactory efficacy on dosage up to 25 mg/wk, or a contraindication to medication use. ‡, DMARD failure: progressive disease or drug intolerance. §, Methylprednisone preferred over prednisone if prednisone has been used previously. DMARD, disease-modifying antirheumatic drug; MTX, methotrexate; NSAIDs, nonsteroidal anti-inflammatory drugs; OT, occupational therapy; PT, physical therapy. (Reproduced with permission from Sweiss NJ, Patterson K, Sawaqed R, et al: Rheumatologic manifestations of sarcoidosis, Semin Respir Crit Care Med 31:463–473, 2010.)
be excluded. For example, symmetric oligoarthritis or polyarthritis and an elevated rheumatoid factor are common features of sarcoid arthritis, but they may also result from reactive arthritis or rheumatoid arthritis (RA). Furthermore, a subset of patients with RA may develop some lung dysfunction, which can cause confusion in differentiating RA from sarcoidosis. In addition, some patients with RA have been reported to eventually develop coincident sarcoidosis, although sarcoidosis coexisting with connective disease disorders is rare. Therefore, it is crucial for physicians to distinguish between the two conditions, or to determine whether the two diseases are comorbid within the same patient.

In patients with migratory polyarthritis, rheumatic fever may be suspected, particularly when joint symptoms manifest before additional sarcoidosis-related symptoms are noted. A biopsy of the synovium or tendon sheaths may be helpful with characteristic granulomas appearing at pathology. Differential diagnoses for patients with monarthritis include gout, septic arthritis, and calcium pyrophosphate dihydrate (CPPD)-positive arthritis (see Tables 117-3, 117-4, and supportive content for detailed consideration of the diagnostic approach).^2

MANAGING SARCOIDOSIS: FOCUSING ON SARCOID ARTHRITIS

No treatments have been U.S. Food and Drug Administration (FDA)-approved for sarcoidosis or for any of its manifestations, including sarcoid arthritis. Furthermore, no randomized trials have been conducted that can guide clinical decision making. To help guide clinicians in management, we have proposed a treatment algorithm, which is shown in Figure 117-2. ^2 Nonsteroidal anti-inflammatory drugs, methotrexate, and local or low-dose systemic corticosteroids are our preferred first-line therapies. Alternatively, hydroxychloroquine may be used. Using information obtained from the diagnostic evaluation, the clinician usually makes treatment decisions after carefully considering the disease severity and its probable clinical course, as revealed by its radiographic progression.

The proposed algorithm requires regular patient visits to monitor disease activity and therapeutic efficacy and tolerability. Responders remain on first-line agents until disease resolution or, alternatively, treatment failure. Before nonresponders are prescribed second-line therapies, higher doses of corticosteroids may be used, depending on tolerability and manifest toxicity. If patients do require second-line medications, two options are available: (1) methotrexate may be given to methotrexate-naïve patients, or (2) biologic therapies consisting of nonmethotrexate disease-modifying antirheumatic drugs (e.g., sulfasalazine, hydroxychloroquine, azathioprine), monotherapy, or combinations thereof may be prescribed to patients who inadequately respond to first-line methotrexate.

For patients for whom this treatment approach fails, one can consider an alternative biologic therapy or an aggressive course of systemic corticosteroids with careful toxicity monitoring. Alternatively, participating in a clinical trial may be appropriate for some patients.

Vitamin D metabolism represents a complex issue in sarcoidosis. Up to 10% of sarcoidosis patients will have hypercalcemia or hypercalciuria. ^52 The mechanism usually attributed has been increased production of 1,25-dihydroxyvitamin D (1,25-OH2D) by epithelioid cells in the granuloma. In one study, elevated 1,25-OH2D was associated with prolonged need for treatment. ^53 However, this same group of patients often requires treatment for osteoporosis. ^49 Bisphosphonates alone may be adequate to treat corticosteroid-induced osteoporosis. ^55 Because of the disassociation between 25-hydroxyvitamin D (25-OHD) and 1,25-OH2D in sarcoidosis, it seems reasonable to measure both levels to ascertain which patients should receive vitamin D supplements.

**Future Directions**

Much remains to be learned about sarcoidosis. Its causes remain unclear, and no medications have been FDA-approved for its treatment. Although increasing attention has been paid to the underlying mechanisms of granuloma formation, full details of sarcoidosis immunopathogenesis have yet to be determined. Appropriate animal models and candidate genes are needed to help advance our understanding of this disease.

Large clinical trials are warranted. We have presented an algorithm for use in treating patients with an established diagnosis of sarcoidosis. However, treatment approaches vary by institution and by individual clinician owing to a myriad of conflicting studies that have been published about sarcoidosis management. Furthermore, therapeutic choices will likely differ according to the type and extent of organ system involvement observed in individual patients.

For relapsed and refractory disease, steroid-sparing agents including cytotoxic drugs and novel biologic therapies such as anti-TNF treatments have been used increasingly. Anti-TNF agents have been investigated in numerous sarcoidosis studies because of the potential role of TNF and other proinflammatory factors in sarcoidosis pathogenesis. Limiting their usefulness in this setting are mounting reports of granulomatous reactions to anti-TNF therapies in patients treated with these agents for nonsarcoidosis indications. Simply put, TNF inhibitors may help treat and may cause sarcoidosis. ^56-59 With this point in mind, it is relevant to our discussion that the presentation of anti-TNF–induced sarcoidosis, similar to all phenotypes of sarcoidosis, is a unique entity, but it does overlap with that of other autoimmune diseases. Furthermore, anti-TNF agents have been reported to induce autoimmune diseases other than sarcoidosis, including systemic lupus erythematosus (SLE), vasculitis, and interstitial lung disease. Diagnosing these conditions in patients with anti-TNF–induced sequelae is critical.

**References**


30. Deleted in proofs.
Hemochromatosis refers to the presence of excess iron in body tissues because of increased iron absorption. Primary or hereditary hemochromatosis (HHC) is an autosomal recessive disease, whereas secondary hemochromatosis refers to iron overload as a result of increased iron availability, ineffective erythropoiesis, or inherited abnormalities of iron metabolism.

Hemochromatosis was first recognized in the 1880s in a series of case reports that described “bronze diabetes” and “pigmented cirrhosis,” but von Recklinghausen is credited with the first use of the term in 1889. In 1935, the familial pattern of HHC was described by Sheldon, who suggested that the disease was due to an inborn error of metabolism. Finch and Finch in 1955 showed that HHC was caused by abnormal iron absorption in the presence of a normal diet. At that time, premorbid recognition of the problem was uncommon, however, and most cases were diagnosed at autopsy. In 1972, serum ferritin became available as a measure of iron stores. Three years later, Simon and colleagues discovered that the HHC gene was present on chromosome 6, close to the HLA-A locus. It took 21 more years for the mutated gene, HFE, to be described, and in the last decade, it has been recognized that other gene mutations also can cause iron overload. Genetic testing has revolutionized the diagnosis of HHC, although the phenotype of any given mutation may vary greatly. Nevertheless, the detection of such genes has greatly enhanced the overall prognosis of this condition by allowing the disease to be diagnosed at a preclinical stage in high-risk individuals. This discovery has helped many patients with HHC achieve a normal life expectancy. Improved knowledge at a molecular level has further aided our understanding of the pathogenesis and therapeutic implications of hemochromatosis, permitting many patients with this disease to achieve a normal life expectancy.

### Normal Iron Metabolism

The average total body iron content in adults is 3 to 4 g, mostly contained within hemoglobin, but also present in myoglobin and cytochromes, in addition to the storage proteins ferritin and hemosiderin (Table 118-1). Of a typical daily Western diet of 10 to 20 mg of iron, 1 to 2 mg is absorbed by duodenal enterocytes each day. Heme dietary sources from fish and meat have a higher bioavailability than nonheme sources, such as vegetables. The addition of ascorbic acid to the meal increases absorption of nonheme iron, whereas tannins, bran, and phytates inhibit iron absorption.

Iron homeostasis is tightly controlled at cellular and molecular levels, influenced by numerous mechanisms, including recent dietary iron intake, the extent of iron stores in the body, and key regulatory peptides. Communication between sites of iron uptake (enterocytes), storage (liver and macrophages), and utilization (erythroid cells) is essential, and an antimicrobial peptide, hepcidin, plays a key role in this regard. Hepatic synthesis of hepcidin is stimulated by increases in the body’s iron requirements, as in situations of anemia, hypoxia, or inflammation. Hepcidin prevents iron loss by reducing the entry of iron into the bloodstream via inhibition of ferroportin, a membrane-bound iron exporter protein found on macrophages, hepatocytes, and enterocytes. Hepcidin production is downregulated when iron requirements return to normal. In the presence of normal iron stores, iron is retained in the intestinal cells by the protein, mobilferritin, and is subsequently excreted when these cells are shed.

When body iron stores reach an adequate level, ferritin production is increased to facilitate storage, and the transferrin receptor is downregulated to minimize the entry of iron into the cells. The iron responsive element binding protein mediates this process by detaching from ferritin mRNA so that more ferritin can be produced. With increasing iron stores, circulating transferrin becomes saturated, and iron is preferentially offloaded to tissue sites.
that contain cells with high levels of transferrin receptors, such as liver, heart, thyroid, gonads, and pancreatic islet cells.\textsuperscript{17}

**GENETICS OF HEMOCROMATOSIS**

Four types of HHC have now been described, all linked to gene mutations (Table 118-2).\textsuperscript{18,19} Classic HHC (type 1) is an autosomal recessive disorder, with a mutation of the HFE gene, located on chromosome 6 (HFE-related HHC). Although numerous such mutations have been described, the most common is a single amino acid substitution of tyrosine for cysteine at position 282 (C282Y). This particular mutation is thought to have arisen in a Celtic/Viking ancestor more than 2000 years ago and is now one of the most common genetic defects in individuals of Northern European origin. This anomaly had no reproductive implications, but may have had survival advantages by protecting against iron deficiency in a susceptible population. Homozygosity for this mutation is a risk factor for organ damage secondary to iron deposition, although phenotypic expression varies widely. Other mutations of the HFE gene include the replacement of histidine with aspartic acid at position 63 (H63D) and the substitution of serine for cysteine at position 65 (S65C). Clinical manifestations of the latter mutations seem to be less serious, although compound heterozygosity of such defects may be associated with evidence of iron overload.

Unlike mutations of the HFE gene that may become clinically obvious in middle age, hemojuvelin (HJV-related HHC, type 1A) or hepcidin (HAMP-related HHC, type 2B) mutations result in juvenile HHC (type 2), which may manifest in the teens or twenties. The rate of iron accumulation seems to be greater than in adult HHC and is often associated with widespread organ involvement and early mortality.\textsuperscript{20} In contrast to the Northern European inheritance of HFE mutations, juvenile HHC has been most commonly reported in Italy.\textsuperscript{21} Clinical manifestations of transferrin receptor mutations (TIR2-related HHC, type 3) seem to resemble manifestations of the classic HFE-related HHC. Such mutations are rare, and few cases have been described.\textsuperscript{22,23}

Type 4 or ferroportin-related HHC is an autosomal dominant condition, described in European and Australian families.\textsuperscript{24,25} Two types of ferroportin mutations have been reported. In the first, loss of surface localization of ferroportin results in a decreased ability of cells to export iron, causing iron to build up predominantly in macrophages. In the second, hepcidin-induced ferroportin dysfunction leads to iron accumulation in parenchymal cells of the liver and other tissues. Phenotypic expression varies, with some patients manifesting the effects of iron overload in a similar manner to classic HHC, and others showing minimal evidence of organ damage.\textsuperscript{26}

**EPIDEMIOLOGY**

Although HHC previously was thought to be a rare condition, genetic testing has revealed that it is one of the most common heritable disorders. Although 5 out of every 1000 individuals of Northern European origin are homozygous for the HFE mutation, phenotypic expression varies, and clinical cases are much fewer in number. In a study of nearly 100,000 individuals from primary care practices in the United States, the prevalence of C282Y homozygosity was as follows: white, 0.44%; Native American, 0.11%; Hispanic, 0.027%; African-American, 0.014%; Pacific Islander, 0.012%; and Asian, 0.0004%.\textsuperscript{27} Peak age at the time of diagnosis is 40 to 60 years for classic HHC.
PHENOTYPIC DISEASE EXPRESSION

Clinical manifestations of HHC vary greatly between individuals with similar mutations, suggesting that other factors influence disease expression. One study demonstrated that among C282Y homozygotes, up to 82% have hyperferritinemia, while approximately 28% of male and 1% of female individuals ultimately develop clinical manifestations of HHC—defined as the presence of liver disease, hepatocellular carcinoma, and arthritis of the second and third metacarpophalangeal joints—by the age of 65. In contrast, compound heterozygotes with C282Y/H63D mutations exhibit higher serum ferritin and transferrin saturation levels compared with normal controls, but are at very low risk of clinical HHC. Other genes may play a role in modifying the phenotypic expression of iron overload. The presence of the gene CYBRD1, which encodes duodenal reductase DCYTB, has been shown to be associated with lower serum ferritin levels in C282Y homozygotes. Furthermore, mutations of other iron-related genes, such as hepcidin, hemojuvelin, haptoglobin, and bone morphogenetic protein, may influence disease manifestations. In addition, profibrotic genes (e.g., TGF) may accelerate the onset of cirrhosis in susceptible individuals.

Why HHC disease penetrance is more evident in C282Y homozygote males than females may be explained by recurrent menstrual blood loss and consequent slower accumulation of iron stores in women. However, genetically determined sex differences in ferritin levels may occur, as distinct HLA A*03B*07 and A*03B*14 haplotypes have been reported in men and women with clinical evidence of HHC.

Environmental factors, including diet, smoking, alcohol intake, and comorbid diseases, also influence clinical expression of HHC. The metabolic syndrome is associated with insulin resistance–associated iron overload, which, in the presence of HHC, may have a synergistic effect on liver damage. Concurrent liver disease, due to hepatitis or steatosis, may exacerbate the process of fibrogenesis. Excess alcohol, meat consumption, and high citrus fruit intake also contribute to increased iron loading. However, ingestion of noncitrus fruits may have a protective effect.

PATHOGENESIS

Hepcidin is a key regulatory peptide in the pathogenesis of HHC. Produced by the liver, it acts by binding to ferroportin on enterocytes and macrophages, thereby restricting dietary iron absorption from the gut and inhibiting release of iron recycled by macrophages from aging red cells. In HHC, inadequate hepcidin synthesis leads to increased intestinal iron absorption and the subsequent deposition of iron in tissues. Absence of hepcidin results in early, severe iron loading, and overexpression of this protein can significantly improve iron deposition in a mouse model of HHC.

Chronic iron overload is thought to cause tissue damage via several mechanisms, including weakening of lysosomal membranes and consequent discharge of enzymes into the cytoplasm. Increased free radical formation contributes to lipid peroxidation of cell membranes. The extent and duration of iron deposition correlate with the development of fibrosis, and it is thought that substantial hepatocyte and Kupffer cell iron accumulation precedes organ damage. In HHC, iron first accumulates in parenchymal cells, with reticuloendothelial (RE) involvement a late feature, in contrast to transfusional iron overload, in which RE cells are primarily targeted. Values of serum ferritin exceeding 1000 µg/L are associated with significantly increased risks of liver fibrosis and cirrhosis.

CLINICAL FEATURES

Extra-articular Manifestations

HHC is more common in men than in women and typically manifests in middle-aged adults as iron stores gradually accumulate, often reaching 20 to 30 g. Organ involvement varies and is unpredictable, although the liver, as the major site of iron storage, is typically affected. Commonly, abnormalities of the liver enzymes, checked as part of a routine health screen, are the initial indication of disease. The degree of iron overload has a direct impact on the life expectancy of the affected individual. Without an early diagnosis, progressive fibrosis leading to cirrhosis may occur. The risk of hepatocellular carcinoma is greatly increased in patients with established cirrhosis.

Glucose intolerance tends to be a late finding in HHC and is due to progressive iron accumulation in pancreatic beta cells causing low C-peptide and insulin levels. Alpha cell function is usually preserved, however, and serum glucagon levels are normal or increased. The risk of diabetes mellitus also is higher in C282Y heterozygotes with no clinical evidence of HHC compared with controls.

Iron deposition in the heart can result in conduction system abnormalities and heart failure. Several large population studies of HHC and atherosclerosis have failed to find a link. However, elevated ferritin levels, particularly in the setting of nonalcoholic fatty liver disease, may be associated with vascular damage via hepcidin upregulation.

Pituitary involvement in HHC is due to iron deposition, resulting in reduced serum levels of secreted hormones from this gland. Low levels of gonadotropin hormone cause loss of libido and erectile dysfunction. Hypothyroidism in HHC is thought to be due to a direct toxic effect of iron on thyroid cells and is associated with low thyroxine and elevated thyroid-stimulating hormone. Such endocrine abnormalities may contribute to the development of osteoporosis in these individuals.

Skin discoloration occurs as a result of extra melanin and iron in the epidermis. It is a late finding, and the development of “bronze diabetes” represents the end stage of years of iron accumulation in the tissues.

Patients with HHC have increased susceptibility to certain infections. High serum iron concentrations may increase bacterial virulence, whereas excess iron in macrophages is thought to reduce phagocytosis. Particular caution is advised with uncooked seafood because of the risk of septicemia from Vibrio vulnificus. In addition, Yersinia enterocolitica, Listeria monocytogenes, Salmonella enteritidis serotype typhimurium, Klebsiella pneumoniae, Escherichia coli,
Rhizopus arrhizus, and Mucor species all have been reported to cause severe illness in patients with iron overload.⁸

Apart from the increased risk of malignant hepatoma in patients with established cirrhosis is an independent association of nonhepatic cancers, particularly breast and colorectal tumors, in HFE C282Y homozygotes.⁵⁵ H63D homozygosity confers a threefold increased risk of colorectal cancer in carriers of the MMR gene mutation.⁵⁶ Iron is potentially carcinogenic via several mechanisms, including its immunosuppressive properties and its role as an essential cofactor for tumor cell growth and in catalyzing the formation of hydroxyl radicals.⁵⁵ Furthermore, cancer risk is lower with reduced iron stores.⁵⁷

**Articular Features**

Arthralgia/arthritis is a common presentation in HHC, affecting 50% to 80% of patients and significantly interfering with quality of life.⁵⁸-⁶³ Although it tends to be a late feature, joint pain may nevertheless be the presenting symptom of HHC, alerting a diligent physician to the presence of an underlying metabolic disorder. Articular involvement may be widespread, but changes to the second and third metacarpophalangeal joints are most characteristic (Figures 118-1 and 118-2). Arthritis may be present in the proximal interphalangeal joints, wrists, shoulders, hips, knees, and ankles.⁶¹ Patients notice pain and stiffness of the involved joints, but evidence of synovitis is usually absent. Hip damage develops in approximately 25% of individuals with HHC, and after hip arthroplasty, the risk of aseptic loosening of the prosthesis is increased.⁶⁵,⁶⁶ The differential diagnosis of HHC-related arthropathy includes severe osteoarthritis, rheumatoid arthritis, other forms of inflammatory arthritis, and crystal arthritis. Rheumatoid factor is typically negative, however, and radiographs, in established cases, show distinctive findings, such as joint space narrowing of the second and third metacarpophalangeal joints, hook-like osteophytes on the radial aspect of the metacarpal heads, and chondrocalcinosis, particularly of the triangular
are thought to be due to excessive ferritin production within inherited in an autosomal dominant fashion. The cataracts hyperferritinemia-cataract syndrome involves several mutations. It is thought that a polymorphism of the ferroportin 1 gene is involved. Genetic factors contribute to the development of disease. It is proposed that increased iron absorption and elevation of tissue and serum ferritin levels. Chronic liver disease resulting from these changes are accompanied by low transferrin saturation. Interleukin-6 is a potent stimulator of hepcidin via the STAT3 pathway, resulting in reduced iron absorption and elevation of tissue and serum ferritin levels. However, these changes are accompanied by low transferrin saturation. Chronic liver disease resulting from hepatitis, alcohol excess, or fatty infiltration is associated with hyperferritinemia and normal or elevated transferrin saturation values. Alcohol, independent of liver damage, induces ferritin synthesis.

**DIFFERENTIAL DIAGNOSIS**

Because no physiologic mechanism is in place to increase iron excretion, the inevitable result of increased iron entry into the body is iron overload. Excess iron may accumulate because of high intake of iron by oral or parenteral means. A full history of medication use, including over-the-counter iron tablets, and a review of any blood transfusions should be elicited. Ineffective erythropoiesis in conditions such as the thalassemias or sideroblastic anemia also results in the buildup of iron stores. Porphyria cutanea tarda is associated with hyperferritinemia, high transferrin saturation levels, and an increased incidence of HFE mutations. However, the clinical findings, particularly dermatologic, distinguish this disease from HHC. Elevated ferritin levels may result from severe inflammation. Interleukin-6 is a potent stimulator of hepcidin via the STAT3 pathway, resulting in reduced iron absorption and elevation of tissue and serum ferritin levels. However, these changes are accompanied by low transferrin saturation. Chronic liver disease resulting from hepatitis, alcohol excess, or fatty infiltration is associated with hyperferritinemia and normal or elevated transferrin saturation values. Alcohol, independent of liver damage, induces ferritin synthesis.

Genetic factors not related to HHC have been linked to other iron overload syndromes. The African iron overload syndrome occurs in a few Africans who drink locally brewed beer containing extremely high levels of iron (80 mg/L). Not all Africans who drink this beer develop hemochromatosis, leading to suggestions that additional genetic factors contribute to the development of disease. It is thought that a polymorphism of the ferroportin 1 gene is involved. Separately, a familial association has been noted with a syndrome of very high ferritin levels (>1000 ng/mL) and bilateral congenital cataracts. This hereditary hyperferritinemia-cataract syndrome involves several mutations in the iron responsive element of t-ferritin and is inherited in an autosomal dominant fashion. The cataracts are thought to be due to excessive ferritin production within the lens fibers.

**INVESTIGATIONS**

A high index of suspicion is helpful when a patient presents with joint pain and abnormal liver enzymes. Although the differential diagnosis is wide, the presence of elevated ferritin and transferrin saturation levels (serum iron x 100/total iron-binding capacity) strongly points to the answer. Serum iron should be measured with the patient fasting because concentrations may be increased after a meal. High ferritin levels also may be caused by systemic inflammation or malignancy, but these conditions tend to be associated with reduced transferrin saturation. Other causes of elevated transferrin saturation include high serum iron secondary to hepatic cytolysis or low transferrin levels secondary to liver failure, and these possibilities should be excluded. If ferritin measures greater than 200 µg/L, and transferrin saturation is greater than 45%, genetic screening is recommended. The finding of homozygosity for the C282Y mutation or compound heterozygosity for C282Y/H63D confirms the diagnosis.

Liver biopsy may be considered for prognostic purposes in established cases. HHC can be distinguished histologically from alcoholic cirrhosis by the preferential distribution of iron in the hepatocytes in the former and in the Kupffer cells in the latter. Magnetic resonance imaging of the abdomen also can be used to determine iron overload in the internal organs. Gradient T2-weighted sequences show decreased signal intensity and correlate highly with liver iron concentrations. This imaging method also can identify other locations of iron deposition (e.g., in the spleen, pancreas, lymph nodes, and heart). Because HHC is a systemic condition, other investigations should include a search for diabetes, thyroid disease, hypogonadism, osteoporosis, and cardiomyopathy. Disease mimickers, such as porphyria cutanea tarda, ineffective erythropoiesis, and chronic alcohol excess, should be excluded.

**SCREENING**

Greater disease awareness and the availability of genetic screening have meant that HHC is increasingly likely to be diagnosed before the classic triad of cirrhosis, diabetes, and skin hyperpigmentation develops. Late presentation with evidence of end-organ damage does occur, however, particularly in patients with additional risk factors for iron overload or liver disease. HHC is an attractive clinical target for population screening because of its high prevalence, potential disease severity, availability of effective treatment, and impact of early diagnosis on the morbidity and mortality of affected individuals. Certain groups are more at risk than others, however, and the disease prevalence is higher in white than in nonwhite individuals. Biochemical measures, such as serum ferritin, may serve as a cost-effective method of screening in whites during routine health checks and in individuals who complain of nonspecific symptoms, such as excessive fatigue and arthralgias. Elevation of serum ferritin in the absence of other causes should prompt measurement of transferrin saturation. Levels greater than 45% in men and greater than 35% in premenopausal women, without adequate explanation, warrant further investigation.
Genetic testing should be reserved for patients with suggestive biochemical abnormalities or a family history of HHC. Routine population screening for C282Y or H63D mutations is not recommended because of the variable clinical penetrance of these genes and the potential negative consequences of a positive result in asymptomatic patients, such as financial, legal, insurance, and psychological implications.\(^8\) When a case of HHC is diagnosed, however, and two gene mutations are identified (i.e., C282Y/C282Y or C282Y/H63D), siblings also should be tested for these mutations. H63D/H63D homozygotes are not thought to be at risk of clinical disease. Children of a patient with HHC or of an individual with C282Y/H63D heterozygosity are at risk only if the other parent also carries hemochromatosis gene mutations.

For individuals in whom genetic testing has identified a risk of HHC, but with no clinical evidence of disease, yearly biochemical screening should be done, with measures of ferritin, transferrin saturation, and liver enzymes. Such monitoring allows early detection of organ compromise and timely initiation of treatment (see Figure 118-2).

**MANAGEMENT**

Removing excess iron before the development of organ damage significantly abrogates the adverse consequences of HHC. Target groups for treatment include asymptomatic individuals with biochemical evidence of high iron stores, in addition to patients with overt clinical disease. Some features of HHC improve with bloodletting, including constitutional symptoms, diabetes, and liver enzyme abnormalities. Phlebotomy has no effect, however, on arthritis, hypogonadism, and liver fibrosis.\(^8\) When cirrhosis is established, the risk of hepatocellular carcinoma is greatly increased, even after a satisfactory reduction in iron stores.\(^8\)

Phlebotomy is an effective method of removing excess iron. The use of chelating agents is rarely necessary. Every 500 mL of whole blood contains 200 to 250 mg of iron, depending on the hematocrit. Phlebotomy can be arranged once or twice weekly, as tolerated by the patient, aiming for serum ferritin of 50 ng/mL. It can take longer than 1 year for iron stores to normalize with this regimen. Iron deficiency anemia should always be avoided, and when ferritin levels reach their target, the frequency of bloodletting may be reduced. Transferrin saturation levels are not an accurate measure of therapeutic efficacy because they are relatively resistant to changes in iron stores in C282Y homozygotes.\(^8\)

Phlebotomy continues for life, and the maintenance schedule depends on the patient’s ability to sustain the ferritin level in the low-normal range. It is important to avoid very low ferritin levels because this situation may increase iron absorption via further reduction in hepcidin levels or increased erythropoiesis in C282Y homozygotes.\(^8\) Blood removal in HHC is not without risks. In particular, life-threatening cardiac arrhythmias may develop during rapid mobilization of iron stores. Vitamin C supplementation may precipitate such problems by facilitating iron release and increasing pro-oxidant and free radical activity.\(^18\) Patients undergoing phlebotomy for HHC should not take extra vitamin C, but can continue to eat fresh produce containing this vitamin.

Other dietary recommendations include reduction or avoidance of food containing high doses of iron, such as red meat and internal organs. Uncooked shellfish is a particular hazard because of the risk of contamination with *V. vulnificus*. Some alcoholic drinks contain iron, and all are potentially hepatotoxic. Alcohol should be consumed only occasionally because it seems to have a synergistic effect in the presence of iron overload on the development of cirrhosis and hepatocellular carcinoma.\(^2,3\) Maintenance of normal body weight is important to avoid the hepatic damage associated with steatosis.\(^8\)

Just as the pathogenesis of joint pain in HHC is unclear, treatment of arthritis in this condition is unsatisfactory, and joint symptoms may progress despite effective phlebotomy. Nonsteroidal anti-inflammatory drugs, colchicine, and intra-articular corticosteroids may be helpful in some cases. Osteoporosis is a potential disease complication, particularly in the setting of hypogonadism or reduced thyroid function. Hormone replacement, if indicated, should be instituted, although some patients may require additional treatment with calcium and bisphosphonates. Screening for malignancy, particularly for carcinoma of the breast and colon, should be considered in the holistic care of patients with HHC.

**OUTCOME**

The earlier HHC is diagnosed, the better the prognosis, because morbidity and mortality are directly related to the extent of iron overload and consequent organ damage. The development of cirrhosis is a serious indicator of reduced longevity. For patients with HHC-related hepatic failure who undergo liver transplantation, survival rates are lower compared with individuals who receive liver transplants for other reasons. Postoperative death in these circumstances is typically due to cardiac complications or infection.\(^15,84\)

In the absence of cirrhosis or diabetes, patients with HHC have a normal life expectancy. Given the importance of timely recognition of this common metabolic problem, a vigilant physician can make an enormous difference in the lives of patients who present with early symptoms of this disease. In this context, the rheumatologist has a particularly relevant role in keeping a high index of suspicion for the diagnosis of HHC in patients with atypical osteoarthritis or chondrocalcinosis.

### Future Directions

Cost-effective screening focused on those at greatest risk of clinical disease, in association with a high level of diagnostic suspicion in patients with relevant symptoms and signs, should reduce the number of patients presenting with irreversible organ damage from HHC. In addition, greater understanding of the genetic and environmental modifiers of clinical disease expression may help to delay or mitigate the consequences of iron overload in susceptible individuals. How comorbid disease influences the onset and extent of disease is particularly relevant in the context of the growing numbers of patients with obesity, metabolic syndrome, and hepatic steatosis. Although phlebotomy has been the mainstay of treatment for many years, exact targets for iron stores...
and ferritin levels warrant further investigation. Other commonly used medications, such as proton pump inhibitors, which inhibit absorption of nonheme iron, or calcium channel blockers, which may have a role in reducing iron overload via divalent metal transporter-1, have a potentially therapeutic benefit in HHC.

Currently at the experimental stage is the concept of hepcidin peptides or hepcidin agonists, aimed at correcting the physiologic deficit linked to increased iron absorption and deposition.

References

KEY POINTS

Severe hemophilia, if not aggressively treated, is often complicated by recurrent hemarthrosis.

Recurrent hemarthrosis causes chronic arthropathy with overlapping clinical and pathologic features of osteoarthritis and rheumatoid arthritis.

Septic arthritis should be considered in hemophilic patients with risk factors (previous arthrocentesis, intravenous drug use, human immunodeficiency virus infection) and acute monoarticular arthritis.

Soft tissue and muscle hemorrhage are frequent complications of hemophilia.

With continuous factor infusion, surgical procedures, including total joint replacements, can be done safely in hemophilic patients.

The best treatment for hemophilic arthropathy is prevention of recurrent hemarthrosis through regular prophylactic factor replacement.

Hemophilic Arthropathy

KATHERINE S. UPCHURCH • DOREEN B. BRETTLER

CLINICAL FEATURES

The spectrum of articular disease in hemophilic patients has been the subject of numerous comprehensive reviews and includes acute hemarthrosis, subacute or chronic arthritis, and end-stage hemophilic arthropathy. The usual distribution of joint involvement is shown in Figure 119-1. Involvement of the small joints of the hands and feet also may occur, although infrequently.

Acute Hemarthrosis

Nearly all patients with severe hemophilia A or B (<1% activity of the deficient factor) and half of patients with moderate disease activity experience hemarthrosis. Acute hemarthroses generally first occur when a child begins to walk, and they continue, usually cyclically, into adulthood, when the frequency diminishes. Patients frequently have premonitory symptoms, such as stiffness or warmth in the affected joint, followed by intense pain, which may be due in part to rapid joint capsule distention.

Pain is accompanied by objective clinical findings of warmth, a tense effusion, tenderness, limitation of motion, and a joint that is often held in a flexed position. Joint pain responds rapidly to replacement of the deficient clotting factor. If hemostasis is achieved early after onset of hemarthrosis, full joint function may be regained within 12 to 24 hours. If the hemorrhage is more advanced, however, blood is resorbed slowly over 5 to 7 days, and full joint function is regained within 10 to 14 days.

Subacute or Chronic Arthritis

Recurrent hemarthroses, particularly in patients with severe factor deficiency, may lead to a self-perpetuating condition in which joint abnormalities persist in intervals between bleeding episodes. The involved joint is chronically swollen, although painless and only slightly warm. Chronic synovitis, including prominent synovial proliferation with or without effusion, may be present. Mild limitation of motion may be noted, often with a flexion deformity. Factor replacement does not modify these findings.

End-Stage Hemophilic Arthropathy

Long-standing end-stage hemophilic arthropathy has features in common with degenerative joint disease and advanced rheumatoid arthritis. The joint appears enlarged and “knobby,” owing to osteophytic bone overgrowth. Synovial thickening and effusion are not prominent, however. Range of motion is severely restricted, and fibrous
ankylosis is common. Subluxation, joint laxity, and malalignment are frequently present. Hemarthroses decrease in frequency, however.

**Septic Arthritis**

Until the early 1980s, septic arthritis rarely occurred in hemophilic patients. With widespread occurrence of human immunodeficiency virus (HIV) infection as a result of contaminated factor concentrates, the incidence of this complication has increased significantly.\(^6,12\) Septic arthritis is seen more often in adult than in pediatric hemophilic patients and is most commonly monoarticular, usually involving the knee. In contrast to spontaneous hemarthrosis, septic arthritis is significantly associated with a temperature greater than 38\(^\circ\) C within 12 hours of presentation and articular pain that does not improve with replacement therapy.\(^11\) Peripheral leukocyte count may not be elevated, particularly in HIV-positive patients.\(^11\) A predisposing factor other than hemophilic arthropathy is often identifiable, including previous arthrocentesis or arthroplasty, intravenous drug use, and an infected indwelling venous access catheter. *Staphylococcus aureus* is the most frequently identified organism even in HIV-infected patients, followed by *Streptococcus pneumoniae*.\(^13\)

**Muscle and Soft Tissue Hemorrhage**

Bleeding into muscles and soft tissue is common in hemophilic patients and may be more insidious than hemarthrosis because of the lack of premonitory symptoms. Bleeding into the iliopsoas and gastrocnemius muscles and the forearm results in well-described syndromes with which the rheumatologist should be familiar. Iliopsoas hemorrhage produces acute groin pain with marked pain on hip extension and a hip flexion contracture. Rotation is preserved, in contrast to intra-articular hemorrhage. If untreated, the expanding soft tissue mass may compress the femoral nerve, causing signs and symptoms of femoral neuropathy.\(^5,14\) Bleeding into the gastrocnemius muscle can cause an equinus deformity from heel cord contracture.\(^7\) Finally, hemorrhage into closed compartments can cause acute muscle necrosis and nerve compression.\(^15\) Of particular importance is bleeding into the volar compartment of the forearm, which can cause flexion deformities of the wrist and fingers. If a compartment syndrome is suspected, compartment pressures should be measured to confirm the diagnosis.

A large intramuscular hemorrhage uncommonly results in the formation of a simple muscle cyst, which clinically appears to be an encapsulated soft tissue area of swelling overlying muscle. Cyst formation in this setting is confined by the muscular fascial plane and most likely results from inadequate resorption of blood and clot. Subperiosteal or intrasosseous hemorrhage, in contrast, may lead to a pseudotumor, a rare skeletal complication of hemophilia. Hemophilic pseudotumors are of two types: the adult type, which occurs proximally, usually in the pelvis or femur; and the childhood type, which occurs distal to the elbows or knees and carries a better prognosis.\(^16,17\)

Conservative early management of muscle cysts and childhood-type pseudotumors is indicated, including immobilization and factor replacement. In adult-type pseudotumors, which usually are refractory to conservative therapy, and in progressive childhood pseudotumors, surgical removal is indicated to prevent serious complications, such as spontaneous rupture, fistula formation, neurologic or vascular entrapment, and fracture of adjacent bone. Aspiration of a pseudotumor or cyst is contraindicated.

**DIAGNOSTIC IMAGING**

**Radiographs**

The earliest radiographic changes in hemophilic arthropathy are confined to the soft tissue and reflect acute hemarthrosis. The joint capsule is distended with displacement of fat pads, and an increased hazy density caused by intrarticular blood is noted. Hemarthrosis before epiphyseal plate closure may result in epiphyseal overgrowth and irregularity. Occasionally, premature epiphyseal closure is seen.

With progression of chronic proliferative synovitis, irreversible radiologic changes appear.\(^18\) These changes reflect the inflammatory and degenerative nature of chronic hemophilic arthropathy (Table 119-1 and Figure 119-2A). Certain changes unique to hemophilic arthropathy occur as well (see Table 119-1 and Figure 119-2B). Radiographic findings in hemophilic arthropathy have been recently reviewed.\(^19\)

### Table 119-1

<table>
<thead>
<tr>
<th>Percentage of joints with:</th>
<th>Any hemarthrosis</th>
<th>Many hemarthroses</th>
<th>Chronic pain</th>
<th>Synovitis</th>
<th>Limitation of motion</th>
<th>Any radiologic abnormality</th>
</tr>
</thead>
<tbody>
<tr>
<td>34.5</td>
<td>13.3</td>
<td>13.9</td>
<td>—</td>
<td>16.9</td>
<td>21.6</td>
<td></td>
</tr>
<tr>
<td>54.0</td>
<td>38.5</td>
<td>13.8</td>
<td>9.8</td>
<td>27.0</td>
<td>52.6</td>
<td></td>
</tr>
<tr>
<td>28.6</td>
<td>8.0</td>
<td>5.4</td>
<td>—</td>
<td>19.8</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>63.1</td>
<td>50.9</td>
<td>26.8</td>
<td>11.6</td>
<td>27.0</td>
<td>50.2</td>
<td></td>
</tr>
<tr>
<td>60.8</td>
<td>42.8</td>
<td>15.2</td>
<td>2.2</td>
<td>34.2</td>
<td>52.4</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 119-1** Distribution of acute hemarthrosis based on a study of 139 patients with hemophilia. Clinical and radiologic features of chronic arthritis in hemophilia. (Adapted from Steven MM, Yogarajah S, Madhok R, et al: Haemophilic arthritis, QJM 58:181, 1986.)
(the latter in patients whose radiographs showed widened intercondylar notches). As preventive measures improve, it will continue to be important to use the most advanced techniques, including MRI, both to stage hemophilic arthropathy and to assess the benefits of new therapies such as prophylactic infusion. Additionally, MRI and ultrasound are useful in the detection and quantitation of soft tissue bleeding, cysts, and pseudotumors.

**PATHOLOGIC FEATURES AND PATHOGENESIS**

Pathologic studies of human hemophilic arthropathy have been limited to synovial specimens obtained at surgery or at postmortem examination and reflect changes of advanced disease only. Studies of experimentally produced hemarthrosis in animals, post-traumatic hemarthrosis in nonhemophilic humans, and canine and murine models of hemophilia A have provided an understanding of the earliest changes induced by acute hemarthrosis and their evolution to chronic arthritis.

As reviewed more recently, the process most likely includes catabolic activation of synovial cells by exposure to blood components with subsequent cartilage destruction and a direct destructive effect of intra-articular blood on cartilage. A single synovial hemorrhage induces serial changes in the synovial membrane, including early focal villous synovial proliferation and subsynovial diapedesis of erythrocytes, followed by the appearance of perivascular inflammatory cells, patchy subsynovial fibrosis, and intracellular iron accumulation in synovial cells and subsynovial macrophages. With repeated hemorrhages, the synovium becomes grossly hypertrophied and hyperpigmented, with eventual organization into a pannus that invades and erodes marginal cartilage. On histologic examination, villous

---

**Table 119-1  Radiologic Manifestations of Chronic Hemophilic Arthropathy**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Also Seen in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periarticular soft tissue swelling</td>
<td>RA</td>
</tr>
<tr>
<td>Periarticular demineralization</td>
<td>RA</td>
</tr>
<tr>
<td>Marginal erosions</td>
<td>OA</td>
</tr>
<tr>
<td>Subchondral irregularity and cyst formation</td>
<td>RA, OA</td>
</tr>
<tr>
<td>Decreased joint space</td>
<td>OA</td>
</tr>
<tr>
<td>Osteophyte formation</td>
<td>OA*</td>
</tr>
<tr>
<td>Chondrocalcinosis</td>
<td>CPPD</td>
</tr>
</tbody>
</table>

Specific

| Femoral intercondylar notch widening | RA            |
| Squaring of distal patellar margin (lateral view) | OA            |
| Proximal radial enlargement (see Figure 119-2B) | OA            |

CPPD, calcium pyrophosphate deposition disease; OA, osteoarthritis; RA, rheumatoid arthritis.


A study of serial radiographs of symptomatic joints in hemophilic patients suggests that serial scoring with conventionally accepted techniques may be a cost-effective alternative to magnetic resonance imaging (MRI) in predicting progressive synovial hypertrophy.

**Other Imaging Methods**

MRI is now routinely used to stage hemophilic arthritis accurately to determine optimal treatment and to follow response to therapy. A scoring system based on MRI has been proposed. MRI has been demonstrated to be superior to conventional radiography and computed tomography in the detection of erosions and cysts, bone marrow edema and hemorrhage, and synovial hypertrophy and torn ligaments.
hypertrophy and subsynovial fibrosis progress, but inflammatory cells are scarce (Figure 119-3). Seventy-five percent of synoviocytes contain siderosomes (electron-dense, iron-filled deposits within lysosomes), in contrast to 10% in normal synovium and 25% in rheumatoid synovium. Iron deposits are associated with the production of proinflammatory cytokines and synovial inhibition of the formation of human cartilage matrix. Although the inflammatory synovial changes are mild, the synovial production of proinflammatory mediators, including interleukin-1 and interleukin-6 and tumor necrosis factor, approaches that of rheumatoid synovium. The articular cartilage is grossly and microscopically abnormal in the setting of recurrent hemorrhesis. Areas of cartilaginous fissuring and rarefaction expose sclerotic bone. The remaining cartilage is thin and unevenly distributed, often freely protruding into the joint cavity. Bone erosions appear at weight-bearing surfaces. Loss of matrix glycosaminoglycan occurs and is also seen in degenerative arthritis.

Current studies suggest that recurrent hemorrhesis induces joint destruction in hemophilic arthropathy through direct and indirect effects of iron on the synovium and cartilage, by the degradative effect of proliferative synovium, and through an alteration in cartilage biochemical composition similar to that seen in degenerative arthritis. A relationship between hemorrhesis-induced overexpression of oncogenes (e.g., c-myc and mdm-2) and the dysregulated, tumor-like proliferation of hemophilic synovium has been noted. Angiogenesis induced by growth factors such as vascular-derived endothelial growth factor may be important in inducing and sustaining a synovial proliferative response to hemorrhesis.

DIAGNOSIS

In most cases of congenital coagulopathy, the diagnosis has been made before presentation to a rheumatologist. In the case of hemophilia, if an affected family member is identified, prenatal diagnosis is possible. Because the spontaneous mutation rate in hemophilia is significant, the diagnosis may not be suspected until infancy, when recurrent, large ecchymoses or sustained oral hemorrhages commonly develop in most affected patients. In the case of hemophilia A or hemophilia B, hemorrhesis is usually a later manifestation, but it may be the initial symptom of other, less severe coagulopathies, even in adulthood. When a coagulopathy is suspected, baseline screening tests, including prothrombin time, activated partial thromboplastin time, and platelet count, should be performed. In patients with hemophilia, prothrombin time and platelet count are normal, and activated partial thromboplastin time is prolonged, denoting a defect in the intrinsic clotting cascade. Referral to a hematologist, who obtains the appropriate factor assays, is the next step.

Individuals with factor VIII or IX levels of 1% or less of the normal level have joint and muscle hemorrhages requiring therapy an average four or five times per month. Such patients are classified as having severe hemophilia. Individuals with factor VIII or IX levels greater than 5% of normal are considered to have mild hemophilia and usually bleed only with trauma or at surgery. Occasional "spontaneous" hemorrhesis may occur in such patients, especially in joints damaged by previously undertreated hemorrhage.

Patients whose factor VIII or IX levels fall between these two ranges are considered to have moderately severe hemophilia, and their clinical picture falls somewhere between the extremes. If such patients have had multiple untreated or suboptimally treated hemorrhoses with subsequent joint damage, the anatomic instability of these joints will cause frequent and severe bleeding, and the condition will appear clinically more severe than the factor VIII or IX assay might suggest.

TREATMENT OF HEMOPHILIA

Until recent years, factor replacement therapy has been given on demand in most hemophilia centers, that is, factor concentrate has been infused at the earliest sign of a hemorrhage. Because of the introduction of highly purified, safe concentrates, prophylactic treatment is now much more common in countries where this product is available, especially in pediatric patients. Instead of providing infusion when a hemorrhage has occurred, factor concentrate is given regularly three times per week to prevent bleeds. Prophylaxis is started before any joint damage has occurred, usually at approximately 2 years of age, with the goal of minimizing bleeding episodes to no more than four to six per year. Indwelling catheters, such as Port-A-Cath and Hickman lines, are required for factor administration because frequent venipunctures are painful and cumbersome. A recent multicenter prospective randomized controlled trial confirmed earlier work and showed that regular prophylactic factor infusion as opposed to episodic factor replacement at the time of hemorrhesis was associated with significantly less joint damage from bleeding episodes with attendant reduction in lifetime disability. Unfortunately, the cost of treatment was enormous, and specifics such as the optimum time to initiate prophylaxis were not addressed. At present, these issues remain
unanswered and act as “a barrier to widespread acceptance of prophylaxis,” even in the United States, but especially in the developing world.

With adequate factor replacement, all types of surgery, including joint replacements, can be done. Surgical intervention should be provided for patients with hemophilia, but only at specialized centers with blood bank and coagulation laboratory support and with the participation of a hematologist who specializes in clotting disorders. A surgeon who feels comfortable operating on patients with clotting disorders also is essential. Constant-infusion techniques for administering factor concentrate during and after surgery have made adequate factor levels easier to maintain and have decreased overall perioperative use of factor concentrates. Many types of commercial factor VIII concentrate are available, most of which are manufactured with recombinant technology.

**Factor VIII Replacement**

All plasma-derived factor concentrates are virally inactivated by various methods, including exposure to solvent detergent, heat, and pasteurization. Recombinant factor VIII concentrates, manufactured by inserting the human factor VIII gene into a mammalian cell line, are widely available and are used almost exclusively, especially in developed countries. Because human plasma is not used in their production, transfusion-transmitted diseases, such as hepatitis and HIV, are no longer a risk. Recombinant concentrates at doses similar to those of plasma-derived concentrates have been efficacious in the treatment of hemorrhage. Half-life and recovery times for infused factor VIII are similar to those for plasma-derived concentrates. Current prices range from $0.35 to $0.90 per unit for factor VIII plasma-derived concentrates and from $1.00 to $1.20 per unit for recombinant factor VIII. In most hemophilia centers in the United States, recombinant factor concentrates are the only concentrates used, although high-purity, plasma-derived concentrates are still available. Because these concentrates have made early and intensive home therapy possible, overall costs of health care have greatly declined for patients treated with these materials.

Arginine vasopressin (desmopressin), a vasopressin analogue, can be used in the treatment of mild hemophilia A to increase the endogenous factor VIII level. Desmopressin increases the baseline factor VIII level about threefold, so a baseline level of at least 10% is required for efficacy. Because this is not a blood product, it poses no danger of transmitting blood-borne viruses. Although cryoprecipitate contains factor VIII, its use has been discouraged because it is not virally inactivated. It is less safe than concentrates.

**Factor IX Replacement**

Factor IX is not found in cryoprecipitate or factor VIII concentrate; these two materials are totally ineffective for the treatment of hemophilia B. Fresh-frozen plasma does contain factor IX and has been used in the past. Most fresh-frozen plasma products are not virally inactivated, however, and are less safe than factor IX concentrates.

The principles of treatment are similar to those for factor VIII replacement. Because the half-life of factor IX is longer, however, it can be given less frequently. Demand therapy is still commonly used; as for factor VIII deficiency, however, prophylaxis is beginning to be used in pediatric patients. Several plasma-derived factor IX concentrates are available, all virally inactivated. In the past, all such concentrates also contained factors II, VII, and X (prothrombin complex concentrates). Currently, only pure factor IX concentrates are used to treat factor IX deficiency. As with factor VIII concentrates, a recombinant factor IX concentrate is available and is widely used. Recovery is less than that of its plasma-derived counterpart, however, and higher doses (approximately 1.5 times calculated levels) must be infused to reach appropriate levels.

**Complications of Factor Replacement Therapy**

**Inhibitor Antibodies**

Inhibitor antibodies may develop after exposure to factor concentrate. They occur most often in patients with severe hemophilia after 9 to 30 exposures of replacement therapy, usually before the age of 5 years. A familial predisposition to the development of this complication may be noted. Because bleeding cannot be reliably controlled in patients with inhibitor antibodies, elective surgery in these patients should be done only after careful deliberation.

Inhibitor antibodies in factor VIII–deficient hemophilic patients are immunoglobulin (Ig)G antibodies (usually IgG4) and may have an unpredictable natural history. Low titer and clinically weak antibodies sometimes are easily neutralized by factor VIII and do not undergo anamnestic increases in titer after multiple factor VIII challenges. Such antibodies may rarely become high in titer. In other patients, antibody titers increase after each exposure to factor VIII. Still other patients seem to lose antibody spontaneously despite multiple subsequent factor VIII challenges. The type of antibody response to factor VIII infusion and the patient’s clinical response dictate therapy.

Therapy for patients with inhibitor antibodies has been reviewed more recently. Induction of immune tolerance through frequent administration of factor VIII successfully eliminates inhibitors in 80% of patients. In patients in whom immune tolerance therapy is unsuccessful, several approaches are available for management of acute bleeding episodes, including administration of activated prothrombin complex concentrate or, more recently, recombinant activated factor VIIa (rVIIa, NovoSeven; Novo Nordisk, Bagsvaerd, Denmark). rVIIa is thought to function directly at the site of injury, causing activation of factor IX and the extrinsic clotting system locally. Porcine factor VIII, which has limited cross-reactivity with the human antibody, was used previously, but has been removed from the market because of contamination with porcine parvovirus. A recombinant form of this protein is being investigated.

Use of immunosuppressives or glucocorticoids has been abandoned in most centers owing to lack of efficacy in this condition and serious side effects. Regimens of regular factor VIII infusions for induction of tolerance have been successful in eliminating the antibody. It has been suggested by some groups that an immune tolerance regimen should be started as early as possible after an inhibitor develops.
Rituximab may be useful for suppressing inhibitor titers in refractory patients.26

Inhibitor antibodies against factor IX are exceedingly rare. No efficacious therapy has been generally accepted. Treatment usually includes large and frequent doses of factor IX concentrate. Induction of immune tolerance with elimination of the antibody also has been done, but with less success than with antibodies to factor VIII. Use of large doses of purified factor IX concentrate in some patients with inhibitor antibodies to factor IX has resulted in anaphylactic reactions and nephrotic syndrome secondary to immune complex formation and deposition in the kidney.57,58

**Human Immunodeficiency Virus**

HIV was introduced into the U.S. blood supply in the 1970s. By the late 1970s, factor concentrate was widely contaminated. By 1982, approximately 50% of patients with hemophilia were infected with HIV.59 Currently, approximately 10% to 20% of American hemophilic patients are infected with HIV. As with other infected individuals, CD4+ lymphocyte counts and HIV titers are used to guide treatment regimens. Since 1985, in the manufacture of plasma-derived concentrates, a triple barrier to viral contamination of plasma-derived concentrates has been applied: (1) self-exclusion for donors, (2) donor screening with serologic tests for HIV, and (3) viral inactivation during concentrate production. Recombinant concentrates also are now widely available. Acquisition of HIV through factor concentrate in patients with hemophilia has been virtually nonexistent since 1985.

**Viral Hepatitis**

A second infectious side effect of cryoprecipitate or factor concentrate is hepatitis, which may result from parenterally transmitted hepatitis A, B, C, or G virus; cytomegalovirus; or another as yet unidentified pathogens. In most series, most patients with hemophilia treated before the 1980s have plasma levels of hepatitis B virus surface antibody, and a few (2% to 5%) carry hepatitis C virus surface antigen. Approximately 80% of hemophilic patients transfused before 1990 have antibody to hepatitis C virus,60 which, in contrast to hepatitis B virus antibody, is a marker for ongoing infection. Virucidal concentrate treatment methods have reduced, but not eliminated, parenteral transmission of hepatitis B and C viruses. Transmission of hepatitis A and G viruses has been reported with the use of plasma-derived concentrates. Vaccination against hepatitis B and hepatitis A is now recommended for infants born in the United States, and vaccination against hepatitis A is recommended for infants with hemophilia. Transmission of hepatitis has decreased dramatically because almost all pediatric patients are treated with recombinant products.

**Therapy for Musculoskeletal Complications of Hemophilia**

**Acute Hemarthrosis**

The most important measure in therapy for acute hemarthrosis is prompt correction of the clotting abnormality by administration of the deficient factor. Arthrocentesis, if it is accomplished within 24 hours of the onset of symptoms (but after factor replacement), may be symptomatically beneficial in advanced acute hemarthrosis; however, for diagnostic and potentially therapeutic purposes, it should be considered mandatory at any time if suspicion of infection is high.12,14 Analgesia and brief joint immobilization for no longer than 2 days often aid in pain control. Subsequently, passive range-of-motion isometric exercise should be initiated to reduce the likelihood of joint contracture (Figure 119-4).

**Chronic Hemophilic Arthropathy**

Nonsurgical management of hemophilic arthropathy has been reviewed.61

**Conservative.** A variety of conservative measures can bring remarkable benefit in the setting of chronic hemophilic arthropathy,62-65 including the following:

- **Prophylactic factor infusions**
- **Intensive physical therapy for muscle building and increased joint stability**
- **Periods of avoidance of weight bearing to allow regression of synovitis**
- **Correction of flexion contractures with wedging casts, night splints, or the judicious use of traction**
- **Training in sports to allow future maintenance of muscle mass**

In modern treatment programs, aspiration of joints with chronic synovial effusions is rarely necessary or of lasting benefit. Failure of these conservative modalities to relieve symptoms or produce regression of synovitis should prompt consideration of other options, including local corticosteroid injections (which have been described as useful),66 use of nonsteroidal anti-inflammatory drugs (NSAIDs), synovectomy, and joint replacement in the end stage.

Despite the obvious theoretical contraindications to the use of NSAIDs in hemophilia (i.e., the antiplatelet effects), several NSAIDs may be used safely for short periods as...
adjuncts to the conservative regimen. Ibuprofen, salsalate, and magnesium salicylate have been shown in a few patients to be safe and efficacious in reducing joint pain and analgesic dependence, although long-term regression of synovitis and modification of the course of chronic hemophilic arthropathy have not been shown with any NSAID. Selective cyclooxygenase-2 inhibitors of NSAIDs do not have significant antiplatelet effects and theoretically should be safer than conventional NSAIDs in patients with hemophilia. Rofecoxib and valdecoxib have been withdrawn from the market because of a causative link to increased risk of cardiovascular events. Although others are in development, celecoxib, the only remaining cyclooxygenase-2 inhibitor on the market, has not been specifically tested in hemophilic patients and, similar to other NSAIDs, should be used with caution.

**Synovectomy.** Synovectomy in the setting of hemophilic arthritis has been shown to reduce the incidence of recurrent hemarthrosis and the severity of synovitis. This procedure can be accomplished surgically, arthroscopically, or through intra-articular injection of radioactive colloids or chemical substances. Patients should be considered for synovectomy if, despite aggressive conservative measures as outlined previously, persistent hemarthroses continue with ongoing chronic synovitis. At our center, specific indications for synovectomy include persistence of at least two hemarthroses per month in the same joint accompanied by symptoms and signs of chronic synovitis despite at least 4 months of conservative therapy, including intensive factor replacement. The major drawback to surgical synovectomy remains the observation, confirmed in most series, that joint motion is reduced postoperatively compared with preoperative baseline joint motion, despite intensive rehabilitation.

To overcome this finding and the high cost of hospitalization and factor replacement therapy attendant with surgical synovectomy, arthroscopic synovectomy has been used for chronic hemophilic arthritis in recent years. Most follow-up series report that this technique is as successful as surgical synovectomy and results in less loss of motion, particularly when continuous passive motion is used in the postoperative period. The total cost of the procedure is less than that of surgical synovectomy, as is the rehabilitation period. Postoperative bleeding after arthroscopic synovectomy has been associated with poor results.

An alternative to surgical or arthroscopic synovectomy is ablation of the synovium using radioisotopic or chemical agents, as reviewed more recently. Such a nonoperative approach has been successful in reducing bleeding episodes by 70% to 80% in patients with hemophilia and is especially useful in patients with circulating factor inhibitors, in whom surgery is relatively contraindicated. Commonly used radioisotopes in the United States have included colloidal 32P chromic phosphate, yttrium (89Y), and radioactive colloidal gold (198Au). Rhenium (186Re) and erbium (169Er) have also been used with success. Theoretical long-term carcinogenic and teratogenic effects remain major concerns associated with this technique in patients who may have long life expectancies and are still of reproductive age; these effects have limited the use of radioisotopes in the United States, but less so in Europe. Chemical synovectomies using osmic acid, rifampicin, and hyaluronic acid have been attempted in some European centers with modest success, especially in children. Short-term results of radioactive and chemical synovectomies are similar, although long-term outcomes may be superior in radioisotopic synovectomy. Radioactive and chemical synovectomies remain experimental in the United States. Both offer the advantages of being minimally invasive, requiring little factor replacement, and resulting in little morbidity, and both are much less expensive than operative procedures.

**Total Joint Replacement.** Major orthopedic procedures, including total joint replacement, have been performed safely and successfully in end-stage hemophilic arthropathy, including in patients with inhibitor antibodies. The primary indication for total joint replacement is pain in an involved joint that is refractory to all conservative measures. Careful preoperative planning is imperative, including assessment for the presence of inhibitors, planning for factor replacement, and planning for a multidisciplinary rehabilitative program. It is a matter of concern, however, that most hemophilic patients in need of total joint replacement are young and may, if they are not infected with HIV, have a long life expectancy. If the procedure is performed at a young age, this virtually ensures the need for one or more revisions during the patient’s lifetime. Long-term follow-up studies of total knee replacement in hemophilic patients have shown improved functional scores but conflicting results with regard to prosthetic survival and the incidence of postoperative infection. Although an increased incidence of prosthetic infection and loosening has been noted in hemophilic patients in one large series (in keeping with many previous smaller reports), a recent 25-year follow-up study encouragingly suggests that the outcome of total knee replacement can approach that of nonhemophilic patients. This was believed to be due in large part to the use of continuous factor infusions in the perioperative period, in addition to a coordinated multidisciplinary approach by specialists within a hemophilia center. In this setting, total knee replacement has been suggested to be the treatment of choice for advanced hemophilic arthropathy. Despite this, however, concerns with regard to perioperative complications (paradoxically including deep venous thrombosis) infection, and late loosening in this population of patients remain. A comprehensive review details the many orthopedic procedures that are now available for alleviating the pain and deformity resulting from hemophilic arthropathy.

**CONCLUSION**

Gene therapy or repair to cure hemophilia may someday be a reality, although this approach currently is still fraught with serious safety concerns. Until that time, the best therapy for hemophilic arthropathy remains its prevention, and prevention is now achievable in many patients. With improvements in the safety and availability of factor concentrates, prophylactic infusion, although expensive, is now feasible, has been shown to reduce hemorrhosis and improve long-term joint outcomes, and at some time in the future may be widely available at an affordable cost. In the meantime, through a combination of prevention of hemorrhosis or correction of the hemostatic defect at the earliest
symptom of joint hemorrhage, education of the patient, application of comprehensive care, and emphasis on the importance of physical activity to maintain muscle mass, the incidence of new or progressive arthropathy can be significantly reduced.

References

KEY POINTS

The hemoglobinopathies differ in their pattern and severity of musculoskeletal manifestations.

Sickle cell disease is characterized by painful vaso-occlusive crises, which stem in part from bone and muscle ischemia.

Patients with sickle cell disease (rarely sickle trait) suffer frequent bone infarctions in the long bone shafts, especially in the humeral and femoral heads.

Sickle cell synovial infarctions may mimic acute septic or gouty arthritis.

Salmonella osteomyelitis is a rare complication of sickle cell disease; septic arthritis is usually due to Staphylococcus (a significant minority of patients have gram-negative infections).

The thalassemias rarely cause arthritis; however, the iron chelator deferiprone has been associated with severe joint pain.

CLINICAL BIOLOGY OF THE HEMOGLOBINOPATHIES

The hemoglobinopathies comprise a group of genetic disorders characterized by the presence of variant hemoglobins in circulating erythrocytes. Clinical manifestations differ dramatically based on the physicochemical behavior of the abnormal hemoglobin, and how the variant protein chains interact with other hemoglobin chains, the erythrocyte membrane, ion transporters, and oxygen. Some variant hemoglobins exhibit abnormal oxygen-dissociation curves or erythrocyte fragility resulting in dyspnea, fatigue, or hemolysis. Relevant to rheumatologic manifestations, the red cells of patients with sickle cell variants are structurally rigid, and as they traverse the microvasculature are unable to deform in small vessels; they occlude the vascular lumina and abnormally adhere to endothelial cells. As a result, patients with homozygous hemoglobin S disorder (sickle cell disease) exhibit a broad range of clinical syndromes characterized by vaso-occlusive crises involving painful infarction in tissues with low oxygen tension and lower pH. Ischemic phenomena, sometimes with lasting sequelae, seem to be particularly common in kidney, eye, bone, synovium, muscle, and skin. Other striking organ involvement may include stroke, mitral valve damage, splenic infarction with hypofunction, and acute pulmonary syndromes. Under certain conditions, these complications may mimic systemic inflammatory, autoimmune, or primary thrombotic disorders.

As a result of chronic hemolysis, some of the hemoglobinopathies such as the thalassemias (imbalanced synthesis of normal globin chains) result in hyperplastic marrow with skeletal abnormalities. Patients may require frequent transfusions, which can lead to syndromes associated with iron overload, or painful joint complications from the required use of iron chelators.

CLINICAL FEATURES OF MUSCULOSKELETAL SYNDROMES LINKED TO HEMOGLOBINOPATHIES

The recurrent painful crisis is the hallmark of sickle disease, with marked pain in the muscles and joints, as well as in the abdomen and chest. These episodes involve muscles and joints and are sometimes difficult to localize; frequently, the pattern of pain is repeated in recurrent crises. Pain that is clearly localized to an isolated area, such as a single shoulder, right upper quadrant, buttock, or groin, warrants aggressive and rapid evaluation. The pain may be severe enough to require narcotics, but extreme diligence must be paid to the development of individualized pain control care plans to limit the expectation and delivery of escalating narcotics in all patients. Painful crises may begin as early as 6 months of age as the concentration of fetal hemoglobin F wanes, permitting sickling to occur more readily.

Avascular necrosis (AVN; osteonecrosis) and ischemic bone syndromes are common in both adults and children with sickle cell anemia (but are not usually seen in patients with heterozygous Hgb S disease—sickle trait). In very young children, repeated vaso-occlusive crises involving growing small bones of the hands and feet cause the hand-foot syndrome, characterized by diffuse painful hand/foot swelling, which may be accompanied by low-grade fever. Radiographic changes may be seen approximately 10 days after the onset of symptoms; these consist of subperiosteal new bone formation in the hands and feet. Cortical thinning, multiple irregular intramedullary deposits, and areas of spotty destruction and formation of periosteal new bone may appear later. These changes can lead to a “moth-eaten” appearance. AVN is often multifocal but characteristically...
affects the femoral heads. The pain may be of acute onset or may progress insidiously from intermittent pain to pain with use, and finally to constant discomfort or severe pain at rest. Frequently, this condition is diagnosed in young to middle-aged patients who have experienced years of vaso-occlusive crises. Presumably, it occurs as the result of multiple ischemic and occlusive episodes, resulting in bone death as well as progressively increased intraosseous pressure. The prevalence of femoral head AVN is likely greater than 40%. Humeral heads are frequently affected. AVN can mimic a true arthritis with the presence of significant synovial effusion. The pathogenesis of these noninflammatory effusions is not certain, but they may result from hydrostatic pressure increases in vessels draining into the necrotic bone.

Noninfectious arthritis is well described in the setting of acute vaso-occlusive crisis. Most patients have homozygous SS disease, but a few may have hemoglobin SC (Hgb SC) disease or hemoglobin S-beta thalassemia (β-thalassemia). Crystal-induced arthritis and infection should be excluded by synovial fluid analysis and culture as appropriate. Some of the effusions (monoarticular or oligoarticular) accompany painful crises and joint examination may reveal that adjacent bone is more tender than the joint capsule itself. Almost all synovial fluids are noninflammatory, hence the possibility that bone infarction with a “sympathetic” or transudative effusion offers a pathologic explanation for these effusions. Polyarthritis with minimal or mildly inflammatory fluid has been described in the setting of crisis. In a series of 70 patients with sickle disease (including Hgb SC disease and Hgb S-thalassemia) followed prospectively, 32 had joint manifestations (30 with Hgb SS disease), indicating that if carefully looked for, articular involvement is fairly common. Most cases of acute monoarticular and oligoarticular arthritis happened in the setting of a painful crisis, but a very significant subgroup (44%) exhibited arthritis and fever as major symptoms (infection was excluded). In this series, synovial fluid analyses were striking. Sickle cells were seen in many of the synovial fluid samples—an observation that has been noted by others. Whereas cultures were negative and no crystals were observed, many of the synovial fluid samples were inflammatory in cellular appearance with neutrophilic predominance. Nonetheless, arthritis lasted only a mean of 5 days. An ankle arthritis associated with new or worsening distal leg ulcers has also been described; both may result from occlusive small vessel disease. The synovial histopathology in sickle cell–associated arthritis is generally fairly bland—minimal inflammation with some intimal proliferation and vascular congestion with occasional thrombosis has been reported.

Joint and bone infections are well-recognized complications in patients with sickle cell disease, although they are not particularly common. Salmonella is an unusual cause of musculoskeletal infection associated with sickle disease. It is more frequent in children than in adults, and more frequently will result in osteomyelitis rather than septic arthritis. Reasons for susceptibility to this specific organism remain unclear. Patients with sickle cell anemia exhibit potentially altered gut mucosal integrity due to small vessel ischemic injury, splenic infarction with dysfunction, and decreased complement activation, all of which may render patients at increased relative risk. Moreover, damage to joints from repetitive ischemic injury provides a nidus for bacterial “seeding.” The diaphysis is most commonly involved in osteomyelitis, although infection may involve the epiphysis, and bacteria may thereby migrate to the joint space.

In a recent retrospective study of 2000 consecutive adults with sickle cell disease, 59 (3%) were found to have had septic arthritis. The hip was involved in 61% of recorded cases—a disturbing observation in that this is a joint that physicians are often reluctant to aspirate and for which it is often tempting on clinical grounds to attribute acute pain to hip AVN. Also of note, in this study from France, no cases of Salmonella were reported. Most infections were due to Staphylococcus aureus, as observed in other series of patients with septic arthritis without sickle disease. A relatively large proportion of gram-negative infections theoretically could be due to increased bacterial translocation across the bowel mucosa. The authors documented a strong association in these adults with the presence of previous childhood osteomyelitis or AVN. Overall, Salmonella infection seems to be more strongly associated with osteomyelitis in patients with sickle cell disease than in those with arthritis. It is important to note that osteomyelitis may be multifocal.

Sickle cell disease is associated with hyperuricemia, likely due to increased uric acid generation associated with hemolysis and erythroid proliferation. Gout has been described in young patients with sickle disease. Although an infrequent occurrence in young patients, it is generally unexpected and can be confused with other more common causes of acute joint pain in these patients.

Patients with thalassemias have rarely been described as having arthritis or joint pain associated with their dysregulated hemoglobin synthesis. However, chelation therapy with deferiprone, which is often needed to reverse the iron load from frequent transfusions, has been associated with multiple musculoskeletal complications. Arthralgias seem to be particularly common, perhaps in 20% of patients, although arthritis has been described. Whether some of these patients may have been experiencing a reaction to periarticular and synovial iron overload, and not directly to the drug, is difficult to ascertain given short-term follow-up after drug withdrawal. Successful response to nonsteroidal anti-inflammatory therapy has been suggested by several authors to indicate a useful palliative therapy.

**DIFFERENTIAL DIAGNOSIS AND DIAGNOSTIC TESTS**

The approach to patients with hemoglobinopathies who experience acute monoarticular or oligoarticular arthritis should be no different than that taken with other patients. However, as noted previously, concern about and therefore vigilance concerning infection, AVN, and “true” sickle cell arthritis should be greater. Thus, arthrocentesis should be performed early in the diagnostic process. Patients with sickle cell disease and ongoing hemolysis often have a leukocytosis, and older literature suggests that some patients may exhibit a blunted erythrocyte sedimentation rate (ESR) response to inflammatory stimuli; however, C-reactive protein (CRP) should be reasonably reliable. Note that not all patients with septic arthritis have markedly elevated...
acute phase reactants; hence no reliable substitute for arthrocentesis has been found with fluid culture performed to diagnose or exclude an infected joint. Indirect serologic studies, such as rheumatoid factor, anticyclic citrullinated protein antibodies, and antinuclear antibodies, generally should be avoided in the initial workup of acute monoarticular and oligoarticular arthritis; this is no different in the patient with a hemoglobinopathy. No imaging test will reliably distinguish infection from crystalline arthritis or sickle cell arthritis (synovial infarction).

In patients experiencing diffuse crisis pain, careful examination must performed to find specifically affected joints. Infection can trigger a generalized crisis. It should be remembered that joint pain not worsened by provocative physical examination may be due to referred pain, and such causes as splenic infarction (left shoulder), necrotic gallbladder or pulmonary infarction with effusion (right shoulder), and periarticular osteomyelitis or bone infarction should be considered in the differential diagnosis.

AVN is diagnosed as in other patients with suspected necrosis. Synovial fluid, if present, is bland. Radiographs are insensitive (flattening and the "crescent sign" are not early findings). Magnetic resonance imaging (MRI) is much more sensitive in demonstrating the necrotic process and marrow edema.

**TREATMENT**

Treatment is generally supportive and analgesic based, once infection and gout have been excluded. Advances have been made in limiting the number of vaso-occlusive crises, but in general, most patients continue to experience painful attacks. For most patients with transient arthritis associated with crises, treatment usually involves analgesics and hydration. Intra-articular corticosteroids are not useful, and as noted, the risk of AVN is already heightened. If infection is suspected by detection of inflammatory synovial fluid, empiric antibiotics should be instituted until culture results have been returned; it should not be assumed that the inflammatory fluid is due to sickle arthritis until cultures are found to be negative. It is far more critical to adequately cover for staph species (including methicillin-resistant Staphylococcus aureus [MRSA]) than to cover for salmonella alone, although a recent study suggests the need for broad gram-negative coverage as well. In the setting of osteomyelitis, every effort should be made to obtain culture definition of the infecting organism before a prolonged course of antibiotic therapy is initiated.

No “proven” treatment approaches for AVN are known (see Chapter 103); conservative methods, such as avoidance of weight bearing with crutches and bed rest, have been used widely, but without proven success. Core decompression is a procedure that is commonly used to treat AVN at early stages; however, its use is based mainly on orthopedic case series, and formal trial-based evidence for this approach is lacking. Moreover, use of non-weight-bearing crutches may be problematic because AVN of the humeral heads may also be present, making the use of crutches difficult. Nonetheless, given the young age of patients with AVN, delaying the initial total hip arthroplasty is believed by many to be rational; core decompression may relieve some of the pain, at least temporarily. Authors have expressed concern that the failure rate for total hip arthroplasty is especially high among patients with sickle disease—another reason why some surgeons prefer to delay this procedure in younger patients with sickle disease. Ultimately, explanation, patient education, and early involvement of the multidisciplinary team will offer additional support for this patient group, given the paucity of clinical evidence supporting therapeutic decisions.

**References**


The references for this chapter can also be found on www.expertconsult.com.
Musculoskeletal manifestations are a common presenting feature of endocrine disease; therefore, a high level of vigilance should be maintained during initial and ongoing assessment of patients with bone, muscle, and soft tissue complaints. Similarly, endocrine disease can arise in the treatment of rheumatic diseases, especially with the use of glucocorticoids. Whereas the underlying mechanisms of many such presentations remain uncertain, their association is sufficiently common as to merit detailed consideration in the history and examination of all new rheumatic disease presentations, and superficial endocrine assessment should form a part of the routine early investigation of new-onset arthralgia and myalgia patients. This chapter addresses some of the more common musculoskeletal elements of endocrine syndromes.

**HYPOTHYROIDISM**

Deficiency of thyroid hormone leads to the state of hypothyroidism. The most common cause is Hashimoto’s thyroiditis, an autoimmune process in which lymphocytic infiltration and fibrous tissue accumulation cause replacement of normal thyroid tissue and thereby gland dysfunction.1 The incidence of autoimmune hypothyroidism (Hashimoto’s thyroiditis) is increased in patients with systemic sclerosis, as well as with systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), mixed connective tissue disease, Sjögren’s syndrome, and polymyositis.2 Most patients develop antibodies against thyroid peroxidase and/or thyroglobulin.3 Hashimoto’s thyroiditis is associated with HLA-B8, HLA-DR3, HLA-Aw30, and HLA-DR5.4 Some patients with Hashimoto’s thyroiditis also exhibit antinuclear antibodies, and many have anti-DNA antibodies; despite this, overt SLE is uncommon. Other potential causes of thyroid gland dysfunction include treatment with radioactive iodine (131I) for Graves’ disease and drug-induced hypothyroidism, for example, associated with amiodarone, iodine (i.e., Wolff-Chaikoff effect), or other drugs.5 In addition, hereditary disorders of the iodothyronine synthesis pathway (thyroxine [T4] and triiodothyronine [T3]), as well as pituitary tumors and related surgical resections, are possible causes.5,10

Hypothyroidism can cause a broad range of symptoms associated with mild (e.g., fatigue, weight gain, cold intolerance, mental slowing, muscle cramping, bradycardia) to severe complications (e.g., heart enlargement, myxedema coma [rare]). Neuromuscular and musculoskeletal manifestations are observed in many patients. These manifestations can occur at any time in the disease process and include weakness, joint and muscle pain, aching, and stiffness.11 Overall, 30% to 80% of hypothyroid patients manifest neuromuscular symptoms, depending on the severity of hypothyroidism, whereas weakness is observed in almost 30% of patients.12 More rarely, hypothyroid myopathy manifests as a polymyositis-like illness with proximal muscle weakness and an increased creatinine phosphokinase level. It may also present as muscle enlargement (pseudohypertrophy); in adults, this condition is called Hoffmann’s syndrome.13 In children with hypothyroid disease (cretinism), a pattern of proximal weakness and diffuse muscle enlargement is known as Kocher-Debre-Semelaigne syndrome.14 Several case reports describe rhabdomyolysis associated with hypothyroidism; in these cases, hypothyroidism is thought to have been causative.15 Myxedema represents a phenomenon in which thickening of muscle tissue occurs after light percussion in approximately 30% of patients; however, it is not entirely specific for the latter.16 Myxedema is likely caused by delayed calcium reuptake by the sarcoplasmic reticulum, which thereby prolongs muscle contraction. This prolongation of muscle contraction is thought to be related to the development of muscle hypertrophy. A further neuromuscular manifestation concerns peripheral neuropathy. In particular, carpal tunnel syndrome is found in 15% to 30% of patients with hypothyroidism.17
HYPERTHYROIDISM

Hyperthyroidism (Graves’ disease) may affect the musculoskeletal system in several ways; the most common manifestation is osteopenia and potentially osteoporosis. The latter may occur in patients with idiopathic Graves’ disease but also in those with iatrogenic hyperthyroidism. Failure to recognize the declining need for thyroid replacement with age is a further important cause of iatrogenic hyperthyroidism in older women. For this reason, patients with hyperthyroidism who are on thyroid hormone and estrogen therapy should have thyroid-stimulating hormone levels monitored and the dose readjusted to keep the thyroid-stimulating hormone level within the normal range. Bone density has been shown to increase after correction of the hyperthyroid state.

Pretibial myxedema is a syndrome of painless nodules that occur over the pretibial areas; virtually all affected patients have concomitant Graves’ ophthalmopathy. Cutaneous lesions vary in size, ranging from nodules with a diameter of 1 cm to very large lesions covering most of the pretibial surface. The lesions are colored differently—from pink to a light purple hue—and can be misdiagnosed as erythema nodosum. In contrast to erythema nodosum, however, the lesions are painless. They are caused by the accumulation of hyaluronic acid in the skin and in some cases exhibit a shiny appearance resembling systemic sclerosis or morphea.

Hyperthyroidism may be associated with changes in the nails, including onycholysis, or elevation of the nail from the nail bed, and clubbing. Clubbing usually is part of the condition known as thyroid acropathy, a rare manifestation of hyperthyroidism associated with periostitis around the metacarpal joints and distal soft tissue swelling of the digits. This condition is not clearly related to levels of thyroid hormone, and it may be seen after the patient has reverted to the euthyroid state.

Proximal muscle weakness, a common complication of hyperthyroidism, is present in most patients. Most patients have lost weight and have other evidence of loss of muscle mass. Proximal muscle weakness reverts rapidly with correction of the hyperthyroid state, suggesting a direct metabolic link to thyroxine effector function. Perhaps related to the proximal myopathy, adhesive capsulitis of the shoulder seems to be increased in patients with hyperthyroidism. In these patients, the condition can be insidious and difficult to treat, with frozen-shoulder syndrome often the initial manifestation.

Strong relationships have been noted between Graves’ disease and Hashimoto’s thyroiditis and other rheumatic diseases. Seventy-five percent to 90% of patients with Graves’ disease have antinuclear antibodies, and a proportion also express anti-DNA antibodies, despite the fact that overt SLE is uncommon. Graves’ disease is associated with HLA-B8, HLA-A1, HLA-Cw7, and HLA-DR3, and combinations of these antigens correlate with persistent disease.

HYPOPARIPTHYROIDISM

Hyypoparathyroidism is usually secondary to surgical removal of the parathyroid glands and commonly is characterized by proximal muscle weakness related to the degree of hypocalcemia. This condition responds rapidly to treatment with vitamin D and calcium. Other common musculoskeletal manifestations of hypoparathyroidism include osteomalacia and rickets, which are discussed elsewhere. Idiopathic hypoparathyroidism is a rare disorder, seen as part of DiGeorge’s syndrome with thymic hypoplasia. Pseudohypoparathyroidism, known as Albright’s hereditary osteodystrophy, is caused by end-organ resistance to the effects of parathyroid hormone. Pseudohypoparathyroidism is due to a defect in GNAS1 (guanine nucleotide binding protein, alpha-stimulating activity polypeptide 1). Patients show persistent hypocalcemia and hyperphosphatemia, but parathyroid hormone levels are consistently elevated. Type Ia pseudohypoparathyroidism, which is autosomal dominant, is characterized by short stature, calcification of perisinal ligaments, and, usually, mental retardation. Patients may present with shortening of the fourth metacarpal and metatarsal bones, and instead of the usual knuckle appearance over the fourth metacarpal head, these individuals have a dimple. Patients affected by type Ib pseudohypoparathyroidism also show resistance to parathyroid hormone but have normal phenotype. Type Ia pseudohypoparathyroidism is almost always inherited maternally, and type Ib is inherited paternally.

HYPERPARATHYROIDISM

Primary hyperparathyroidism results from hyperfunction of the parathyroid glands themselves. The condition is characterized by hypersecretion of parathyroid hormone (PTH) caused by adenoma, hyperplasia, or, rarely, carcinoma of the parathyroid glands. Secondary hyperparathyroidism comprises the reaction of the parathyroid glands to hypocalcemia caused by pathologic conditions other than parathyroid pathology (e.g., chronic kidney disease). The most important joint manifestation of primary hyperparathyroidism is calcium pyrophosphate deposition (CPPD) disease, a disorder with varied clinical manifestations attributed to precipitation of calcium pyrophosphate dihydrate crystals in the connective tissues. Chondrocalcinosis refers to radiographic evidence of calcification in hyaline and/or fibrocartilage. Pseudogout refers to clinically evident acute synovitis that results from shedding of crystals in the joint space after rupture of a CPPD deposit; it is characterized by red, tender, and swollen joints that may resemble gouty arthritis (see Chapter 96). CPPD crystal deposition disease is a polyarticular arthritis (i.e., it leads to inflammation of several joints), although it can initially present as monoarthritis. Diffuse idiopathic skeletal hyperostosis (DISH) and pseudoankylosing spondylitis are considered additional possible forms of calcium pyrophosphate dehydrate crystal deposition disease (CPPD CDD), but these syndromes are considered in detail in Chapter 96. Some patients with long-standing hyperparathyroidism report proximal muscle weakness—a condition rapidly reversed by removal of the parathyroid adenoma. In these patients, the muscle enzymes are normal, and electromyography and muscle biopsy reveal a picture most consistent with denervation. In secondary hyperparathyroidism associated with advanced renal disease, numerous bone and articular abnormalities are described. The musculoskeletal
changes of renal osteodystrophy resulting from secondary hyperparathyroidism include erosive arthritis in the hands, resorption of the distal clavicle, and erosions in the axial skeleton. In children, widespread bone deformities of osteitis fibrosa cystica can be very disabling. Other musculoskeletal manifestations of advanced renal failure include aluminum-induced osteomalacia and β2-microglobulin amyloidosis.

ADRENAL GLAND DISORDERS

Primary Cushing’s disease is caused by a pituitary adenoma, which is a benign tumor resulting in excess adrenocorticotropic hormone (ACTH). It may also arise from adrenal adenomas. Iatrogenic Cushing’s syndrome (hypercortisolism/hyperadrenalism) secondary to exogenous high-dose long-term glucocorticoid therapy is the most common condition involving adrenal hormones that mediates adverse effects on the musculoskeletal system. Osteonecrosis, a common late complication of high-dose glucocorticoid therapy, may first become evident months or years after glucocorticoid therapy has been discontinued. It is observed less commonly after short courses of therapy, however, or after intermittent high-dose intravenous therapy, although clinical vigilance is required. Because rheumatic patients receiving high-dosage glucocorticoid therapy have associated diseases characterized by joint pain, the risk of osteonecrosis should be considered, particularly in the differential diagnosis. Use of the lowest acceptable dose of glucocorticoids has reduced the risk of osteonecrosis.

Steroid-induced myopathy may be difficult to recognize in patients being treated for primary or secondary inflammatory rheumatic conditions, particularly myopathies. However, steroid myopathy is characteristically more severe in the pelvic girdle. It may come on gradually or abruptly, starting with weakness and muscle aching, and can be sufficiently severe to render patients bedbound. Biopsy specimens or T2 relaxation times are compatible with muscle fiber atrophy, whereas muscle enzymes are normal. Long-acting and fluorinated glucocorticoids are more likely to induce myopathy, and treatment usually requires ultimate discontinuation of glucocorticoids before any improvement is seen; weeks or months may pass before muscle strength begins to return.

Osteopenia as a secondary effect of hypercortisolism (primary or secondary) is independent of degree or duration of hypercortisolism (adenoma) but may be related to the total dose and duration of glucocorticoid therapy and is more frequent with dosages greater than the equivalent of 7.5 mg/day of prednisone. Prophylaxis is recommended, and regimens shown to be effective in preventing or treating glucocorticoid-induced osteoporosis must include calcium, vitamin D₃, and bisphosphonates.

Cushing’s disease may be confused with primary musculoskeletal disease, including polymyalgia rheumatica, polymyositis, or statin myopathy, or it may be mistaken for back pain that may arise from osteoporotic fractures or other pathologies. In Cushing’s syndrome secondary to ectopic adrenocorticotropic hormone production, glucocorticoid serum levels may be extremely high, inducing severe myopathy and additional complications such as steroid psychosis. Several other side effects of glucocorticoids on the musculoskeletal system are less well understood. Some patients describe intense joint pain, frequently most severe in the knees, when high doses of glucocorticoids are started. This pain typically resolves even if the dose is left unchanged.

Adrenal insufficiency is classified into three subtypes based on where the abnormality is based in the hypothalamic-pituitary-adrenal (HPA) axis. Primary insufficiency is caused by adrenal gland damage (Addison’s disease). The secondary form is related to insufficient corticotropin (ACTH) from the pituitary gland. The tertiary form is related to insufficient corticotropin-releasing hormone (CRH) generated from the hypothalamus. Acute adrenal insufficiency, or adrenal crisis, is severe and is characterized by shock. Primary adrenal insufficiency (Addison’s disease) is almost exclusively autoimmune (now only rarely related to tuberculosis) and is characterized by weakness, weight loss, abdominal pain, hyperpigmentation, nausea, and hypotension. Secondary adrenal insufficiency can be related to destruction of the pituitary gland or to deficiency of ACTH. Classically, it is associated with hemorrhage of the pituitary gland, or thrombosis, or it may be noted during infiltrative processes such as those seen when sarcoidosis affects the pituitary gland. Glucocorticoid use and subsequent withdrawal can cause secondary or tertiary adrenal insufficiency. Iatrogenic Addison’s disease can be subtle because mineralocorticoids are still being produced; salt wasting, hyperkalemia, and postural hypotension are usually less impressive; and hyperpigmentation is not seen because the pituitary is suppressed. Adrenal insufficiency, particularly in such circumstances, can mimic fibromyalgia syndrome. Tertiary adrenal insufficiency is most commonly related to withdrawal of glucocorticoids. Glucocorticoid-induced adrenal insufficiency can be caused by several mechanisms, including decreased hypothalamic synthesis of CRH, blockade of the actions of CRH on the anterior pituitary, and, after prolonged or profound deficiency of ACTH, adrenal atrophy. As in idiopathic Addison’s disease, features may not be evident unless the individual undergoes a new exogenous stressful condition, such as surgery or infection. Adrenal insufficiency (adrenal crisis) is a rare disorder that usually manifests with gradually evolving clinical symptoms and signs. Occasionally, acute adrenal insufficiency (crisis) can become a life-threatening condition as the result of acute interruption of a normal or hyperfunctioning adrenal or pituitary gland, or sudden interruption of adrenal replacement therapy. Addisonian crisis has been seen even in individuals still receiving physiologic or “replacement” doses of glucocorticoids. It should be assumed that individuals taking glucocorticoids at greater than the equivalent of 5 mg/day of prednisone have a pituitary-adrenal axis unable to respond to severe stress (medical and surgical stress, concomitant use of certain medications); consequently, increases in the glucocorticoid dose should be considered in the acute setting.

Musculoskeletal Manifestations and Steroid Deficiency

The so-called steroid withdrawal syndrome consists of widespread arthralgias, myalgias, malaise, and sometimes...
low-grade fever.\textsuperscript{68} It may be seen when high-dose glucocorticoids have been used for nonrheumatic conditions, such as asthma or inflammatory bowel disease, and it arises from suppression of the pituitary-adrenal axis. Moreover, abrupt reduction of the dose of glucocorticoid can cause a severe rebound flare of the underlying disease, at least in rheumatic diseases.\textsuperscript{80} This condition may arise even though the dose of glucocorticoid remains in the pharmacologic range. It is important to recall that treatment with glucocorticoids should not be stopped until endogenous synthesis of glucocorticoids is fully efficient. Administration of low-dose "modified-release" glucocorticoids, which addresses appropriate timing of administration (at night, reflecting HPA circadian rhythms), has been shown to reduce the severity of this syndrome.\textsuperscript{70-72}

Subclinical hypocortisolism associated with insufficient production of cortisol may arise in conditions of chronic stress (e.g., interpersonal conflict, chronic inflammatory disease state), especially in the elderly.\textsuperscript{73,74} In this circumstance, a new stressor may induce the development of polymyalgia rheumatica\textsuperscript{75,76} (see Chapter 88).

**DIABETES MELLITUS**

Diabetes mellitus (DM) may affect the musculoskeletal system in myriad ways. Many rheumatologic disorders have been observed more frequently among individuals with DM than in the general population.\textsuperscript{77} The metabolic perturbation characterized by diabetes (including glycosylation of proteins; microvascular abnormalities with damage to blood vessels and nerves; and accumulation of extracellular matrix proteins in skin and periarticular structures) results in overall changes in the connective tissue.\textsuperscript{78} Musculoskeletal complications are most commonly seen in patients with a long-standing history of type 1 diabetes, but they also may be observed in patients with type 2 diabetes.\textsuperscript{79} Some of these complications have a direct association with diabetes, whereas others have a suggested but unproven association (Table 121-1).

**Table 121-1  Rheumatic Complications of Diabetes Mellitus (DM)**

<table>
<thead>
<tr>
<th>Conditions Limited to DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetic muscle infarction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditions Occurring More Frequently in DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetic cheiroarthropathy, or limited joint mobility, or stiff hand syndrome</td>
</tr>
<tr>
<td>Flexor tenosynovitis, or trigger finger</td>
</tr>
<tr>
<td>Dupuytren's contracture</td>
</tr>
<tr>
<td>Carpal tunnel syndrome</td>
</tr>
<tr>
<td>Adhesive shoulder capsulitis, or frozen shoulder</td>
</tr>
<tr>
<td>Calcific shoulder periartthritis (tendinitis)</td>
</tr>
<tr>
<td>Reflex sympathetic dystrophy, or shoulder-hand syndrome, or complex regional pain syndrome</td>
</tr>
<tr>
<td>Diabetic osteoarthritis, or Charcot's arthropathy, or neuropathic arthropathy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditions Sharing Risk Factors of DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffuse idiopathic skeletal hyperostosis</td>
</tr>
<tr>
<td>Gout/Pseudogout</td>
</tr>
<tr>
<td>Osteoarthritis</td>
</tr>
</tbody>
</table>

**Hands**

Diabetic cheiroarthropathy, also known as diabetic stiff hand syndrome or limited joint mobility syndrome, has been reported in 8% to 50% of all patients with type 1 diabetes and is also seen in type 2 diabetic patients.\textsuperscript{80} Prevalence increases with the duration of diabetes, and the disorder is associated with and is predictive of other diabetic complications. Diabetic cheiroarthropathy is characterized by thick, tight, waxy skin reminiscent of systemic sclerosis with limited joint range of motion (inability to fully flex or extend the fingers) and possible sclerosis of tendon sheaths. Once again, the underlying causes are thought to be multifactorial and include increased glycosylation of collagen in the skin and periarticular tissue, decreased collagen degradation, and diabetic microangiopathy.\textsuperscript{81} As a consequence, flexion contractures of the fingers may develop at advanced stages, and the classic indication of the presence of this condition is known as the “prayer sign,”\textsuperscript{82} which is seen as a patient's inability to press the palms together completely without a gap remaining between opposed palms and fingers.

Another frequent complication affecting the hands is flexor tenosynovitis (or trigger finger); patients describe a catching sensation or a locking phenomenon that may be associated with pain in the affected fingers.\textsuperscript{83} Physical examination reveals a palpable nodule, usually in the area overlying the metacarpophalangeal joint, and thickening along the affected flexor tendon sheath on the palmar aspects of the finger and hand. Flexor tenosynovitis is thought to have the same pathogenesis as diabetic cheiroarthropathy; its prevalence is similarly related to the duration of diabetes.\textsuperscript{84} Abnormalities of matrix metalloproteinases and their tissue inhibitors have been demonstrated in diabetic fibroblast culture systems ex vivo (Brown I and McInnes IB, personal communication). Dupuytren's contracture results from thickening, shortening, and fibrosis of the palmar fascia; nodule formation along the fascia is often detected.\textsuperscript{85} Flexion contractures of the fingers may result, usually at the fourth finger but sometimes involving any of the second through fifth digits. Dupuytren's contracture has been reported in 16% to 42% of diabetic patients, with pathogenesis that is thought to be the same as that for cheiroarthropathy. Once again, the prevalence of this condition increases with disease duration, but Dupuytren’s contracture also may be seen early in the course of the disease.

Another frequent complication of DM is carpal tunnel syndrome (CTS), which is observed in up to 20% of diabetic patients; the specific relationship to diabetes is thought to be median nerve entrapment caused by diabetes-induced connective tissue alterations, as previously mentioned.\textsuperscript{86} CTS usually is diagnosed on the basis of patient history and classic clinical findings, such as Tinel's sign (tapping over the median nerve on the volar aspect of the wrist) and results of Phalen's test (the wrist flexion test).\textsuperscript{87} It is important to examine patients for possible motor weakness caused by median nerve compression; electromyography/nerve conduction velocity (EMG/NCV) testing can confirm the diagnosis of CTS in uncertain cases, helping to localize the site of nerve entrapment. In any case, management of CTS is the same for diabetic patients as for nondiabetic patients.
Shoulder

The most common shoulder involvement in diabetes occurs as adhesive capsulitis, or frozen shoulder, which has been reported in 19% of diabetic patients. Adhesive capsulitis refers to a stiffened glenohumeral joint, usually caused by a reversible contraction of the joint capsule; patients report shoulder stiffness, along with decreased range of motion. Calcific shoulder periartitis (tendinitis) occurs at least three times more frequently than in people without DM and most commonly affects the shoulder in which calcium hydroxyapatite crystals deposit, predominantly in periarticular areas. Shoulder radiographs show calcium deposits outside of the joint, often in the area of the rotator cuff tendons; however, in up to 60% of cases, this condition is asymptomatic in DM.

Reflex sympathetic dystrophy, also known as shoulder-hand syndrome or complex regional pain syndrome (CRPS), has been reported in diabetic patients, although whether it occurs with increased frequency is controversial, and it may be associated with adhesive capsulitis (with or without calcific periartitis). Patients describe pain from shoulder to hand in the affected limb, and classic symptoms include swelling of the affected limb/area, skin changes (changes in hair growth, shiny skin, color and temperature changes), increased sensitivity to temperature and touch (hyperesthesia), and vasomotor instability. Transient, patchy osteoporosis may be seen.

Feet

Diabetic osteoarthropathy (also known as Charcot’s or neuropathic arthropathy) is a condition involving destructive, lytic joint changes that most commonly affect the pedal bones. It presents as a severe, destructive form of degenerative arthritis caused by loss of sensation (brought on by underlying diabetic neuropathy) in involved joints, leading to inadvertent (often unappreciated) repeated microtrauma to the joints with consecutive degenerative changes. Diabetic osteoarthropathy is rare, affecting only 0.1% to 0.4% of diabetic patients, and is seen in both type 1 and type 2 diabetes. The average duration of Charcot’s arthropathy in affected patients is 15 years, and physical examination reveals peripheral neuropathy. Additional symptoms include skin changes such as erythema, swelling, hyperpigmentation, or purpura and soft tissue ulcers over the affected area, as well as joint loosening or instability and joint deformities. The diagnosis is made on the basis of radiographic findings, with symptoms often milder than would be expected from a view of the radiographs. Usually, no history of overt trauma is reported. Depending on the stage and severity of Charcot’s arthropathy, radiographs can show degenerative changes with subluxation, bone fragments, osteolysis, periosteal reaction, deformity, and/or ankylosis.

Computed tomography (CT) scans are insensitive in evaluating disease activity, whereas magnetic resonance imaging (MRI) and bone scintigraphy studies are valuable adjuncts to plain films in this regard. Diabetic peripheral neuropathy is thought to play the greatest pathogenic role in diabetic osteoarthropathy. Differential diagnosis includes inflammatory and degenerative processes, infections, tumors, deep venous thrombosis or thrombophlebitis, and neuropathic arthropathies secondary to other conditions.

Muscles

Diabetic muscle infarction is a rare condition characterized by spontaneous infarction with no history of trauma that tends to affect patients with a long history of poorly controlled DM. It is seen more frequently in patients with insulin-requiring diabetes, and most affected patients show multiple microvascular complications (e.g., neuropathy, nephropathy, retinopathy). The condition is characterized by an acute onset of pain and swelling over days to weeks in affected muscle groups (usually the thigh or calf), along with varying degrees of tenderness; creatinine phosphokinase levels may be normal or elevated. However, laboratory investigations need to exclude other conditions such as tumor, muscle infection/abscess, thrombophlebitis/thrombosis, localized myositis, and osteomyelitis. CT scans are nonspecific, and MRI may show high signals of the involved muscle on T2-weighted images.

Incisional muscle biopsy may be needed to confirm the diagnosis; primary findings on biopsy include muscle edema and necrosis. Because excisional muscle biopsy may worsen the muscle lesion, this procedure should be done only to rule out infection or malignancy. If such a biopsy is performed, it is important that tissue is sent for culture, and that consideration is given to the presence of atypical organisms and of mycobacteria.

Diffuse Skeletal Disease

Diffuse idiopathic skeletal hyperostosis (DISH) is characterized by metaplastic calcification of spinal ligaments (diagnosed on lateral spine radiographs), along with osteophyte formation. However, disk spaces, apophyseal joints, and sacroiliac joints are spared, and the thoracic spine is most commonly affected. DISH may be accompanied by more generalized calcification of other extra-axial ligaments and tendons (Figure 121-1). The underlying pathophysiology...
is unclear; however, DISH clearly has greater prevalence among diabetic patients than among people without diabetes, particularly in association with type 2 diabetes, and in obese patients. Classically, patients complain of stiffness in the neck and back with decreased range of motion, and pain generally is not a prominent symptom.

References


The references for this chapter can also be found on www.expertconsult.com.
Musculoskeletal syndromes may be associated with malignancy in a variety of ways. Cause and effect are difficult to define clearly in many situations, however. Certain chronic rheumatic diseases have been associated with increased risk of the subsequent development of malignancy; one example is lymphoma in an individual with primary Sjögren’s syndrome. The converse situation also exists, in that certain rheumatic diseases, such as dermatomyositis, are seen more frequently in association with neoplastic disease. Other factors can contribute to the association of musculoskeletal syndromes and malignancy. Many of the medications used to treat rheumatic diseases modulate the immune system and may be associated directly or indirectly with increased risk for the subsequent development of malignancy. In unusual circumstances, musculoskeletal involvement occurs as a paraneoplastic process, defined as a hormonal, neurologic, hematologic, or biochemical disturbance associated with malignancy, but not directly related to invasion by the neoplasm or its metastases.

PARANEOPlastic SYNDROMES

Musculoskeletal syndromes can develop as a manifestation of a paraneoplastic process and occasionally can be the first presentation of an underlying malignancy. Hematologic malignancies, lymphoproliferative disorders, and solid tumors are associated with a wide variety of paraneoplastic syndromes. Older age of onset, atypical features of rheumatic disease, and absence of response to glucocorticoids or other conventional therapy may suggest a paraneoplastic process. Knowledge of the associations with rheumatic syndromes and underlying malignancy is crucial when caring for these patients. Hypertrophic osteoarthropathy, amyloidosis, and secondary gout are reviewed in Chapters 102, 116, 94, and 95. Table 122-1 lists common paraneoplastic associations.

Carcinomatous Polyarthritis

The term carcinomatous polyarthritis is used to describe the development of arthritis in association with malignancy, but it is distinct from arthritis associated with metastasis or direct tumor invasion. Table 122-2 lists common features of carcinomatous polyarthritis. It generally occurs in patients who are older; it has an explosive onset and often develops in close temporal correlation with discovery of the malignancy. Although it can have various presentations and may mimic the appearance of rheumatoid arthritis (RA) or asymptomatic migratory polyarthritis, carcinomatous polyarthritis is more often a seronegative asymmetric disease with predominant involvement of the lower extremities and some sparing of the small joints of the hands. No evidence of direct tumor extension or metastasis has been found, and no specific histologic or radiographic appearance has been identified. Carcinomatous polyarthritis can occur in association with many types of malignancy, but has been reported in greatest frequency in association with breast, colon, lung, and ovarian cancers and with lymphoproliferative disorders. The underlying pathogenesis of this process has not been elucidated; however, the arthritic symptoms may be improved with successful treatment of the malignancy.

Vasculitis

Vasculitis in association with malignancy is uncommon and has a reported prevalence of only 8% among patients with malignancy. The association seems to be significantly
### Table 122-1 Paraneoplastic Syndromes

<table>
<thead>
<tr>
<th>Connective Tissue Disease</th>
<th>Malignancy</th>
<th>Clinical Setting</th>
<th>Clinical Alert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carcinomatous polyarthritis</td>
<td>Multiple types of solid tumors, including breast; lymphoproliferative disorders</td>
<td>See Table 122-2</td>
<td>See Table 122-2</td>
</tr>
<tr>
<td>Vasculitis</td>
<td>Lymphoprolietic and hematopoietic malignancies</td>
<td>Cutaneous vasculitis most common; systemic vasculitis rare</td>
<td>Vasculitis not related to infections, medications, or autoimmune disease</td>
</tr>
<tr>
<td>Mixed cryoglobulinemia</td>
<td>Non-Hodgkin’s lymphoma</td>
<td>Immune complex–mediated disease with cutaneous vasculitis, neuropathy, fatigue, and visceral organ involvement</td>
<td>Usually appears 5-10 yr after diagnosis of cryoglobulinemia</td>
</tr>
<tr>
<td>Panniculitis</td>
<td>Hematologic malignancies; pancreatic, breast, and prostate cancers</td>
<td>Induration of skin and deeper tissues; eosinophilia often present</td>
<td>Usually refractory to prednisone</td>
</tr>
<tr>
<td>Fasciitis</td>
<td>Ovarian, breast, gastric, and pancreatic cancers</td>
<td>Palmar fasciitis with inflammatory polyarthritis; similar in presentation to reflex sympathetic dystrophy</td>
<td>Bilateral presentation; severe fibrosis and contractures; aggressive course</td>
</tr>
<tr>
<td>Reflex sympathetic dystrophy syndrome</td>
<td>Multiple cancer types; Pancoast tumors</td>
<td>Tumors may invade stellate ganglion or brachial plexus on affected side</td>
<td>Absence of typical antecedent factors; failure to respond to conventional therapies</td>
</tr>
<tr>
<td>Erythromelalgia</td>
<td>Myeloproliferative disorders</td>
<td>Often seen in setting of thrombocytosis</td>
<td>—</td>
</tr>
<tr>
<td>Atypical polymyalgia</td>
<td>Renal, lung, and colon cancer; multiple myeloma</td>
<td>—</td>
<td>Age &lt;50 yr; asymmetric involvement; poor response to prednisone</td>
</tr>
<tr>
<td>Digital necrosis</td>
<td>Gastrointestinal and pulmonary tumors</td>
<td>Severe Raynaud’s phenomenon with onset &gt;50 yr old</td>
<td>Asymmetric features; digital necrosis</td>
</tr>
<tr>
<td>Remitting seronegative symmetric synovitis with pitting edema</td>
<td>Several tumor types</td>
<td>Abrupt onset of arthritis and edema surrounding wrists and small joints of hands</td>
<td>Presence of fever, weight loss; poor response to prednisone</td>
</tr>
<tr>
<td>Multicentric reticulohistiocytosis</td>
<td>Lung, stomach, breast, cervix, colon, and ovarian carcinomas</td>
<td>—</td>
<td>Rare associations with malignancy limited to case reports</td>
</tr>
<tr>
<td>Lupus-like syndromes</td>
<td>Variety of solid tumors and lymphoproliferative disorders</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Antiphospholipid antibodies</td>
<td>Multiple cancer types</td>
<td>Association between antibodies, cancer, and risk of thrombosis unclear</td>
<td>Higher presence of antibodies found in patients with malignancy</td>
</tr>
<tr>
<td>Osteogenic osteomalacia</td>
<td>Solid tumors and tumors of mesenchymal origin</td>
<td>Bone pain and muscle weakness</td>
<td>Diligent search is indicated in all patients with late-onset apparent idiopathic osteomalacia</td>
</tr>
<tr>
<td>Sarcomiosis</td>
<td>Cervical, bladder, gastric, lung, breast, and renal cancers and cutaneous and pulmonary squamous cell carcinomas</td>
<td>Highest incidence of “malignancy” during first 4 yr after detection of granuloma</td>
<td>Malignant tumors can cause sarcomoid-like tissue reactions leading to mistaken diagnosis of sarcoidosis before recognition of malignancy</td>
</tr>
<tr>
<td>Lymphomatoid granulomatosis</td>
<td>Lymphoma</td>
<td>Unusual granulomatous form of vasculitis with angiodestructive infiltration of various tissues</td>
<td>—</td>
</tr>
</tbody>
</table>

### Table 122-2 Features of Carcinomatous Polyarthritis

- Close temporal relationship between onset of arthritis and discovery of malignancy
- Late age of onset of arthritis
- Asymmetric joint involvement
- Explosive onset
- Predominant lower extremity involvement with sparing of wrists and small joints of hands
- Absence of rheumatoid nodules
- Absence of rheumatoid factor
- No family history of rheumatoid disease
- Nonspecific histopathologic appearance of synovial lining
- No periosteal reaction

Higher with lymphoproliferative and myeloproliferative disorders than with solid tumors, and vasculitis commonly predates the identification of malignancy. The vasculitic process is most often small vessel and cutaneous and only rarely involves significant organs. Up to 5% of patients with cutaneous vasculitis have an underlying neoplasm. Treatment often requires the use of glucocorticoids and therapy directed against the underlying malignancy, although it seems that this is often ineffective. Table 122-1 shows malignancies associated with vasculitis. In the setting of malignancy, it is believed that persistent antigen stimulation from the tumor results in T cell activation or immune complex formation and deposition.
The development of small vessel vasculitis has been reported to antedate and postdate the development of lymphoproliferative and myeloproliferative diseases. One group looked retrospectively at 222 patients with vasculitis and identified 11 who had developed an associated malignancy. Of these 11 patients, 7 had hematologic neoplasia, and 4 had malignant solid tumors. Nine of the patients manifested cutaneous vasculitis, and the remaining 2 had vasculitic involvement in the bowel. In 4 patients, the development of vasculitis antedated the diagnosis of malignancy. Similar findings were reported by investigators, who found an underlying malignancy in 8 of 192 patients with cutaneous vasculitis. Most malignancies were hematologic (6 of 8) and predated (5 of 8) the diagnosis of cancer. In a retrospective analysis of 23 patients with cutaneous vasculitis and hematologic malignancies, the authors were able to attribute the presence of vasculitis to the malignancy itself in 61% of cases.

Systemic vasculitis is much less commonly associated with underlying malignancy. Case reports and small series have found antineutrophil cytoplasmic antibody (ANCA)-negative and ANCA-positive vasculitis associated with hematologic malignancies. Granulomatosis with polyangiitis (GPA) (formerly Wegener's granulomatosis) has likewise been associated with the development of several types of malignancies, including lymphoproliferative disorders, bladder cancer, and renal cell carcinoma. In some cases, the malignancy was diagnosed within months of the diagnosis of GPA, and in other reports, cancer developed many years after diagnosis and treatment of vasculitis, making it unclear whether the malignancies were a result of the vasculitis or possibly the treatment. A group from the Cleveland Clinic did a retrospective study to assess directly the temporal relationship between vasculitis and cancer. During an 18-year study period, the authors found only 12 cases of vasculitis and cancer diagnosed within the same 12-month period: Six patients had lymphoproliferative disorders, and 6 had solid tumors. In most cases, the vasculitis responded partially to immunosuppressive therapy, but investigators observed a more impressive improvement in vasculitis with definitive treatment for the underlying malignancy. A more recent study found 20 cases of malignancy among 200 patients with ANCA-positive vasculitis; 6 were diagnosed concurrently with a diagnosis of vasculitis, and 14 predated vasculitis by a median of 96 months. Only 4 of 20 malignancies in this series were lymphoproliferative; the remaining malignancies were solid organ tumors.

Vasculitis associated with underlying malignancy is often poorly responsive to conventional therapy directed against the vasculitis. In one series of 13 patients with cutaneous vasculitis and lymphoproliferative or myeloproliferative disorder, symptoms of vasculitis were poorly responsive to therapy with nonsteroidal anti-inflammatory drugs (NSAIDs), glucocorticoids, antihistamines, and antiserotonin agents. Although investigators described lessening of the severity of vasculitis after chemotherapy directed against the malignancy, they generally found chemotherapy to be ineffective. Of 13 patients identified, 10 died as a direct result of the malignancy. Similarly, Hutson and Hoffman found general concurrence between improvement in vasculitic syndrome and definitive treatment for the associated underlying malignancy. More recently, a study of cutaneous small vessel vasculitis associated with solid tumors (most commonly lung, breast, and prostate) found a significant response to immunosuppressive therapy directed against the vasculitis, although concurrent treatments for the underlying malignancy were undertaken.

**Cryoglobulinemia**

Cryoglobulins are immunoglobulins that precipitate at reduced temperature. Cryoglobulinemia can be characterized by hyperviscosity symptoms or by vasculitis. Patients often have fatigue, arthralgia or arthritis, cutaneous vasculitis or purpura, neuropathy, digital ischemia, and visceral organ involvement (renal or pulmonary). Three types of cryoglobulins have been identified:

Type I: Monoclonal immunoglobulin, either IgG or IgM; this type is associated with lymphoproliferative disorders.

Type II: Monoclonal IgM directed against polyclonal IgG; type II cryoglobulins were initially thought to be idiopathic and were known as mixed essential cryoglobulinemia. With the identification of hepatitis C virus (HCV), it has been discovered that most of these patients have HCV infection that is directly involved in the pathogenesis of the cryoglobulins. Specific epitopes of HCV antigens are recognized by IgG components of immune complexes, and viral particles are found in the cryoprecipitate. One study found clonal B cell populations in the peripheral blood of 48% of HCV-positive patients with type II cryoglobulinemia, many of whom were eventually diagnosed with a B cell malignancy. Overall, it is estimated that approximately 5% to 8% of patients with mixed cryoglobulinemia may go on to develop non-Hodgkin lymphoma, usually after 5 to 10 years of cryoglobulinemia. The risk of developing non-Hodgkin's lymphoma among HCV-positive cryoglobulinemic patients may be 35 times higher than in the general population. Other data suggest that HCV infection may be associated with other hematologic malignancies. At this time, the subset of patients with mixed cryoglobulinemia who will develop lymphoma cannot be predicted.

Type III: Mixed polyclonal IgG and IgM; type III cryoglobulins are commonly seen with a variety of illnesses, including connective tissue diseases (systemic lupus erythematosus [SLE] and RA) and infections. In one study of 607 patients diagnosed with mixed cryoglobulinemia, 27 cases of hematologic malignancy were identified. Of these, systemic autoimmune diseases were detected in 56% of cases of non-Hodgkin's lymphoma.

**Panniculitis**

The fasciitis-panniculitis syndrome, which includes eosinophilic fascitis, is characterized by swelling and induration of the skin that extends into deeper subcutaneous tissues and is associated with fibrosis and chronic inflammation. Patients may develop arthritis and subcutaneous nodules similar to those seen in erythema nodosum. The arthropathy seems to be secondary to periarticular fat necrosis, can
be monoarticular or polyarticular, and may mimic RA or juvenile RA. Blood and tissue eosinophilia is commonly, but not always, present. This syndrome can be idiopathic and have a benign course, or it can be secondary to a variety of infectious, vascular, or traumatic events. In a few patients, the fasciitis-panniculitis syndrome is associated with an underlying malignancy. Hematologic malignancies are most often associated with this syndrome and are usually diagnosed concurrently or within the first year. Pancreatic cancer and pancreatitis also can be associated with this syndrome. Patients with cancer-associated fasciitis-panniculitis syndrome are predominantly female and are generally refractory to prednisone.

**Palmar Fasciitis**

Palmar fasciitis with arthritis is a syndrome characterized by progressive bilateral contractures of the digits, fibrosis of the palmar fascia, and inflammatory polyarthritis. The metacarpophalangeal and proximal interphalangeal joints are most commonly affected; other affected joints include the elbows, wrists, knees, ankles, and feet. Indurated reticulate palmar erythema can also be seen as part of the palmar fasciitis spectrum. Palmar fasciitis is almost uniformly associated with the presence of an underlying malignancy, most often ovarian, breast, gastric, and pancreatic tumors. Although initially thought to be an atypical variant of reflex sympathetic dystrophy, the severity of manifestations, bilateral presentation, and strong association with occult malignancy suggest that, in these cases, palmar fasciitis is a distinct entity that behaves as a paraneoplastic syndrome. Glucocorticoids, chemotherapy, or both do not seem to result in improvement, although fasciitis occasionally regresses with treatment of the underlying malignancy.

**Reflex Sympathetic Dystrophy**

Reflex sympathetic dystrophy and a variant, shoulder-hand syndrome, are characterized by regional pain, swelling, vaso-motor instability, and focal osteoporosis in a given limb; this condition is thought to be caused by sympathetic dysfunction. Absence of associated antecedent factors, such as stroke, myocardial infarction, or trauma, and failure to respond to conventional therapy warrant a search for an underlying malignancy. A variety of malignancies have been associated with the development of reflex sympathetic dystrophy or its variants. Pancoast tumor of the lung apices and other malignancies that infiltrate the stellate ganglion or brachial plexus have been described in patients with reflex sympathetic dystrophy. Therapy directed against the underlying malignancy may lead to some amelioration of symptoms associated with reflex sympathetic dystrophy.

**Erythromelalgia**

Erythromelalgia is an enigmatic condition characterized by attacks of severe burning, erythema, and warmth of the extremities with symptoms predominantly involving the feet. Symptoms are often exacerbated when the extremities are placed in a dependent position, during ambulation, or during exposure to increased temperatures. Partial relief can be obtained through elevation or cooling of the extremity. This disorder can occur idiopathically (60%) or secondary to another disease (40%). Myeloproliferative disorders, including polycythemia vera and essential thrombocytoysis, are common primary causes and have been found to precede the diagnosis of erythromelalgia by several years. The underlying pathophysiology of this disease is unknown; however, it is often associated with thrombocytoysis. In the largest published retrospective cohort, 168 patients at the Mayo Clinic were identified with this diagnosis between 1970 and 1994. The authors found that after a mean follow-up of 8.7 years, 31.9% of patients reported worsening of disease, 26.6% reported no change, 30.9% reported improvement, and 10.6% reported complete resolution of symptoms. Kaplan-Meier survival curves revealed a significant decrease in survival compared with controls. A history of myeloproliferative disease was found in 15 of 168 patients. The exact cause of the symptoms is unclear, but microvascular arteriovenous shunting has been hypothesized. In cases of secondary erythromelalgia, platelet breakdown products and platelet microthrombi may underlie disease pathogenesis. The most effective therapy seems to be the use of daily aspirin, leading to significant relief of symptoms, which is believed to be related to inhibition of cyclooxygenase-1. A host of other therapies have been tried with varying success. Because of the association with myeloproliferative diseases, routine monitoring with complete blood counts is prudent.

**Polymyalgia Rheumatica**

Polymyalgia rheumatica is a disorder affecting older adults that manifests with discomfort and stiffness in the shoulder and hip girdle, fatigue, anemia of chronic disease, and elevated erythrocyte sedimentation rate (ESR). Classically, this condition responds to moderate doses of prednisone within 48 hours. A variety of other conditions can have presentations that mimic polymyalgia rheumatica, including other rheumatic disorders, systemic infections, and malignancy. Although the association between polymyalgia rheumatica and malignancy has been controversial, atypical features of polymyalgia rheumatica may suggest the presence of occult malignancy, including age younger than 50 years, limited or asymmetric involvement of typical sites, ESR less than 40 mm/hr or greater than 100 mm/hr, severe anemia, proteinuria, and poor or delayed response to 20 mg daily of prednisone. Kidney, lung, and colon cancer and multiple myeloma are most often found in patients presenting with atypical polymyalgia rheumatica. One study of patients undergoing evaluation for possible polymyalgia rheumatica found 10% to have a diagnosis of malignant neoplasms. In contrast, several prospective studies have shown that patients who present with classic polymyalgia rheumatica or temporal arthritis do not seem to have an increased risk of developing malignancy over age-matched controls.

**Raynaud’s Phenomenon and Digital Necrosis**

The development of digital necrosis or profound Raynaud’s phenomenon may suggest the presence of infection, inflammatory disease, or an underlying malignancy. In patients
older than 50 years, the development of Raynaud’s phenomenon, particularly in an asymmetric fashion or in association with digital necrosis, should raise the possibility that this is a paraneoplastic process. These features often antedate the diagnosis of the malignancy by an average of 7 to 9 months. A variety of solid tumors and lymphoproliferative disorders have been associated with this syndrome. Certainly, the presence of digital necrosis in patients with dermatomyositis is highly suggestive of the presence of an underlying malignancy. Mechanisms proposed include cryoglobulinemia, immune complex–induced vasospasm, hypercoagulability, marantic endocarditis with emboli, and necrotizing vasculitis. Therapy with interferon-α also has been reported in association with the development of Raynaud’s phenomenon and digital necrosis.

Remitting seronegative symmetric synovitis with pitting edema

Remitting seronegative symmetric synovitis with pitting edema (RSPE) is an uncommon disorder primarily affecting the metacarpophalangeal joints and the wrists. Although the underlying cause and pathogenesis of this illness are unclear, lymphoma, myelodysplastic syndromes, and several solid tumors, mostly adenocarcinoma, all have been reported in association with it. One retrospective study from a single center in the United States found that 3 of 14 patients with RSPE developed cancer. Characteristics that suggest possible underlying malignancy include the presence of systemic features, such as fever or weight loss, and a poor response to glucocorticoids.

Multicentric reticulohistiocytosis

Multicentric reticulohistiocytosis is a rare condition characterized by the presence of cutaneous papules often associated with a destructive arthritis of the interphalangeal joints of the hands. The papules are flesh-colored to brown-yellow and are classically present in the periungual region and on the dorsal hands and face. Arthritis mutilans may develop in 50% of cases. The characteristic histologic appearance of tissue infiltration with histiocytes and multinucleated giant cells can be found in affected skin, joints, and occasionally internal organs. Multicentric reticulohistiocytosis has been reported in association with hyperlipidemia, malignancies, and autoimmune diseases. Malignancy has been associated in 25% to 31% of cases, although most literature consists of individual case reports. The most frequently seen malignancies include carcinoma of the lung, stomach, breast, cervix, colon, and ovary.

Lupus-like Syndromes

Lupus-like syndromes are rarely associated with underlying malignancy. Isolated case reports have described lupus-like syndromes with ovarian carcinoma and hairy cell leukemia; subacute cutaneous lupus was reported in a patient with breast carcinoma. Studies on the presence of antinuclear antibodies (ANAs) in patients with cancer have yielded mixed results. Two studies were unable to find a significantly increased prevalence of ANAs in patients with solid tumor or lymphoma compared with healthy controls. In contrast, one smaller study found an increased prevalence in patients with non-Hodgkin’s lymphoma compared with controls (21% vs. 0), and another study found a prevalence of ANAs of 27.7% in 274 patients with various malignancies compared with 6.45% of 140 healthy controls. No predictive features seem to suggest occult malignancy in patients presenting with lupus-like syndromes or positive ANAs.

Antiphospholipid Antibodies

Antiphospholipid antibodies and their association with thromboses have been described as a primary syndrome and a secondary phenomenon in autoimmune diseases, primarily SLE. More recently, antiphospholipid antibodies have been associated with a variety of malignancies. Correlations between antiphospholipid antibodies in cancer patients and thromboembolic events have been less clear, however. Several studies have shown the presence of antiphospholipid antibodies in patients with solid tumors and lymphoproliferative disorders at a higher frequency than the 1% to 5% seen in the general population. An early study of 216 consecutive patients with cancer found 22% positive for anticardiolipin antibodies compared with 3.4% in controls. This study found a two-fold increase in the development of thromboembolism in patients with positive antibodies compared with patients with negative serologies; it also indicated that most thromboembolic events occurred in patients with higher antibody titers. Other studies have confirmed the association between malignancy and antiphospholipid antibodies (12.5% to 68%), but have been unable to show a correlation with thromboembolic events. A correlation between antibody titer and disease activity has been shown in some studies, and decreased survival in others. A review of the literature concluded that antiphospholipid antibodies resolve in one-third of cancer patients after treatment for the underlying malignancy.

Studies of the prevalence of antiphospholipid antibodies in unselected patient populations have shown an association with underlying malignancy. A prospective study in France found that 7% of 1014 consecutive patients admitted to a medical ward had antiphospholipid antibodies. In antibody-positive patients, cancer was the most frequently associated disease. A more recent study in patients presenting with a first ischemic stroke found a significantly higher rate of development of cancer within 12 months in patients who had anticardiolipin antibodies (19% vs. 5%).

Osteomalacia

Osteomalacia is the softening of bones often associated with failure of adequate calcification secondary to renal dysfunction or to lack of vitamin D. Osteomalacia has been associated with benign and malignant solid tissue and mesenchymal tumors. Tumors causing oncogenic osteomalacia have been shown to overproduce fibroblast growth factor 23 (FGF-23), and elevated serum levels of FGF-23 can be detected in patients with this paraneoplastic condition. Octreotide scintigraphy may be a useful tool for
identifying occult tumors. With removal of the tumor, there often seems to be resolution of the osteomalacia and normalization of serum FGF-23 levels.

**Sarcoidosis**

Noncaseating granulomata can occur in numerous settings and are not pathognomonic for sarcoidosis. Granulomata resembling those of sarcoidosis may be found in lymph nodes that drain sites of malignancy. These tumor-related tissue reactions resulting in granuloma formation have been described with many types of malignant lesions, including solid tumors and lymphomas. The clinical and radiographic presentation of sarcoidosis and cancer can be virtually indistinguishable, making it important to pursue aggressive evaluation in a patient with sarcoidosis.

The risk of malignancy developing in patients with an established diagnosis of sarcoidosis is controversial. Some studies have shown an increased risk of developing lung cancer and lymphoma, whereas others have shown no increased risk of cancer over the general population.

**Lymphomatoid Granulomatosis**

Lymphomatoid granulomatosis is a rare disorder with angiodestructive and lymphoproliferative features involving the lung and, less often, the skin and central nervous system. Although lymphocytic infiltration of vessels is a hallmark of the disease, lymphomatoid granulomatosis now seems to fall within the spectrum of lymphoproliferative disorders. Despite the predominance of T cells within inflammatory infiltrates, studies have suggested that an Epstein-Barr virus (EBV)-associated B cell proliferation may underlie the pathogenesis of the disease. Prognosis is generally poor, with a median survival from diagnosis of 14 months, although more recent reports suggest some response to rituximab therapy. Frank lymphomas evolve in 25% of cases.

**INFLAMMATORY MYOPATHIES**

The inflammatory myopathies in adult populations encompass a group of illnesses characterized by an idiopathic immune-mediated attack on skeletal muscle resulting in muscle weakness. Many associations between the inflammatory myopathies and the presence of malignancy have been noted, but the reason for the association remains unknown. Dermatomyositis has classically been associated with occult malignancies, whereas associations between polymyositis and inclusion body myositis are becoming increasingly recognized. A further issue is whether the inflammatory myopathy predates the malignancy and can be considered a primary rheumatic disease with known risks of developing malignancy, or whether it simply represents a manifestation of a paraneoplastic process.

On average, the prevalence of malignancy in association with inflammatory myopathies has been approximately 15% to 25%, which appears to be consistent across populations studied. The frequency of malignancy has ranged, however, from 6% to 60% in patients with dermatomyositis, and from 0 to 28% in patients with polymyositis. Other estimates have placed the incidence of cancer in patients with inflammatory myopathies at five to seven times that of the general population.

Dermatomyositis has been associated with a wide range of malignancies; solid tumors are most often seen in cancer-associated myositis as opposed to lymphoid malignancies. Ovarian, lung, and gastric tumors are most common in European populations, and nasopharyngeal malignancies in Asian populations. Studies have confirmed a strong association between dermatomyositis and malignancy. Hill and colleagues studied a pooled cohort of patients from Sweden, Denmark, and Finland and found 198 cases of cancer in 618 patients with dermatomyositis. The standardized incidence ratio (SIR) for malignancy with dermatomyositis was 3. Similar results were found in a Scottish cohort of 286 patients with dermatomyositis. Of these patients, 77 were found to have underlying malignancies, with an SIR of 3.3 to 7.7. Buchbinder and co-workers used strict histopathologic criteria to classify myositis in patients from Victoria, Australia. This group found 36 cases of cancer in 85 patients diagnosed with dermatomyositis and an SIR of 4.3 to 6.2. All of these studies have suggested that cancers are most commonly diagnosed within 1 to 2 years of diagnosis of dermatomyositis.

In contrast to previous work, studies of Asian populations have shown a higher association of nasopharyngeal carcinomas with dermatomyositis. In a nationwide Taiwanese study, 9.4% of 1012 patients with dermatomyositis were found to have an underlying malignancy, the most common of which was nasopharyngeal cancer, followed by lung cancer. In a smaller study, 66.6% of 15 dermatomyositis patients in Singapore had malignancies, most of which were nasopharyngeal carcinoma. Eight white patients with nasopharyngeal carcinoma within 1 year of diagnosis of dermatomyositis were reported more recently in Tunisia. A similar study from Japan found that cancer was diagnosed in 24% of dermatomyositis patients; most cases involved gastric cancer, a common malignancy in the Japanese population.

In polymyositis, the relative risk for developing internal malignancies seems to be lower than that for dermatomyositis, but it is consistently increased over that expected in the general population. Studies have found a 14% to 30% prevalence of cancer among patients with polymyositis, with SIRs increased to 1.2 to 2.1. The nationwide Taiwanese study found a 4.4% incidence of cancer among 643 patients with polymyositis, with an SIR of 2.15 compared with the general population. Small numbers of many types of cancers were found. These studies confirm results reported in previous large studies of Swedish and Finnish populations and a 1994 meta-analysis of all published case-control and cohort studies of malignancy and myositis, which identified an odds ratio for the association of cancer with dermatomyositis of 4.4, and of cancer with polymyositis of 2.1.

In amyopathic dermatomyositis, a variant of dermatomyositis in which typical cutaneous manifestations are present with subclinical or no identifiable muscle disease, the association with underlying malignancy is now...
becoming clear. Whereas previous published reports were limited to small groups of patients,118-119,134-136 recently published cohort studies and a systematic review of the myopathic dermatomyositis literature have suggested similar associations between myopathic dermatomyositis and internal malignancies.127,132,133

Far less is known about the association of inclusion body myositis with underlying malignancy. In Buchbinder’s study from Northern Europe, 52 patients were identified with inclusion body myositis.122 Of these patients, 12 were found to have internal malignancies, with an SIR of 2.4. The numbers of each type of cancer seen are too small to reveal specific associations.

Not all studies concur regarding the association between the inflammatory myopathies and malignancy.134,135 In a study done at the Mayo Clinic, patients with myositis did not seem to be at statistically significant risk for the development of malignancy. No clinical differences were seen between patients who developed a malignancy and patients who did not.135

Despite the negative results of some studies, it seems that most work supports the notion of an increased risk of malignancy in association with dermatomyositis and polymyositis. For patients in whom an inflammatory myopathy has been diagnosed, a workup for the presence of malignancy should be done. The extent of this workup has been debated, however, because extensive undirected searches often result in a very low yield. It is probably rare for an undirected workup to yield evidence of malignancy in polymyositis and dermatomyositis patients;136; any workup should be tailored to the individual patient’s age, symptoms, and signs. Studies have suggested that imaging of the chest, abdomen, and pelvis may increase the potential for discovery of underlying malignancy.137,138 Other studies have suggested the use of serum tumor markers (CA125 and CA19-9) to augment detection of patients with dermatomyositis or polymyositis at highest risk for associated malignancy.139 More recently, a prospective study of whole body positron emission tomography (PET)/computed tomography (CT) was found to be comparable with broad conventional screening (including chest, abdominal, and pelvic CT scans, among other tests).140 Malignancies associated with inflammatory myopathies have been known to develop many years after the diagnosis of muscle disease, so continued vigilance and repeated screening for malignancy are warranted.

In certain cohorts, the risk of malignancy may be higher, including those with older age at diagnosis,141 evidence of distal extremity weakness,142 and prominent pharyngeal and diaphragmatic involvement.142 Patients with myositis-associated autoantibodies may be at less risk for the development of malignancy.142-144 More recent work has suggested that the presence of leukocytoclastic vasculitis145 and cutaneous ulceration,146 as well as the absence of pulmonary involvement,141 increases further the possibility that an underlying malignancy is present.

Although the pathogenesis is unknown, the types of malignancy associated with inflammatory myopathies have been varied, including adenocarcinomas of the breast, ovaries, and stomach. Most cases of dermatomyositis and malignancy seem to occur within 1 year of each other, with myositis diagnosed first in most cases.148 When identified, removal of the malignancy may result in improvement of the myopathic process, which further supports the paraneoplastic nature of myositis in some cases.146

**RISKS OF DEVELOPING LYMPHOPROLIFERATIVE DISORDERS IN RHEUMATIC DISEASES**

Since the 1960s, increasing numbers of reports have described the association between rheumatic disease and the development of malignancy, particularly lymphoproliferative disorder. Table 122-3 shows pre-existing connective tissue diseases that have been associated with malignancy. Much of what is known about associations between rheumatic disease and malignancy is drawn from retrospective and prospective cohort studies, registry linkage studies, small series, and case reports. This is thought to be mediated, at least in part, by chronic immune stimulation and hyperactivity that may lead to malignant transformation. In addition, certain confounding factors need to be considered when the risk of development of malignancy is assessed, including the potential oncogenic properties of many of the immunosuppressive and cytotoxic medications prescribed to treat autoimmune diseases. Lymphoproliferative disorders have developed in patients with rheumatic diseases and in recipients of solid organ transplantsations treated with immunosuppressive agents. EBV has been implicated in the development of lymphoid neoplasia in immunosuppressed patients. In the following sections, many of the rheumatic diseases and the therapies used to treat them are discussed.

**Sjögren’s Syndrome**

Sjögren’s syndrome, an autoimmune exocrinopathy, is characterized by a benign lymphocytic infiltrate of salivary and lacrimal glands that leads to the development of sicca syndrome (keratoconjunctivitis and xerostomia).147 The development of lymphoproliferative disorders in the setting of Sjögren’s syndrome is perhaps the prototypic example of chronic autoimmune disease with increased risk of malignancy. In 1964, investigators first reported the development of four cases of lymphoproliferative disorders in a cohort of 58 patients with Sjögren’s syndrome.148 In 1978, 7 of 136 patients with sicca syndrome were identified as having developed non-Hodgkin’s lymphoma. Compared with the expected incidence of cancer among women of the same age range, a 44-fold increased risk of developing non-Hodgkin’s lymphoma was noted.149 These findings have been reproduced numerous times in other cohorts. Lymphoproliferative disorders complicate approximately 4% to 10% of cases of primary Sjögren’s syndrome.150,151 The relative risk for development of lymphoproliferative disorders in patients with primary Sjögren’s syndrome ranges from 6 to 44,149,152,153 and a meta-analysis of cohort studies has found a pooled SIR of 18.8.149 Most lymphoproliferative disorders were non-Hodgkin’s lymphoma, specifically mucosa-associated lymphoid tissue (MALT) lymphoma, other marginal-zone lymphomas, and diffuse large B cell lymphoma.155,156 Waldenström’s macroglobulinemia, chronic lymphocytic leukemia, and multiple myeloma were more rarely reported.149,153,154,159
Table 122-3 Pre-existing Connective Tissue Diseases Associated with Malignancy

<table>
<thead>
<tr>
<th>Connective Tissue Disease</th>
<th>Malignancy</th>
<th>Associated Factors</th>
<th>Clinical Alert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sjögren’s syndrome</td>
<td>Lymphoproliferative disorders</td>
<td>Glandular features: lymphadenopathy, parotid or salivary enlargement</td>
<td>Clues to progression from pseudolymphoma to lymphoma include worsening of clinical features, disappearance of rheumatoid factor, and decline of IgM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extraglandular features: purpura, vasculitis, splenomegaly, lymphopenia, low C4 cryoglobulins</td>
<td></td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Lymphoproliferative disorders</td>
<td>Presence of paraproteinemia, greater disease severity, longer disease duration, immunosuppression, Felty’s syndrome</td>
<td>Rapidly progressive, refractory flare in long-standing rheumatoid disease may suggest an underlying malignancy</td>
</tr>
<tr>
<td>SLE</td>
<td>Lymphoproliferative disorders</td>
<td>—</td>
<td>Non-Hodgkin’s lymphoma should be considered in SLE patients who develop adenopathy or masses; lymphoma of the spleen is another cause of splenic enlargement in SLE Poorly healing skin lesion within discoid plaques should be evaluated</td>
</tr>
<tr>
<td>Discoid lupus erythematosus</td>
<td>Squamous cell epithelioma</td>
<td>Found in oldest plaques, ≥20 yr after onset of discoid lesion, primarily in men 30-60 yr old</td>
<td>Annual chest radiograph after fibrosis is detected Changes in skin features or poorly healing lesions should be evaluated Esophagoscopy and biopsy, if indicated, of distal esophageal constricting lesions</td>
</tr>
<tr>
<td>Systemic sclerosis (scleroderma)</td>
<td>Alveolar cell carcinoma</td>
<td>Pulmonary fibrosis, interstitial lung disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nonmelanoma skin cancer</td>
<td>Areas of scleroderma and fibrosis in the skin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adenocarcinoma of the esophagus</td>
<td>Barrett’s metaplasia</td>
<td></td>
</tr>
<tr>
<td>Paget’s disease of bone</td>
<td>Osteogenic sarcoma</td>
<td>Development of severe pain; increasing incidence with age</td>
<td></td>
</tr>
<tr>
<td>Dermatomyositis</td>
<td>Ovarian, lung, and gastric cancer in Western populations; nasopharyngeal carcinoma in Asian populations</td>
<td>Older age, presence of cutaneous vasculitis; less likely in setting of myositis-specific antibodies or interstitial lung disease</td>
<td>Malignancy evaluation needs to be tailored to individual patient’s age, symptoms, and signs</td>
</tr>
</tbody>
</table>

SLE, systemic lupus erythematosus.

Generally, the development of lymphoma is a late manifestation of Sjögren’s syndrome, often seen after 6.5 years of disease. Clinical and laboratory features seem to be associated with or predictive of development of lymphoproliferative disorders, including palpable purpura, cutaneous ulcerations, cryoglobulinemia, low serum complement levels, monoclonal gammopathies, cytopenias, splenomegaly, and adenopathy. Progression to high-grade lymphoma portends a poor prognosis. In contrast, the incidence of other malignancies or all-cause mortality was not increased in patients with Sjögren’s syndrome compared with the general population. It is believed that chronic B cell stimulation may lead to the malignant transformation of clonal lines characteristic of Sjögren’s syndrome. The presence of a viral trigger accounting for malignant transformation is one possible theory. EBV, among other viruses, has been implicated, but studies have failed to find EBV or other viral particles in lymphoma specimens associated with Sjögren’s syndrome.

Other reports have described the presence of chromosomal translocations with increased frequency in patients with Sjögren’s syndrome who have developed lymphoma. One group of investigators identified the presence of translocations of the proto-oncogene \( bcl-2 \) in 5 of 7 patients with Sjögren’s syndrome and lymphoma by the use of polymerase chain reaction. Such translocations were found in peripheral blood or bone marrow in 5% of unselected patients with Sjögren’s syndrome without evidence of lymphoma in another study. Conversely, no evidence of \( \text{bcl-2} \) translocations was present in 50 salivary gland biopsy specimens of patients with Sjögren’s syndrome without evidence of lymphoma. Analysis of biopsy specimens taken before the development of lymphoma from the 7 patients previously mentioned revealed no evidence of \( \text{bcl-2} \) translocation. Translocation seemed to correlate with the development of lymphoma in at least a subset of patients with Sjögren’s syndrome, and the use of polymerase chain reaction technology may allow early detection of malignant transformation.

Rheumatoid Arthritis

Data from numerous studies since the 1970s are persuasive that RA is associated with a twofold to threefold increased risk for the development of lymphoproliferative disorders,
the magnitude of which has remained constant despite dramatic changes in therapy. Many factors, including chronic inflammation and immune dysregulation, in addition to potential oncogenic properties of immunosuppressive therapies for the treatment of RA, must be considered when the risk of development of hematologic malignancies is evaluated. It is often difficult to separate the effects of medication use from the underlying severity of inflammation that makes medication use necessary or indicated, a concept termed confounding by indication. This association has been highlighted further by widespread use of tumor necrosis factor (TNF) inhibitors for patients with refractory disease and the potential for these medications to interfere with innate immune tumor surveillance.

In 1978, an SIR of 2.7 for lymphoma was reported in a group of 46,101 Finnish RA patients compared with the general population. A similarly increased risk of 2.4 for lymphoma was seen later in a group of 20,699 Danish patients, and an SIR of 1.9 to 2 was reported in a large cohort of 76,527 Swedish patients. In the United States, an increased risk of 1.9 was found in a cohort of 18,527 patients, and an SIR of 2.2 was noted in a separate cohort of 8458 patients 65 years old and older. In the United Kingdom, an SIR of 2 to 4.4 for lymphoma was observed in an inception cohort of 2015 patients with inflammatory arthritis compared with the general population, and an SIR of 2.04 to 2.39 was seen for non-Hodgkin’s lymphoma in a cohort of 26,623 RA patients in Scotland. A meta-analysis of nine cohort studies of RA patients found a pooled SIR of 3.9 for lymphoma using a random effects model. Canadian investigators found an increased risk of leukemia (SIR, 2.47) among RA patients, but were unable to confirm elevated rates of lymphomas compared with the general population. Data from case-control studies of patients with non-Hodgkin’s lymphoma have shown similar results: Odds ratios of 1.3 to 1.5 were found for underlying RA. These associations have been recently confirmed in other populations, including patients from Japan, Taiwan, California, and Spain. In general, lymphomas in patients with RA seem to be predominately diffuse large B cell type and recently have been shown to favor nongerminall center subtypes.

Most studies have suggested that the risk for development of lymphoma is related to the degree of inflammation. The Swedish group identified high inflammatory activity (defined by ESR, swollen and tender joint counts, and the physician’s global assessment of disease activity) as a significant risk factor, with an odds ratio of 25.8 compared with low disease activity. No association between any specific drug and the development of lymphoma was identified; however, the cohort examined was treated between 1965 and 1983, and few of these patients were apparently treated with immunosuppressive drugs, making the lack of association less certain. In a follow-up case-control study of 378 lymphomas in a Swedish group of RA patients published more recently, a 71-fold increased risk of lymphoma was reported in patients with high cumulative disease activity compared with low disease activity. Immunosuppressive therapy did not seem to modify risk for lymphoma in this study. Patients did not have increased rates of lymphoma development before disease onset. Patients with Felty’s syndrome (a variant of RA associated with neutropenia and splenomegaly) were found in a Veterans Affairs study of 906 men to have a twofold increase in total cancer incidence, but a 12-fold increase in risk of non-Hodgkin’s lymphoma.

Disease-Modifying Antirheumatic Drug Therapy

Several studies have looked at the contribution of disease-modifying antirheumatic drug (DMARD) therapy to the elevated risk of malignancy in RA patients. A prospective, observational study was performed in a group of Canadian RA patients enrolled in a DMARD registry. Although this study found an increased rate of lymphoproliferative disorders in this cohort compared with the general Canadian population (SIR, 8.05), no significant differences in DMARD exposure were noted between patients who developed malignancy and patients who did not. A second group of Canadian investigators similarly identified an increased risk for the development of lymphoma and myeloma in RA patients overall compared with control groups. In this study, the risk of lymphoma and myeloma seemed to be fourfold greater in the RA group when DMARD use was not controlled for and 3.4-fold greater when individual DMARD use was controlled for. Despite the low level of DMARD exposure in this population, no strong effect of DMARD use was seen.

Similar effects of DMARD use were seen in the study of Swedish patients with RA and lymphoma: Treatment with any DMARD (odds ratio [OR], 0.9) or specific use of methotrexate (OR, 0.8) did not seem to be associated with increased risk of lymphoma compared with DMARD-naïve RA patients; however, no patients had been treated with TNF inhibition. In contrast, a European cohort of RA patients enrolled in a DMARD registry was evaluated longitudinally for the development of malignancies. Investigators found an increased risk of lymphoproliferative disorders in patients with the highest cumulative exposure to DMARDs compared with patients with less than 1 year of exposure (SIR, 4.82).

Although inconclusive, data from these studies when taken together suggest a possible increased risk for the development of lymphoproliferative disorders in RA patients treated with DMARDs. More recent studies have suggested, however, that this increased risk may be due to the duration and severity of the underlying disorder, rather than to specific medication use. Associations of specific immunosuppressive therapy for the treatment of autoimmune diseases are discussed further in Chapters 61 and 62.

Risk of Solid Tumor in Patients with Rheumatoid Arthritis

Despite persuasive evidence of increased risks of lymphoproliferative disorders associated with underlying RA, overall rates of all-site malignancies do not seem to be higher compared with the general population. The overall “null” result of all malignancies is due to the combination of increased risk of lymphoproliferative disorders offset by an apparently decreased risk of colorectal malignancies. The decreased risk of colorectal
cancer has been attributed to long-term use of NSAIDs among RA patients. Aside from lymphoproliferative disorders, only a few solid tumors, including lung cancer and skin cancer, have been associated with RA.

Increased risk of lung cancer in RA patients has been seen in multiple studies.173-177,182-187,196-198 A study evaluating three separate RA cohorts (an inpatient registry of 53,067 prevalent cases of RA, an inception cohort of 3703 incident RA cases, and a registry of 4160 RA patients treated with TNF inhibitors) found a consistently increased risk of lung cancer in all cohorts (SIR, 1.48 to 2.4) compared with the general population.81,199 This association may be related to tobacco use, which seems to be a common risk factor for the development of RA, in addition to its well-known association with lung cancer,199 although the particular association of lung cancer among RA patients who smoke is unknown. Similarly, a study of lung cancer in a cohort of 8768 U.S. veterans (92% male) with RA found an increased risk of lung cancer (OR, 1.43; 95% confidence interval [CI], 1.23 to 1.65) after adjustment for covariates such as age, gender, and tobacco exposure.200

A slightly increased risk for the development of nonmelanoma skin cancer (most commonly basal cell carcinoma and squamous cell carcinoma) has been noted in several studies.175,176,196,201,202,203 although the significance of these tumors, which carry a low probability of metastasis, is unclear. Unfortunately, nonmelanoma skin cancer is rarely captured in national cancer registries, so its incidence among the general population or subpopulations such as RA is difficult to quantify or compare. Furthermore, important risk factors for nonmelanoma skin cancer, including ultraviolet light exposure, are almost impossible to quantitate in observational studies. What is of greater concern, however, are newer data suggesting an increased rate of melanoma among patients with RA.182,185,202-204 Because of the suggestion of increased risk of skin cancer, whether attributed to underlying disease or immunosuppressive therapies, it is reasonable to suggest periodic skin examinations in RA patients, particularly those with other risk factors, including smoking and increased ultraviolet light exposure. Certainly, all suspicious lesions should be evaluated by a dermatologist and biopsy strongly considered.

**Systemic Lupus Erythematous**

The risks of developing malignancy in association with SLE have been difficult to estimate in the past. Small series and cohort studies have noted that patients with SLE might be at increased risk for malignancy, including non-Hodgkin’s lymphoma, sarcoma, and breast carcinoma.205-209 Other small series have not found differences,210 however, or have found infrequent associations211 in number or type of malignancy between patients with lupus and the general population.210 Conflicting results also are seen in case-control studies of larger groups of patients, with some cohorts showing an increased overall risk of malignancy,212-214 although others have failed to do so.215-218 Some studies that did not find increased risk of overall malignancies in patients with SLE have shown increased risk of lymphoproliferative disorders, however.160,162,216-218 Confounding factors complicating interpretation of these studies include possible incomplete ascertainment of malignancies, inclusion of nonrepresentative cohorts of patients with SLE, and selection of inappropriate control populations.199

To determine more adequately whether individuals with SLE are at increased risk, studies of large multinational cohorts, systematic reviews, and meta-analyses of pooled data are necessary. The SIR of individual studies has ranged from 1.1 to 2.6.219 A meta-analysis of six of the clinical cohort studies found a slightly increased risk of overall malignancies in cohorts of patients with SLE, with an SIR of 1.58.220 This analysis showed an increased risk of lymphoma in these cohorts, with an SIR of 3.57 for non-Hodgkin’s lymphoma and 2.35 for Hodgkin’s disease. A separately performed meta-analysis of the incidence of lymphoma in patients with SLE found an SIR of 7.4.221 Individual hospital discharge database studies have shown a consistently higher risk of non-Hodgkin’s lymphoma (SIR, 3.72 to 6.7), but these studies examined only hospitalized patients with SLE.200 Pooled analysis showed a slightly elevated risk for the development of breast cancer, with an SIR of 1.53, but they did not find an increased risk of lung or colorectal cancer in these patients.221 The same confounding factors influencing individual studies are a factor in interpreting these pooled data.

A more recent series of studies analyzing nearly 9500 lupus patients (>7,000 patient-years of observation) in a multinational cohort study has helped to define better potential associations with malignancy.222,223 The authors found a slightly increased risk of malignancies overall (SIR, 1.15) with higher risks for the development of hematologic malignancies (SIR, 2.75), particularly non-Hodgkin’s lymphoma (SIR, 3.64).222 Forty-two cases of non-Hodgkin’s lymphoma were identified, most of which were of aggressive histologic subtypes.224 The incidence of lymphoma in this study was evidently early in the course of SLE, rather than after many years of chronic disease activity or use of multiple immunosuppressive medications.213,225 The elevated risk of non-Hodgkin’s lymphoma in this cohort seemed to be independent of race or ethnicity, although white patients seemed to have higher rates of malignancy in general compared with patients of other ethnicities.226 In a case-control study within the multisite international SLE cohort, age, disease-related damage, and smoking were found to be associated with increased risk of malignancy; use of immunosuppressant medications (particularly lagged 5 years) may contribute to increased risk of hematologic malignancy.227

Although the exact cause of the association is unknown, several theories have arisen to explain the possible connection between SLE and malignancy, especially B cell lymphoma. Some authors have postulated that certain immunologic defects may predispose patients to SLE and B cell lymphoma, including apoptosis dysfunction, chronic antigenic stimulation, and overexpression of bcl-2 oncogene.220,226 Viruses, EBV in particular, also have been postulated as part of the development of SLE and lymphoma.220,228 Studies have not conclusively validated any of these theories to date, however.

The relative prevalence of cervical cancer in SLE patients is difficult to estimate because national cancer registries often do not record malignancies in situ. Cervical cancer remains an important issue for women with SLE, and an increased risk may come about for different reasons,
including (1) reduced clearance of human papillomavirus (HPV), the causal agent in most cases of cervical cancer; (2) increased risks of cervical cancer associated with immunosuppressive medications; and (3) reduced rates of routine Pap smears and other screening procedures in a population of patients with chronic illnesses. In the large multinational study performed more recently, the SIR for invasive cervical cancer was found to be elevated at 1.26, albeit with confidence intervals that cross the null. Other studies have confirmed increased risks of abnormal Pap smears and cervical dysplasia in women with SLE. Different studies have implicated increased prevalence of HPV infection and other sexually transmitted diseases, and immunosuppression may partly explain this association. As with mammography, women with SLE seem less likely to undergo routine Pap testing than women in the general population.

Several studies have identified a link between SLE and the development of lung cancer. Indeed, the largest multinational cohort of SLE subjects found an increased incidence of lung cancer compared with the general population (SIR, 1.37; 95% CI, 1.05 to 1.76). Further analyses of the cases of lung cancer in this cohort revealed a variety of tumor types, including adenocarcinoma, bronchoalveolar carcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma, and carcinoid tumor. Most cases (71%) occurred in smokers, 25% of cases were in men, and few (20%) had previous exposure to immunosuppressive agents. To date, no particular demographic or clinical features have been found to explain this apparently increased risk.

Overall, the presence of SLE seems to carry a small increased risk for the development of cancer, particularly lymphoproliferative disorders such as non-Hodgkin’s lymphoma, lung cancer, and cervical cancer. The underlying causes of these associations are unknown, but they do not seem to be exclusively related to the use of immunosuppressive or cytotoxic agents. Data suggest that lupus patients may be less likely to receive recommended cancer screening.

Systemic Sclerosis

Although data are conflicting, most evidence suggests that individuals with systemic sclerosis seem to have an increased risk of developing malignancy. Estimates of the prevalence of cancer among scleroderma patients range from 3.6% to 10.7%. However, 13% of deaths are reported among patients with systemic sclerosis. The malignancies that have been implicated are often observed in organs affected by inflammation and fibrosis, including the lung, breast, esophagus, and skin. The SIR of malignancy in the scleroderma population is 1.5 to 5.1 compared with that of the general population. An apparent increase has been reported in the observed number of cases of lung cancer that occur in the setting of pulmonary fibrosis, but not in association with tobacco use.

A different study evaluating 20 lung cancers among 632 scleroderma patients in Australia found that cigarette smoking, but not underlying pulmonary fibrosis, was associated with increased risk of lung cancer. Studies have been mixed regarding increased risk or a temporal relationship between scleroderma and breast cancer. Older age at the time of diagnosis of systemic sclerosis seems to be a significant risk factor for the development of cancer. Data are mixed regarding associations between systemic sclerosis-specific autoantibodies and the development of cancer. Selected studies support potential associations, whereas others do not. A recent study found a close temporal relationship to cancer (within 1 to 2 years of scleroderma diagnosis) in scleroderma patients with positive antibodies for RNA polymerase I/III. Expression of RNA polymerase 1 and RNA polymerase III was detected in the tumors of affected individuals.

Although the SIR for all malignancies is 1.5 to 2.4, the incidence ratio for lung cancer can be as high as 7.8, and for non-Hodgkin’s lymphoma, 9.6. Cases of non-Hodgkin’s lymphoma seem to be more likely to occur within the first year of diagnosis of systemic sclerosis. Elevations in incidence have been found for other specific cancers as well, including nonmelanoma skin cancers (4.2), primary liver cancers (3.3), and hematopoietic cancers (2.3)—all having a higher incidence than is seen in the general population. The greatest risk seems to correspond to areas commonly affected by fibrosis, particularly the lung and skin. Esophageal involvement, common to limited and diffuse systemic sclerosis, is the likely cause for an increased incidence of Barrett’s esophagus (12.7%) and development of esophageal cancer (SIR, 9.6). In contrast to the data previously described, one study found no increase in overall or specific malignancies in patients with systemic sclerosis (SIR, 0.91 overall). Localized scleroderma, including morphea and linear scleroderma, does not seem to convey an increased risk of malignancy. Several reports have described the development of postirradiation morphea in patients treated for breast cancer.

**PRIMARY TUMORS AND METASTATIC DISEASE**

Primary Musculoskeletal Tumors

This section does not provide in-depth knowledge of the primary tumors of the musculoskeletal system. Rather, it provides a reference to the most common primary malignant musculoskeletal tumors and symptoms that may arise in association with them. Primary tumors of bone, including benign and malignant tumors, are discussed in greater detail in Chapter 123.

A primary malignant bone cancer is any neoplasm that develops from the tissues or cells found within bone that has the ability to metastasize. Neoplasms may develop or arise from any of the types of cells present within the bone—osteoblasts, chondrocytes, adipose and fibrous tissue, vascular cells, hematopoietic cells, and neural tissue. A neoplasm developing from any of these tissues is called a sarcoma, which signifies that it is derived from mesenchymal tissue. The bone sarcomas are named for the predominant differentiated tissue type, such as osteosarcomas, chondrosarcomas, liposarcomas, and angiosarcomas.

The most common manifestation of these tumors is the development of pain in the area of the lesion, which may

---

*References 205, 216, 218, 220, 227, 239.*
be accompanied by a sympathetic effusion or stiffness in the surrounding joint. This discomfort does not seem to be activity related and is often worse at night. These tumors can manifest, however, as painless masses or as pathologic fractures. Systemic features, such as fatigue, malaise, weight loss, fevers, and night sweats, are rare with all of these tumors except for Ewing’s sarcoma. Primary malignant bone tumors are uncommon, particularly compared with other types of cancer. They have their highest incidence in childhood and adolescence and constitute 3.2% of childhood malignancies that occur before age 15 years. The incidence has been reported in this age group as 3 per 100,000 individuals. These tumors commonly arise out of areas of rapid growth, with the most common site of primary bone sarcomas being the metaphysis near the growth plate.

Table 122-4 lists the most common types of primary malignant bone tumors. Osteosarcoma is the most common of the tumors and generally occurs in individuals in the second decade of life or in elderly individuals. Osteosarcoma can occur secondary to radiation therapy delivered as treatment for other malignancies. Paget’s disease of bone can rarely (<1% of cases) proceed to malignant transformation. Severe pain in the setting of Paget’s disease may signal transformation to osteogenic sarcoma. Tumors most frequently affect the femur, humerus, skull, and pelvis and can result in pathologic fractures. Survival is usually less than 1 year. Differentiating malignancy from Paget’s disease may require a biopsy.

Chondrosarcoma has been reported as the second most common of the malignant bone tumors. This tumor may occur as a primary tumor or as a malignant transformation in the setting of benign lesions, such as enchondromas or osteochondromas. Fibrosarcoma is significantly less common than the previously mentioned tumors and accounts for less than 4% of primary malignant bone tumors.

As a group, round cell tumors include primary lymphoma of bone, Ewing’s sarcoma, and metastatic neuroblastoma. Ewing’s sarcoma is a common primary bone tumor of childhood. Giant cell tumors as a group account for 4.5% of bone tumors. They usually arise from the metaphysis or epiphysis of long bones, generally around the knee. Most are benign, but a few are malignant lesions, usually arising out of a previously irradiated benign giant cell tumor.

In addition to the primary malignancies of bone, a plethora of malignant tumors can arise from mesenchymal connective tissue; these are also known as sarcomas. They can result in joint complaints, but more often result in soft tissue complaints. They are very rare. Rhabdomyosarcoma is a malignant tumor arising from muscle tissue. It is the fourth most common solid tumor in children and is responsible for more than half of all soft tissue sarcomas in children. Rhabdomyosarcoma rarely occurs in adults. It can appear at any site, and the symptoms most often are referable to the site involved. Most commonly, rhabdomyosarcoma affects the orbits, genitourinary tract, limbs, head and neck, and parameningeal areas. It commonly metastasizes to lymph nodes, lungs, and bone.

### Metastatic Disease

When bone lesions are identified, primary tumors need to be considered, although most malignant lesions in bone are metastatic. Metastasis rarely affects muscles, joints, or adjacent connective tissue. More commonly, it affects bone. The most common sites of metastasis are the spine and pelvis. It is uncommon to find metastatic lesions distal to the elbow, and, although rare, metastasis to the foot is more common than to the hand. When distal or acral metastasis is identified, it is often associated with lung cancer. Primary tumors generally associated with metastases to bone include tumors in the prostate, thyroid, lung, breast, and kidney. Although most skeletal metastases do not produce pain, one of the most common causes of cancer pain is infiltration of bone. The pain can be intense and stabbing or dull. It is often constant rather than intermittent, is worse at night, and often is worse with weight bearing and movement. Rheumatic or arthritic complaints often occur before lesions are easily identified on radiographs. Arthritis associated with metastatic carcinoma is most commonly monoarticular and most commonly affects the knee. Metastases to the hip, ankle, wrist, hand, and foot have been reported, but occur less frequently. Breast and lung cancers are present in most patients. Metastases to the extremities can simulate gout, osteomyelitis, tenosynovitis, or acro-osteolysis. Joint involvement can be related to direct synovial implantation or involvement of the juxta-articular or subchondral bone. Table 122-5 presents the clinical features suggestive of underlying metastases.

Radiographic features of bone tumors can be significant when the duration of disease and the type of malignancy are interpreted. Lesions may be lytic or blastic, and patterns of destruction often reflect the aggressiveness of tumors. Well-circumscribed lesions may be indicative of slower growth, whereas a “moth-eaten” pattern with evidence of cortical destruction typically signifies a more rapid rate of growth. What has been described as a permeative pattern suggests an extremely rapid rate of destruction and is often associated with an extraosseous soft tissue mass. Computed tomography, magnetic resonance imaging, and radionuclide imaging can provide significant information for diagnosis, staging, prognosis, and therapy.

### Table 122-4  Primary Bone Tumors

<table>
<thead>
<tr>
<th>Nonosseous Tumors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple myeloma</td>
<td></td>
</tr>
<tr>
<td>Round cell tumors</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Osseous Tumors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Osteosarcoma</td>
<td></td>
</tr>
<tr>
<td>Chondrosarcoma</td>
<td></td>
</tr>
<tr>
<td>Giant cell tumors</td>
<td></td>
</tr>
<tr>
<td>Fibrosarcoma</td>
<td></td>
</tr>
</tbody>
</table>

### Metastatic Disease

When bone lesions are identified, primary tumors need to be considered, although most malignant lesions in bone are metastatic. Metastasis rarely affects muscles, joints, or adjacent connective tissue. More commonly, it affects bone. The most common sites of metastasis are the spine and pelvis. It is uncommon to find metastatic lesions distal to the elbow, and, although rare, metastasis to the foot is more common than to the hand. When distal or acral metastasis is identified, it is often associated with lung cancer. Primary tumors generally associated with metastases to bone include tumors in the prostate, thyroid, lung, breast, and kidney. Although most skeletal metastases do not produce pain, one of the most common causes of cancer pain is infiltration of bone. The pain can be intense and stabbing or dull. It is often constant rather than intermittent, is worse at night, and often is worse with weight bearing and movement. Rheumatic or arthritic complaints often occur before lesions are easily identified on radiographs. Arthritis associated with metastatic carcinoma is most commonly monoarticular and most commonly affects the knee. Metastases to the hip, ankle, wrist, hand, and foot have been reported, but occur less frequently. Breast and lung cancers are present in most patients. Metastases to the extremities can simulate gout, osteomyelitis, tenosynovitis, or acro-osteolysis. Joint involvement can be related to direct synovial implantation or involvement of the juxta-articular or subchondral bone. Table 122-5 presents the clinical features suggestive of underlying metastases.

Radiographic features of bone tumors can be significant when the duration of disease and the type of malignancy are interpreted. Lesions may be lytic or blastic, and patterns of destruction often reflect the aggressiveness of tumors. Well-circumscribed lesions may be indicative of slower growth, whereas a “moth-eaten” pattern with evidence of cortical destruction typically signifies a more rapid rate of growth. What has been described as a permeative pattern suggests an extremely rapid rate of destruction and is often associated with an extraosseous soft tissue mass. Computed tomography, magnetic resonance imaging, and radionuclide imaging can provide significant information for diagnosis, staging, prognosis, and therapy.

### Table 122-5  Frequent Features of Arthritis Resulting from Metastatic Carcinoma

<table>
<thead>
<tr>
<th>Presence of constitutional symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior history of malignancy</td>
</tr>
<tr>
<td>Protracted clinical course</td>
</tr>
<tr>
<td>Negative culture results, negative crystal analysis</td>
</tr>
<tr>
<td>Medical therapeutic failure</td>
</tr>
<tr>
<td>Rapid reaccumulation of hemorrhagic noninflammatory effusion</td>
</tr>
<tr>
<td>Radiologic evidence of destructive process</td>
</tr>
</tbody>
</table>
Postchemotherapy Rheumatism

Several rheumatic or musculoskeletal manifestations can develop in patients after administration of chemotherapy for the treatment of malignancy. Postchemotherapy rheumatism has been best described in patients treated for breast cancer, but has also been described in other malignancies, including ovarian cancer and non-Hodgkin’s lymphoma.\textsuperscript{277-279} The phenomenon has been described as a noninflammatory, self-limited, migratory arthropathy. Typically, symptoms develop several weeks to several months after the completion of chemotherapy and often include myalgia, stiffness, arthralgia, and arthritis involving the small joints of the hands, ankles, and knees.\textsuperscript{278} It can be mistaken for RA based on its symptoms; however, most patients have little or no evidence of synovial thickening and have no radiographic or serologic evidence to suggest RA. The pathogenesis of this process is unknown; however, it is self-limited, usually lasting less than 1 year, and is best treated in a conservative fashion. Evaluation should be performed to exclude recurrent carcinoma or another inflammatory condition. The medications most frequently implicated in this phenomenon include cyclophosphamide, 5-fluorouracil, methotrexate, and tamoxifen.\textsuperscript{279,280}

Other immunomodulatory agents also have been linked to the development of musculoskeletal findings. Tamoxifen use has been associated with the development of an acute inflammatory arthritis similar to RA;\textsuperscript{281} however, a randomized controlled trial of tamoxifen or placebo in 7145 women at high risk for breast cancer (but without a diagnosis of cancer) did not find an increased prevalence of arthralgia or arthritis symptoms in the tamoxifen arm.\textsuperscript{282} More recently, aromatase inhibitors have been used widely for hormone receptor–positive breast cancer, with increasing reports of associated arthralgia and arthritis. A retrospective, exploratory analysis of the Arimidex Tamoxifen Alone or in Combination (ATAC) trial evaluated the development of joint symptoms in 5433 women with early breast cancer randomized to receive anastrozole alone or tamoxifen alone.\textsuperscript{283} When analysis was limited to women who entered the study without joint symptoms at randomization, 35.2% developed joint symptoms on anastrozole compared with 30.3% on tamoxifen (P < .0001). However, intensity of symptoms was not different between groups. A prospective, nonrandomized study evaluated the prevalence of musculoskeletal symptoms in women who switched from anastrozole to letrozole because of articular symptoms.\textsuperscript{284} Most (61%) were able to continue with letrozole, with 28.5% discontinuing the second aromatase inhibitor owing to severe musculoskeletal symptoms. Therefore, switching of aromatase inhibitors may be a reasonable strategy when an initial agent is intolerable because of arthralgia or arthritis; many women continue to be intolerant to more than one agent.

Biologic agents used for the treatment of chronic viruses or a variety of malignancies can similarly lead to autoimmune phenomenon. Use of interleukin-2 can result in spondyloarthritis or inflammatory arthritis. Interferon-α administration can result in seropositive nodular RA and myalgia and arthralgia.\textsuperscript{285,286} Use of interferon can also result in autoantibody formation and features suggestive of SLE and autoimmune thyroid disease.\textsuperscript{286-288}

Lymphoproliferative and Myeloproliferative Diseases

Leukemia

Leukemia can result in the development of musculoskeletal complaints. Bone pain, the most common musculoskeletal manifestation, has been reported to occur in 50% of adults with leukemia.\textsuperscript{289} Long bone pain is more common in children, whereas axial pain is more common in adults. Generally, bone pain is more common in the lower than the upper extremities.\textsuperscript{290} Overt synovitis can develop in association with acute and chronic leukemia and can lead to the development of monoarticular or polyarticular arthritis.\textsuperscript{291} The pathogenesis seems to be leukemic infiltration of the synovium and subperiosteal tissue. Bleeding or hemorrhage in the joint also may be associated with the process. Most cases of arthritis associated with leukemia are seen in children—14% to 50% compared with 4% to 16.5% in adults.\textsuperscript{292-294}

In a series of adult patients with acute leukemia studied over a 10-year period, 5.8% (8 of 139) of patients presented with rheumatic manifestations. On average, symptoms of arthritis preceded the diagnosis of leukemia by 3.25 months.\textsuperscript{295} The most common patterns of presentation included asymmetric large joint involvement in association with low back pain, followed by symmetric polyarthritis mimicking early RA. Rheumatic manifestations included morning stiffness, low back pain, nonarticular bone pain, pain out of proportion to objective findings, low-grade fever, and elevation of the ESR. The response to NSAIDs, glucocorticoids, and conventional antirheumatic therapy was reportedly poor, but tumor-directed chemotherapy resulted in substantial improvement of rheumatic manifestations. Patients with these manifestations were more likely to exhibit early osteopenia or lytic bone lesions. Ultimately, prognosis and mortality rates were no different between patients presenting with or without rheumatic manifestations.\textsuperscript{295}

In contrast, a large retrospective study of children with leukemia found that 21.4% (36 of 168) with acute lymphoblastic leukemia and 10.5% (6 of 57) with acute nonlymphoblastic leukemia developed symptoms associated with bones and joints. Thirteen of these patients with acute lymphoblastic leukemia had evidence of bony lesions on radiographs.\textsuperscript{296} Many of these children had been incorrectly treated for juvenile RA or osteomyelitis before the diagnosis of leukemia. The group with bone lesions seemed to do very well; their condition might fall into a subgroup of childhood leukemia that has a better prognosis.\textsuperscript{294,296} A more recent study found that the presence of subtle blood count changes and nighttime pain may help distinguish leukemia from juvenile RA.\textsuperscript{294}

Multiple Myeloma

Multiple myeloma is a neoplastic proliferation of plasma cells, causing a nonosseous malignant tumor to arise in the marrow. In contrast to the other primary tumors of bone, which have their highest incidence in children and adolescents, myeloma is a tumor of adults, occurring most
commonly in the fifth and sixth decades of life. The most common musculoskeletal feature of this disease is the development of bone pain. Other hallmark features are diffuse pain and stiffness. Patients characteristically develop osteopenia, and osteolytic lesions are seen on radiographs. Lytic lesions, which can occur in any area of the skeleton, are produced by focal accumulations of plasma cells. Osteosclerotic lesions also have been reported. True arthritis is rare, but cases of arthritis secondary to articular and periarticular invasion with malignant cells have been reported in multiple myeloma and in Waldenström’s macroglobulinemia.

A secondary feature of the disease, which can often lead to additional musculoskeletal complaints, is the development of hyperuricemia and secondary gout. Sjögren’s syndrome and other autoimmune phenomena have also been described in association with multiple myeloma.

### Lymphoma

Musculoskeletal symptoms have been reported in 25% of cases of non-Hodgkin’s lymphoma. The most common musculoskeletal problem associated with lymphoma is the development of bone pain associated with metastasis or lymphoma in the bone. By report, more than 50% of patients have evidence of bone lesions at autopsy; however, few patients actually present with arthritis or bone pain. Nonetheless, non-Hodgkin’s lymphoma has been reported to manifest as a seronegative arthritis with or without other features, such as lymphadenopathy and hepatomegaly, typically seen with this disease. Monarticular and polyarticular involvement can occur. Cases of polyarthritis simulating RA in the setting of non-Hodgkin’s lymphoma have been reported. Although it is unusual to see direct involvement of the synovium, this has also been reported. Cases with radiographic evidence of bone destruction have been associated with non-Hodgkin’s lymphomatosus arthropathy. Suspicion of lymphoma should be heightened in patients in whom severe constitutional symptoms seem out of proportion to the degree of arthritis, especially in patients who are negative for rheumatoid factor.

### Angioimmunoblastic Lymphadenopathy

Angioimmunoblastic lymphadenopathy is a rare lymphoproliferative disorder marked by the clinical features of lymphadenopathy, hepatosplenomegaly, rash, and hypergammaglobulinemia. Patients can develop a nonerosive, symmetric, seronegative polyarthritis concurrent with other features, or as an initial complaint of the disease. Similar features have been reported with intravascular lymphoma, with a report of a patient presenting with a symmetric polyarthritis accompanied by fever. Table 122-6 lists musculoskeletal complaints reported with hematologic malignancy.

### Graft-versus-Host Disease

Graft-versus-host disease is a complication of bone marrow transplantation and a major cause of morbidity and mortality in the transplant population. Numerous musculoskeletal complaints arise in the setting of acute graft-versus-host disease (lasting 0 to 3 months) and of chronic graft-versus-host disease (lasting >3 months after transplantation). The most frequent manifestation is involvement of the skin, which in many cases can progress to resemble the changes of systemic sclerosis. Skin changes consistent with eosinophilic fasciitis also have been reported. Graft-versus-host disease can lead to symptoms of keratoconjunctivitis sicca and xerostomia resembling Sjögren’s syndrome. Other features, including arthralgias, arthritis, myositis, Raynaud’s phenomenon, and serositis, have also been reported.

### SUMMARY

A plethora of factors contribute to the development of musculoskeletal syndromes in the setting of autoimmune disease and malignancy. A great many autoimmune disorders are known, and for most, the underlying cause and pathogenesis have not been elucidated. This incredible diversity often makes understanding the relationships between associated symptoms difficult. To allow any generalizations regarding an association between an autoimmune disorder and the subsequent development of malignancy, large numbers of patients must be studied longitudinally for exceptionally long periods. Other confounders complicate the picture. Many of the agents used in the treatment of connective tissue and autoimmune disorders modulate the immune system. These agents may have direct carcinogenic potential, whereas others may affect the immune system in a way that may decrease tumor surveillance, subsequently leading to the development of a neoplasm. Intricately entwined are the unique differences among individual immune systems, not only in healthy individuals, but also in individuals whose immune systems are already altered because of an underlying autoimmune disorder. Although uncommon, it is plausible that virtually any of the autoimmune-based diseases and the agents used to treat them might be associated with malignancy in certain circumstances. Most important, when musculoskeletal symptoms arise, malignancy or paraneoplastic syndromes should be considered in the differential diagnosis, especially when patients present with atypical features of autoimmune disease or are refractory to conventional treatment. In addition, the potential for any agent to induce a neoplastic process must be weighed against its proposed benefits before it can be initiated as therapy.

### Table 122-6 Musculoskeletal Manifestations of Hematologic Malignancy

<table>
<thead>
<tr>
<th>Malignancy</th>
<th>Pathogenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leukemia</td>
<td>Infiltration of synovium</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>Metastases or invasion of bone, rarely joint</td>
</tr>
<tr>
<td>Angioimmunoblastic lymphadenopathy</td>
<td>Vasculitis, cryoglobulinemia</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>Metastasis or invasion of bone, hyperuricemia</td>
</tr>
</tbody>
</table>

### Selected References


Full references for this chapter can be found on www.expertconsult.com.
References


Joints and periarticular structures are often involved by non-neoplastic, mass-forming lesions, such as synovial cysts and loose bodies. These structures are affected infrequently, however, by benign or malignant neoplasms. Joint neoplasms can be divided into tumors that are primary or arise de novo within the joint and tumors that are secondary and access the joint by invading from neighboring bones and soft tissues, or by spreading from distant sites via the vascular system. Primary joint neoplasms are more common and tend to recapitulate the phenotype of tissues that normally construct the joint and synovial sheath—synovium, fat, blood vessels, fibrous tissue, and cartilage. Regardless of the histologic type, the benign variants greatly outnumber their malignant counterparts, and as a group these tumors tend to develop in the synovium, not in the other periarticular structures. These biologically and morphologically diverse lesions often pose significant challenges in diagnosis and treatment, and their clinicopathologic features are the focus of this chapter.

### NON-NEOPLASTIC LESIONS

**Synovial and Ganglion Cysts**

Cysts are defined as closed compartments or sacs that are lined by epithelium and are frequently filled with fluid. Neither the synovial cyst nor the ganglion cyst is considered a true cyst because each lacks an epithelial lining.

Synovial cysts are common and are formed from the synovial lining of a joint, tendon, or bursa. They are non-neoplastic lesions and are caused by herniation of the synovium through the joint capsule or tendon sheath into neighboring tissues, or by expansion of a pre-existing bursa. In adults, synovial cysts frequently develop in association with a variety of joint disorders, including trauma, osteoarthritis, crystal arthropathies, infection, and rheumatoid arthritis or one of its variants. Most synovial cysts have an anatomic relationship to a joint, and most originate in the posterior aspect of the knee, where they are known as a popliteal or Baker's cyst, followed in frequency by the shoulder and hip. These lesions may also arise in the spine, where they develop from the facet joints, most commonly in the lower lumbar region. The posteromedial region of the knee may be prone to the development of synovial cysts because the synovium-lined joint capsule in this anatomic site may not provide adequate structural support. Synovial cysts of the posterior knee joint are purported to affect 2.4% of children, who, in contrast to adults, are usually asymptomatic and have an otherwise normal knee joint.

Synovial cysts can enlarge as they become increasingly distended with synovial fluid. Consequently, they may manifest as a periarticular mass, produce progressive joint pain and swelling, limit joint mobility, and compress adjacent neurovascular structures. An example of the last-mentioned item occurs in the spine, where synovial cysts that arise from facet joints may impinge or extend into spinal nerves, causing radicular pain. Other complications of synovial cysts, which sometimes can produce dramatic clinical findings, are acute rupture and secondary infection.

A variety of radiographic techniques have been used to image synovial cysts. Imaging modalities that provide the greatest quantity of diagnostic information include arthrography, ultrasonography, computed tomography (CT), and magnetic resonance imaging (MRI). All of these modalities reveal synovial cysts to be simple or septated thin-walled structures associated with joints and periarticular...
structures and filled with fluid, whose density is similar to that of water (Figure 123-1).

Grossly, synovial cysts usually range in size from 1 to 10 cm. Their inner surface is smooth, glistening, and translucent; however, prior hemorrhage or secondary infection may distort this surface by virtue of attached blood clot and inflammatory debris or the generation of granulation tissue. The cyst wall comprises an inner surface lined by flattened or plump cuboidal synoviocytes arranged one or several cell layers thick, which are surrounded by an outer sheath of fibrous tissue (Figure 123-2). Sometimes the synovial lining cells may be hyperplastic and form papillary fronds, and occasionally scattered subsynovial collections of hemosiderin-laden macrophages are indicative of previous hemorrhage. Facet joint cysts often contain abundant amorphous debris surrounded by macrophages and are associated with severe changes of the ligamentum flavum.

Treatment of synovial cysts varies and depends on their location and associated symptoms. These cysts may be managed successfully with conservative therapy; however, in certain situations, surgical excision is required.1-7

Ganglion cysts have been recognized for centuries; Hippocrates described them as being composed of “mucoid flesh.” They are more common than synovial cysts and arise from tendon sheaths, ligaments, menisci, joint capsules, and bursae. Occasionally, they develop de novo in the subchondral areas of bone, and rarely, they arise within nerve or skeletal muscle and lack communication with a joint. Intraneural ganglia have been shown to develop from dissection of joint fluid along an articular branch of a nerve.9 Ganglion cysts are distinguished from synovial cysts by virtue of the fact that they lack a surface lining of synoviocytes. A variety of hypotheses have been proposed to explain their pathogenesis, but none have been proven.8 The most accepted theory is that ganglia develop from mucoid cystic degeneration of periarticular structures. They are commonly associated with repetitive motion activities, inflammatory arthropathies, and trauma.

Most ganglia arise along the dorsal and volar aspects of the wrists and fingers and the dorsum of the feet.8,10 They usually are asymptomatic and typically manifest as a slowly growing, mobile, firm mass that moves with the structure from which it has arisen (Figure 123-3). Ganglia may be painful if traumatized and can compress adjacent neurovascular structures, producing a variety of symptoms. The radiographic characteristics of ganglia are similar to those of synovial cysts, and they appear on images as small, fluid-filled cystic structures.1,10

Macroscopically, most ganglia are round, but they may form elongate cylindrical structures if they track along a tendon sheath. Ganglia are uniloculated or multiloculated, have thin walls, and are filled with translucent mucoid fluid (Figure 123-4). The cyst lacks an inner cell lining, and the bulk of the wall consists of dense fibrous tissue, which is usually surrounded by areolar tissue (Figure 123-5). In many instances, the cyst wall is distorted by variable quantities of reactive myxoid tissue and muciphages, which result from small ruptures and extravasation of fluid.

When ganglion cysts form, they may remain stable for years or may spontaneously resolve, and ganglia that disappear may subsequently redevelop. Treatment is frequently conservative because of their innocuous nature. Ganglia
Loose bodies and joint mice are generic terms for free-floating structures within a joint cavity. They are the most common tumor-like lesions of joints and may be exogenous, such as fragments of a bullet, or endogenous, such as pieces of articular cartilage, osteophytes, menisci, ligaments, or bone. When not otherwise specified, the term loose bodies refers to detached pieces of articular cartilage or subchondral bone (osteoarticular loose bodies) or both that lie free within the joint, or that have become secondarily embedded in the synovium. Loose bodies can cause pain, crepitance, and locking, and they can limit joint range of motion.

Osteoarticular loose bodies are a secondary complication of a variety of conditions, including trauma, osteochondritis dissecans, and arthritides of various causes. When dislodged, the sloughed articular cartilage remains viable because it receives its nourishment from the synovial fluid, but the bone dies because it derives its nutrition solely from blood vessels. Over time, as the loose body tumbles within the joint, its edges become rounded and smooth; however, it eventually becomes embedded within the synovium. When the synovium encompasses the loose body, it may digest and resorb it, or adjacent subsynovial connective tissue cells may undergo a proliferative and metaplastic response.

These cells produce layers of newly formed fibrocartilage and hyaline cartilage, which may undergo endochondral ossification and are deposited on the surface of the loose body (Figure 123-6). These layers of newly formed tissue surround the centrally located loose body, similar to the cambium layers of a tree, and provide a mechanism for the whole structure to increase gradually in size and become significantly larger than the initial osteochondral defect from which they originated (Figures 123-7 and 123-8). As the loose body enlarges, the innermost portion of original articular cartilage cannot be supported adequately by diffusion of synovial fluid, and it dies and calcifies. This combination of events causes the loose body to appear on x-rays as dense speckled and ring-like calcifications (Figure 123-9). Radiographically and histologically, the differential
anterior portions of the joint and frequently in the posterior portions as well. In humans, nodules develop within the substance of the meniscus of the knee joint adjacent to its attachments to the tibia.

The exact origin of such structures is unknown, although they probably are true sesamoid bones, as seen in rodents, or they may represent ossification secondary to local injury. This latter possibility is supported by the fact that previous knee trauma has been noted in numerous reported cases. The main symptom of meniscal ossicles is pain after exertion, such as walking or prolonged standing, with relief when the knee is at rest. Radiographs may reveal an intraarticular calcification that can be confused with a loose body. MRI shows the ossicle to be a corticated marrow-containing structure that has increased signal intensity on T1-weighted images and decreased signal intensity on T2-weighted images.

The ossicle is located within the lateral or the medial meniscus and appears as a small (≈1 cm in diameter), palpable bony nodule (Figure 123-10). If the ossicle is symptomatic, it should be excised; however, if it is an incidental finding, it can be managed conservatively.

NEOPLASMS

Fatty Lesions of the Synovium

Although the subsynovial connective tissue of diarthrodial joints is rich in fat, a true lipoma of the synovium is rare. When these rare tumors develop, they most frequently affect the knee joint and the synovial sheaths of tendons of the hands, ankles, and feet, where they are more common in the extensor than in the flexor synovial sheaths. Synovial lipomas can be sessile or pedunculated, and when pedunculated, they may produce pain if they twist on their stalks and become secondarily ischemic. Synovial lipoma, similar to its subcutaneous counterpart, comprises lobules of mature white adipocytes that are delineated by a thin fibrous capsule.

A more common but still unusual fatty lesion of the joint is lipoma arborescens, also known as villous lipomatous proliferation and lipomatosis of the synovium. This disorder is
characterized by a diffuse increase in the quantity of subsynovial fat, which bulges into the overlying synovial lining, producing a villous architecture. It is uncertain whether the proliferating fat is neoplastic (lipomatosis) or is a manifestation of a hyperplastic or reactive process. Affected patients are usually adults, but sometimes adolescents and rarely children develop the lesion. Lipoma arborescens causes chronic effusions, pain, and swelling, and restricts joint motion. The duration of symptoms is often long, and symptoms have been reported to be present for as long as 30 years; however, acute onset has been documented.

Lipoma arborescens most commonly arises in the knee (Figure 123-11), especially the suprapatellar portion, although it also has been observed in the hip, ankle, and wrist joints. It is typically localized to one joint, but several cases of bilateral knee involvement have been described. Laboratory studies are unremarkable, and the joint fluid is clear and yellow. Plain films show joint fullness, and findings of osteoarthritis are often present. Arthrography reveals multiple lobulated filling defects, which on CT represent a villonodular mass of low signal intensity that on MRI has the density of fat (Figure 123-12). At surgery, the affected synovium has a prominent villous or villonodular architecture and is tan-yellow (Figure 123-13). Histologically, the lesion comprises sheets of mature adipocytes admixed with nutrient blood vessels, all of which are partially compartmentalized by fibrous septa, and is covered on its intraarticular surface by several layers of synovial cells (Figure 123-14). Synovectomy may relieve the symptoms and prevent effusions, but associated osteoarthritis may be progressive.
The clinicopathologic differential diagnosis includes diffuse tenosynovial giant cell tumor, synovial chondromatosis, and synovial hemangiomia. These lesions can be distinguished easily from lipoarborescens by their distinct histologic features. Another disorder that should be included in the differential diagnosis is Hoffa’s disease—a condition of irritation, inflammation, and hyperplasia of the synovial lining in regions where fat is normally present, such as adjacent to the patella or patellar ligament.

Vascular Lesions of the Synovium

Benign vascular tumors of the synovium are rare. Their growth pattern can be localized or diffuse. They tend to predominate in adolescence and young adulthood, but symptoms frequently can be traced back to childhood. The joint most commonly involved is the knee, but hemangiomas also have been described in the elbow, the ankle, the tarsometatarsal and temporomandibular joints, and the tendon sheaths of the wrist and ankle.

Unusual complications of synovial hemangiomas include a secondary destructive arthritis and Kasabach-Merritt syndrome.

Synovial hemangiomas produce a variety of symptoms, including unilateral, intermittent joint pain and enlargement, which may result in limitation of motion, locking, buckling, and hemarthrosis, especially after minimal trauma. Classically, the affected joint diminishes in size if sufficiently elevated to allow the blood to drain out of the lesion. On physical examination, the joint is swollen and doughy, and nearby cutaneous hemangiomas may be evident. Joint aspiration frequently yields bloody fluid. Preoperative diagnosis of localized hemangioma is difficult, and the differential diagnosis includes localized tenosynovial giant cell tumor, and in the knee includes discoid meniscus, meniscal tears, cysts, and ossicles.

Radiographic evaluation may show nothing more than a vague soft tissue shadow indicative of a swollen synovium and a distended joint capsule or regional osteoporosis in patients who have had long-term symptoms and recurrent hemarthrosis. Rarely, calcified phleboliths are apparent; however, they are associated more often with a soft tissue arteriovenous malformation with secondary joint involvement than with an isolated intra-articular hemangioma. Arthrography may show an intra-articular filling defect, and arteriography may be negative in small localized capillary hemangiomas, but contrast material may collect in the more diffuse lesions that contain cavernous or large ectatic vascular spaces. CT reveals a lobulated soft tissue mass with mild enhancement after contrast injection. MRI may show the tumor to have a low to isointense signal on T1-weighted images and high signal intensity on T2-weighted images.

Macroscopically, the localized hemangioma tends to be small, but larger lesions (8 cm) have been documented. It may be poorly defined, well circumscribed, sessile, or stalked, and ranges in color from red to dark blue-purple. Microscopically, the hemangioma is usually of the cavernous or venous type, with large dilated blood-filled vessels lined by cytologically benign endothelial cells. In arteriovenous hemangioma, which grows in a diffuse fashion, the entire synovium may be edematous and beefy red or stained brown by hemosiderin, and consists of prominent tortuous, congested vessels that may penetrate the joint capsule and extend into neighboring soft tissues. Histologically, the vessels recapitulate architecturally abnormal arteries, veins,
and capillaries; have abnormal interconnections; and are arranged in a disorganized tangle.

Therapy for a localized hemangioma is marginal surgical excision, which usually is curative. Diffuse lesions frequently are difficult to eradicate because of their extensive nature. Incomplete excision or debulking may be the only surgical option. Radiation therapy is not indicated.

**Fibroma of Tendon Sheath**

Fibroma of tendon sheath is an uncommon benign neoplasm that clinically mimics tenosynovial giant cell tumor but is morphologically distinct. Fibroma of tendon sheath first was identified as a clinicopathologic entity in 1936, and since that time, more than several hundred cases have been reported. A translocation t(2;11) (q31-32;q12) that has been identified in this tumor is likely related to its molecular genesis.

Fibroma of tendon sheath usually arises from the tendons and sheaths of the flexor surfaces of the distal extremities; approximately 70% of cases involve the fingers or hand. Of the fingers, the thumb is affected most frequently, followed in descending order by the index and middle fingers. Less commonly, large diarthrodial joints such as the knee and rarely the elbow and ankle are sites of origin. Patients range in age from infants to the elderly, but the median age is the early 4th decade of life. Patients present with a slow-growing, painless mass that usually has been noted for several months to a year. In 6% to 10% of cases, a history of antecedent trauma is reported. Tumors developing in large joints can be painful, can restrict range of motion, and may be palpable.

Plain x-rays show soft tissue fullness; rarely is evidence of bony erosion noted. CT or MRI shows a solid, well-circumscribed mass of soft tissue density that usually has low signal intensity on T1-weighted and variable intensity on T2-weighted images. At surgery, the tumors are often attached directly to the tendon or tendon sheath. They are rubbery, oblong, well circumscribed, or encapsulated; average 1.5 to 1.8 cm in greatest dimension; and have a tan-white cut surface (Figure 123-17).

Microscopically, fibroma of tendon sheath is multilobular, with clefts interposed between adjacent lobules. The lobules are composed of spindle and stellate fibroblasts enmeshed in a collagenous and sometimes myxoid stroma. Immunohistochemically, the tumor cells have the staining profile of myofibroblasts, and ultrastructurally, the cells have features of fibroblasts and myofibroblasts.

The natural history of fibroma of tendon sheath consists of slow growth that eventually ceases. The treatment of choice is surgical excision, but a 24% recurrence rate has been reported.

**Synovial Chondromatosis**

Synovial chondromatosis is an uncommon condition characterized by the formation of multiple nodules of hyaline cartilage within the subsynovial connective tissue. If the cartilage nodules undergo endochondral ossification, the term synovial osteochondromatosis is appropriate. It is unclear whether the proliferating cartilage is metaplastic or neoplastic; however, more recent cytogenetic abnormalities involving chromosome 6 found in the cartilage of these lesions support a neoplastic process. Regardless, synovial chondromatosis is benign and does not metastasize.

Synovial chondromatosis most commonly affects middle-aged men, with an average age in the 5th decade of life. Middle-aged women are more likely to develop the disease in the temporomandibular joint. The genders are equally affected with regard to hand and foot involvement, and patients with hand and foot involvement are usually in their 6th decade.

Patients commonly describe joint pain, swelling, stiffness, crepitance, and limitation of motion with a locking or grating sensation on movement. Symptoms usually are long-standing, recurrent, and progressive.

Synovial chondromatosis typically arises in large diarthrodial joints. The knee is affected in more than 50% of cases, usually as a monoarticular condition. Other common sites include the hip, elbow, shoulder, and ankle. Infrequently, synovial chondromatosis arises in the small joints.
of the hands and feet and in the temporomandibular joint. When cartilage nodules develop in the synovial lining of bursae, tendons, and ligaments, this is known as extra-articular synovial chondromatosis. The extra-articular variant most commonly affects the fingers, followed by the toes, hand, wrist, foot, and ankle, and more than one synovial sheath may be involved.

Plain x-ray findings largely depend on whether the cartilage nodules are calcified or ossified, and whether they erode adjacent bony structures. Visible calcifications are absent in 5% to 33% of cases; however, most often, multiple oval intra-articular radiodensities range in size from a few millimeters to several centimeters (Figure 123-19). The pattern of mineralization varies and may appear as irregular flecks that represent calcified cartilage or show a trabecular architecture, which is a manifestation of endochondral ossification. Lesions that are not mineralized can be seen on an arthrogram because they produce multiple filling defects. In approximately 11% of cases, nodules erode the neighboring skeleton, especially along the anterior aspect of the distal femur.

CT may show mass-like nodules in the synovium that have a density similar to skeletal muscle. CT also can detect small calcifications and erosions before they are apparent on plain films. MRI shows that the nodules of cartilage have low signal intensity on T1-weighted sequences and high intensity on T2-weighted sequences; this reflects the high water content of the hyaline cartilage. Areas of calcification or mineralized bone have a low signal intensity on T1-weighted and T2-weighted sequences. CT and MRI scans are helpful in identifying the intra-articular source of the lesion and its anatomic extent. In long-standing disease, the bones adjacent to involved joints may be osteoporotic and may show changes of secondary osteoarthritis.

The cartilage in extra-articular synovial chondromatosis has similar radiographic changes. Nodules of cartilage are frequently mineralized and may appear as a linear arrangement of small calcific densities that are aligned along the sheath and that can span many joints (Figure 123-20).

The radiographic differential diagnosis of synovial chondromatosis includes osteochondritis dissecans, osteoarthritis with loose bodies, tuberculosis, hemopathic arthropathies, pseudogout with extensive synovial calcification, and synovial tumors. In many instances, the clinical presentation and the radiographic picture should lead to the correct diagnosis. However, in many cases, x-rays and the clinical picture are vague, so that only a biopsy specimen can remove all doubt about the diagnosis.

Characteristic of synovial chondromatosis is a thickened synovium containing numerous opalescent firm nodules of cartilage that bulge from the surface in a cobblestone pattern (Figure 123-21). These nodules usually measure less than 5 cm and may lose their attachment to the synovium and form loose bodies, sometimes hundreds of them. The calcified cartilage is white, and areas of ossification manifest as gritty tan trabeculae, which may house fatty marrow. The synovium adjacent to the cartilage may show reactive changes, such as edema, hyperemia, hyperplasia, and villous transformation.
Cartilage develops in the connective tissue of the subsynovial compartment (Figure 123-22). Neoplastic chondrocytes produce the hyaline matrix that eventually forms individual nodules, which peripherally abut surrounding tissues (Figure 123-23). Cartilage varies in its cellularity, and chondrocytes range from small to large; some may be binucleate and hyperchromatic, similar to the chondrocytes in intrasosseous chondrosarcoma (Figure 123-24). Despite these histologic findings of concern, experience has shown that hypercellular variants with atypical chondrocytes usually behave in a benign fashion. Infrequently, the disease manifests as a single, solitary nodule of cartilage, which may be very large and may undergo partial endochondral ossification. Intra-articular osteochondroma can severely limit joint motion and may be confused clinically with other types of neoplasms.38

Over time, nodules of cartilage attached to the synovium are invaded by blood vessels. This invasion results in endochondral ossification, with woven and lamellar bone formation and the development of a medullary cavity with fatty marrow (Figure 123-25). If these nodules lose their synovial attachments and become free-floating, they may continue to increase in size, because the cartilage derives its nourishment from synovial fluid, although the osseous portion and the marrow die.

The treatment of choice for synovial chondromatosis consists of excision of the involved synovium and removal of all loose bodies. The prognosis is good, although recurrences may be noted if removal is incomplete. Most recurrences develop in the setting of diffuse involvement of the synovium. Synovial chondromatosis rarely undergoes malignant transformation into chondrosarcoma, although in one series, this phenomenon occurred in 5% of cases. However, a significant percentage of the few reported cases of synovial chondrosarcomas have shown evidence of underlying synovial chondromatosis.39-41 The chondrosarcomatous component exhibits dense hypercellularity, marked cytologic atypia, and mitotic activity; the stroma is often myxoid.
**Chondroma of Tendon Sheath and Periarticular Structures**

A solitary soft tissue chondroma is considered to be a benign neoplasm. It commonly arises in tendon sheaths and infrequently involves joint capsules or other periarticular structures.

Tendon sheath chondromas usually arise in the flexor tendon sheaths of the distal extremities and are about three times more common in the hands than in the feet. They affect the sexes equally and are detected in early to mid-adulthood, presenting as a painless, slowly growing, firm mass. Radiographically, tendon sheath chondromas appear as an extraosseous, well-delineated soft tissue mass that contains calcifications, which are punctate or ring-like in 33% to 70% of cases. Grossly, the tumors are ovoid, firm, blue-white, well-circumscribed masses of hyaline cartilage that are usually 1 to 2 cm in dimension and, in contrast to synovial chondromatosis, are solitary. Histologically, the hyaline cartilage is well formed with occasional small foci of myxoid change. The cartilage can be cellular, and chondrocytes may show limited cytologic atypia; this sometimes causes confusion with chondrosarcoma. The treatment of choice is simple excision, which is infrequently associated with local recurrence.

The intracapsular and periarticular regions are uncommon sites for soft tissue chondromas. When they occur, they usually originate in the anterior infrapatellar region of the knee (Figure 123-26). In this location, chondromas can achieve a large size (8 cm) and may interfere with knee motion. Their morphology and biologic behavior are similar to those of soft tissue chondromas that arise elsewhere. Intracapsular chondromas have been reported in the knees of three members of a family with familial dysplasia epiphysealis hemimelica. Two other cases have been described in which cartilaginous hamartomas of the volar plates of the proximal and distal interphalangeal joints of the hands and feet were associated with peculiar hypertrophic skin lesions of the hand and hemihypertrophy of the limb.

**Tenosynovial Giant Cell Tumor**

Tenosynovial giant cell tumor comprises a group of benign tumors that affect the synovial lining of joints, tendon sheaths, and bursae. These lesions were previously known as giant cell tumor of tendon sheath, localized nodular synovitis, and pigmented villonodular synovitis. Tenosynovial giant cell tumor may be localized or diffuse and may be locally aggressive in that it may invade into bone, grow through joint capsules, extend along tendons, and infiltrate into adjacent soft tissues. Despite its destructive potential, this lesion does not have the capacity to metastasize.

The common histologic denominator of these lesions is the neoplastic proliferation of synovial-like cells that may form a localized mass or spread along the synovial surface and invade downward into subsynovial connective tissue. The growing cells expand the subsynovial compartment, producing finger-like extensions, villi, and redundant folds. These projections often fuse into nodules and form convoluted lobulated masses admixed with a tangle of hair-like villi. The process may be a local phenomenon involving only part of the synovial lining, or it may be extensive, with the whole synovial surface affected.

Until recently, the origin of tenosynovial giant cell tumor was unknown. Previously considered a reactive process, possibly in response to repeated hemorrhage, many tenosynovial giant cell tumors now have been shown to result from a translocation between chromosomes 1p13 and 2q35 in which the gene encoding colony-stimulating factor-1 is fused to the collagen VI alpha-3 (COL6A3) gene. Consequently, overexpression of colony-stimulating factor-1 occurs in the neoplastic cells, which account for only 2% to 16% of cells in the mass. The remaining cells largely represent non-neoplastic inflammatory cells that are recruited into the tumor because they contain the receptor for colony-stimulating factor. This phenomenon has been termed a landscape effect and is also observed in certain types of lymphomas and sarcomas.

**Tenosynovial Giant Cell Tumor of Joints and Tendon Sheaths: Diffuse Type (Synonym: Pigmented Villonodular Synovitis)**

The diffuse type of tenosynovial giant cell tumor of the joint involves large areas of the synovial lining, although involved areas are invariably present. Its incidence is approximately 1.8 per 1 million. Although it may occur in all age groups, spanning children to the elderly, most affect young adults in the 3rd to 4th decade of life. The sexes tend to be equally affected, although some series have reported a predominance of males or of females.

Diffuse tenosynovial giant cell tumor of the joint usually manifests as monoarticular arthritis. Bilaterality or involvement of multiple separate sites has been infrequently reported. Some patients with polyarticular disease also have had significant congenital anomalies. Primary complaints include pain and mild intermittent or repeated bouts of
swelling. Symptoms develop insidiously and progress slowly over a long time, ranging from months to years.\textsuperscript{32,34} The involved area may be stiff, swollen, and warm, and a palpable mass sometimes can be appreciated. Point tenderness can be detected in approximately 50% of patients. Anatomic instability of the involved joint is uncommon.

The knee joint is affected most commonly and is involved in about 80% of cases.\textsuperscript{32,34} The next most frequent sites are the hip, ankle, calcaneocuboid joints, elbow, and tendon sheaths of fingers and sometimes toes. Occasionally, the palm, the sole of the foot, and unusual locations such as the temporomandibular joint and posterior elements of the spine are involved. Bursal involvement is rare, but if it happens, it usually occurs in the popliteal and iliopsoas bursae and the bursa anserina. Infrequently, the disease affects large tendon sheaths proximal to the ankle and wrist, and produces a periarticular soft tissue mass.\textsuperscript{35,36} It is thought that some of these lesions dissect through a joint capsule or a tendon sheath and extend along fascial planes to produce a soft tissue mass.\textsuperscript{35}

Invasion of bone on either side of a joint can be seen with intra-articular, bursal, or tendon sheath involvement. This most frequently occurs when the tumor involves “tight” joints, such as the hip, elbow, wrist, and feet, or when tendon sheaths are closely apposed to neighboring bones (Figure 123-27).\textsuperscript{37,38} Rarely, only one bone may be invaded by an intra-articular lesion, and in this situation, it may be difficult to distinguish this type of lesion from a primary bone tumor (Figure 123-28).\textsuperscript{39}

Joint aspiration frequently yields blood-tinged brown fluid that lacks diagnostic abnormalities.\textsuperscript{60} Synovial fluid analysis may show a low glucose content, a minimally elevated protein level, and a faint mucin clot. The inflammatory cell count is usually low but may be elevated. Similar findings can be seen in trauma, Charcot’s joint, bleeding disorders, sickle cell disease, and Ehlers-Danlos syndrome.

In at least two-thirds of cases, a soft tissue density due to the tumor or effusion or both can be visualized on a plain film.\textsuperscript{61-64} Joint narrowing or calcification is uncommon. Arthrography may show numerous nodular filling defects that extend into an expanded joint space. Arteriograms are unusually striking owing to the prominent vascularity of the tumor. There tends to be an inverse correlation between the degree of vascularity and the amount of fibrosis or scarring of the lesion.

CT and MRI are useful in delineating the extent of disease and can detect intralesional lipid and hemosiderin deposits, which are important diagnostic features.\textsuperscript{61-63} The tumor has low signal intensity (equal to that of skeletal muscle) on T1-weighted images and is heterogeneous on T2-weighted images. Extension into the bone manifests radiographically as multiple, well-marginated, subchondral cyst–like lucencies or juxtacortical oval pressure erosions (see Figures 123-27 and 123-28).\textsuperscript{58,59} In the knee, the femoral area adjacent to the intercondylar region is the site most frequently invaded as the tumor grows along the cruciate ligamentous insertions. Periarticular osteopenia, periosseal reactions, and joint destruction are unusual because the joint space is preserved until late in the course of the disease.\textsuperscript{52,63} The radiographic differential diagnosis of a given case includes (1) tuberculosis, which generally involves more osteopenia and joint destruction; (2) hemophilia, which is also associated with more extensive joint destruction; (3) synovial chondromatosis, which frequently has calcified radiopaque bodies; and (4) rheumatoid arthritis, which shows more severe osteopenia and joint narrowing.

Grossly, the synovium in diffuse tenosynovial giant cell tumor is red-brown to mottled orange-yellow and looks like a plush Angora rug (Figure 123-29). Matted masses of villous projections and synovial folds are prominent and are admixed with sessile or pedunculated, rubbery-to-soft
nodules (0.5 to 2 cm in diameter). The synovial membrane is thick and succulent, and is often coated with a fibrinous exudate. Red-brown or golden brown tissue may extend deep into subsynovial structures or may invade the joint capsule. If a tendon sheath is involved, a sausage-shaped mass may be evident because the sheath is distended by the proliferating tumor. If the joint capsule is invaded, adjacent soft tissue structures, including nerves and vessels, may be covered by wispy, red-brown tissue. If soft tissue invasion is extensive, the lesion may appear as a soft-to-rubbery, red-brown mass with foci of hemorrhagic cysts. Similar tissue may be present near the chondro-osseous junction or may be wrapped around vascular and ligamentous attachments to bone surfaces, which represent entrance points into the interior of the bone. Although other conditions, such as hemochromatosis and hemosiderosis, also may discolor the synovium brown, the nodular component is usually absent; in addition, microscopic features are definitive for separating these entities (see later).

Microscopic examination reveals marked synovial cell hyperplasia with surface proliferation and, more important, subsynovial invasion by masses of mitotically active polygonal or round cells with moderate amounts of eosinophilic cytoplasm and round nuclei (Figures 123-30 and 123-31). Included among the invading synovial cells are scattered lymphocytes, multinucleated giant cells (osteoclast, Touton, or foreign body type), hemosiderin-laden macrophages, and fibroblasts. Hemosiderin also can be seen between cells and in synovial lining cells and polygonal cells. Foci of hemorrhage are common and are surrounded peripherally by giant cells and macrophages (Figure 123-32). Scattered collections of foamy macrophages (xanthoma cells) filled with lipid also are a frequent finding. These different cell populations fill and distend the synovial villi, causing them to fuse with adjacent ones, forming nodules. In some nodules, abundant collagen deposition with hyalinization may cause confusion with neoplastic bone. Rarely, the tumor may contain focally calcified cartilaginous matrix.65
Immunohistochemical studies have been interpreted to support a synovial cell or fibrohistiocytic phenotype. The tumor cells express clusterin, D2-40, and a minority of large cells stain with antibodies to desmin. It is important to note that cells that harbor the translocation express colony-stimulating factor-1 (CSF1). Flow cytometric analyses have shown that some of these tumors, especially tumors that have large extra-articular soft tissue components, may be aneuploid and have high proliferative indices. Although these flow cytometric findings may help predict which cases would be more locally aggressive, examples of diffuse tenosynovial cell tumors with these attributes have not metastasized.

Treatment of diffuse tenosynovial giant cell tumor of the joint is not standardized and has included radiation therapy, total synovectomy, arthrodesis, bone grafting, and primary arthroplasty. Although no single therapy has been consistently successful, wide synovectomy is currently the recommended treatment. It is difficult to perform an actual complete synovectomy, however; residual involved synovium frequently remains, causing a local recurrence rate of 16% to 48%. Tumors arising in the knee have a higher rate of recurrence compared with tumors arising in other joints. Rarely, recurrent disease or tumors with large extra-articular components may require more radical surgery, such as ray resection or amputation. Studies have shown that moderate doses of radiation may control and even cure patients with such extensive disease, possibly obviating the need for radical surgery or amputation. Finally, drugs targeting CSF1 are being tested to determine their efficacy in clinically challenging cases.

**Malignant Diffuse Tenosynovial Giant Cell Tumor**

Malignant tenosynovial giant cell tumor is a rare lesion; only a handful of cases have been reported. The knee is the joint that has been most commonly affected, and in many cases, coexisting benign-appearing, diffuse tenosynovial giant cell tumor is noted. In the malignant variant, the neoplastic cells may be spindle or polyhedral shaped and cytologically malignant. These tumors have the capacity to behave aggressively; almost 50% of patients die from metastatic disease.

**Localized Tenosynovial Giant Cell Tumor of the Joint (Synonyms: Benign Giant Cell Synovioma, Benign Synovioma, Localized Nodular Synovitis)**

Localized tenosynovial giant cell tumor of the joint manifests as a solitary, well-circumscribed mass. It usually consists of a single sessile or pedunculated, sometimes lobulated, tumor that ranges from 1 to 8 cm in diameter (Figure 123-33). Most commonly, it is unilateral, arises in the knee, and is equally distributed between the sexes.

Symptoms are similar to those of diffuse tenosynovial giant cell tumor, except that in the localized variant, a higher frequency of joint locking is reported because the mass interferes with motion. A few patients may present with acute severe joint pain caused by torsion and infarction of the tumor. Effusions are common, but the synovial fluid tends to be less bloody than in the diffuse variant, and may be clear.

Imaging studies show a heterogeneous nodular mass that contains lipid and hemosiderin deposits (Figure 123-34). In the knee joint, the tumor frequently arises in the suprapatellar notch, in the femoral notch, and between the meniscus and the joint capsule. Usually, no bone invasion occurs. Marginal excision is usually curative. Small lesions can be extirpated arthroscopically.

Histologically, localized tenosynovial giant cell tumor of the joint is identical to the nodules of the diffuse variant. The main difference is that the prominent synovial villi present in the diffuse variant are absent or sparse.
Localized Tenosynovial Giant Cell Tumor of the Tendon Sheath (Synonyms: Giant Cell Tumor of Tendon Sheath, Fibroxanthoma of Tendon Sheath)

Localized tenosynovial giant cell tumor of the tendon sheath usually involves the hand or wrist, and less frequently the foot or ankle. It is the most common soft tissue tumor of the hand and usually arises from the flexor tendon sheaths of the finger; the index finger is affected most frequently, followed in descending order of frequency by the third, fourth, fifth, and first fingers.

Tumors of the digits predominate in females, with a male-to-female ratio of at least 2:1, and tumors of the toes have an equal sex distribution. On average, patients are in the 3rd to 5th decades of life and present with a painless, palpable, firm, mobile mass.

Clinically, the mass usually is solitary and is located on the flexor surface, but it may bulge into the extensor or lateral aspects of the digits (Figure 123-35). The tumors are slow growing, and the intervals between detection and surgical treatment have ranged from several weeks to longer than a decade, with an average interval of slightly longer than 2 years.

Radiographically, the tumors appear as well-circumscribed soft tissue masses, and in about 25% of cases, adjacent extrinsic excavation of the cortical bone has a sclerotic margin (Figure 123-36). MRI reveals the lesions to have a hypointense signal on T1-weighted images and either a hypointense or a hyperintense signal on T2-weighted images. These findings are helpful in distinguishing giant cell tumor of tendon sheath from other soft tissue tumors.

The gross pathology is that of a well-circumscribed, multinodular, round, rubbery, variegated red-brown-tan-yellow mass, generally not larger than 5 cm in diameter, which is firmly attached but easily peeled off the involved tendon (Figure 123-37). Sometimes at surgery, the lesion may “pop out” of the incision. The cut surface reveals a solid mass that ranges from hues of yellow to orange-red-brown, depending on the quantity of lipid and blood pigments present. Often bands or septa of white fibrous tissue subcompartmentalize the lesion. The morphology and immunophenotype of the cell types present are identical to those in the diffuse variant (Figure 123-38). Ultrastructurally, proliferating cells have features similar to type A and type B synovial lining cells. Flow cytometry has been performed on a few cases, and all have been diploid. Cytogenetics shows that many of these tumors have a translocation between 1p13 and 2q35.

These tumors are benign and do not metastasize. Rarely have malignant giant cell tumors of tendon sheath been reported. The treatment of choice is conservative surgical excision, which is usually curative. Local recurrence may be noted if excision is incomplete.
MALIGNANT TUMORS OF THE JOINT

Malignant tumors of joints are uncommon and are classified into primary and secondary types. Primary malignancies are virtually always sarcomas and usually arise within the synovium of large diarthrodial joints, especially the knee. Patients are adults, who present with chronic symptoms of pain, swelling, and effusion, and most tumors are chondrosarcomas or synovial sarcomas. Rarely, other sarcomas originate within a joint; our experience has included cases of intra-articular myxoinflammatory fibroblastic sarcoma, pleomorphic fibrosarcoma,extraskeletal myxoid chondrosarcoma, conventional chondrosarcoma, malignant tenosynovial giant cell tumor, and angiosarcoma (Figure 123-39). Secondary malignant tumors of joints, by definition, originate beyond the confines of the joint; most are sarcomas that extend from neighboring bone or surrounding soft tissue. Although synovial tissue is highly vascular, metastases or involvement of the synovium by carcinoma, lymphoma, or leukemia is uncommon.

Primary Sarcomas of Joints

Conventional Chondrosarcoma

Conventional chondrosarcoma arising in the synovium is unusual; fewer than 50 cases have been reported in the English language. In approximately 50% of cases, the chondrosarcoma arose in association with pre-existing synovial chondromatosis. Patients are usually in the 5th to 7th decade of life and have an equal sex distribution. Typically, patients present with a progressively enlarging mass in the joint that may cause mechanical dysfunction, pain, and stiffness. In patients who have pre-existing synovial chondromatosis, the duration of symptoms is usually long, and in some instances, symptoms may be present for 25 years. Most chondrosarcomas arise in the knee joint, followed by the hip and elbow joints.

Radiographic studies usually show a periarticular soft tissue mass that may have dense irregular or ring-like calcifications. Occasionally, invasion into the medullary cavity of adjacent bone is evident. The radiographic differential diagnosis varies according to the presence of calcification and includes synovial chondromatosis, synovial sarcoma, diffuse tenosynovial giant cell tumor, and chronic synovitis.

Grossly, the involved joint is filled with synovium massively thickened by innumerable nodules of opalescent blue-white cartilage. The nodules of cartilage vary in size and may be free-floating in the joint cavity. In several cases, the tumor has extended into adjacent soft tissue and bone.

Microscopically, the tumor is composed of malignant hyaline and myxoid cartilage. Rarely, the matrix is entirely myxoid and has features of extraskeletal myxoid chondrosarcoma. The neoplastic cartilage is cellular and contains cytologically atypical chondrocytes. The periphery of the lobules of cartilage is typically the most cellular, and in this region, some of the tumor cells are spindled. Other findings include necrosis and permeation of invaded bone. Coexisting synovial chondromatosis can be identified by its well-formed nodules of hyaline cartilage, which are less cellular, containing cytologically banal-appearing chondrocytes and a matrix that is frequently mineralized.

Treatment is usually surgical extirpation with consideration given for chemotherapy in high-grade lesions or lesions that have metastasized. Inadequate surgical removal virtually ensures local recurrence, which may necessitate subsequent radical excision. Metastases have occurred in approximately one-third of reported patients; the lung is the most common site for systemic spread.

Synovial Sarcoma

Synovial sarcoma is a common sarcoma that accounts for approximately 6% to 10% of soft tissue sarcomas. It usually develops in the deep soft tissues and rarely arises in joints (Figure 123-40), but it may secondarily invade articular synovium from neighboring soft tissues. Earlier descriptions of this tumor attest to its wide spectrum of morphology.
Synovial sarcoma commonly affects adolescents and young adults. In a series of 121 cases, the age range was 9 to 74 years, with a median age of 34 years; however, it occurs in children with significant frequency.\textsuperscript{83,84,86}

The term \textit{synovial sarcoma} implies that the tumor originates from the synovium; however, less than 10% of cases are intra-articular or show continuity with a synovial lining.\textsuperscript{83,86}

Approximately 60% to 70% of synovial sarcomas arise in the extremities, especially the lower limb, in the vicinity of large joints, particularly the popliteal areas of the knee and foot.\textsuperscript{87,88} Regions of the thigh, hand, leg, and digits may be affected, and in the distal extremities, tumors are often adjacent to joint capsules or tendon sheaths or both. Tumors also have been reported in the neck, torso, craniofacial region, retroperitoneum, orbit, tongue, mediastinum, soft palate, heart, kidney, lung, pleura, and prostate.

No clinical features specific to synovial sarcoma distinguish it from other sarcomas. The most common complaint is the development of a slowly enlarging, deep-seated, palpable mass that is painful in about 50% of cases.\textsuperscript{84,88} Symptoms may be present for an unusually long time, ranging from months to 25 years with an average of about 6 months to 2.5 years, before medical evaluation is sought.\textsuperscript{84,88} Delay in diagnosis is more frequent with tumors that are located in the deep soft tissues than in tumors that are based in the more superficial and clinically noticeable regions. In some cases involving the knee region, vague mild pain over several months may occur before a mass is appreciated, and if the tumor reaches a large size, limitation of motion finally may be noted. Head and neck lesions produce symptoms related to their specific sites, such as hoarseness and breathing or swallowing difficulties. Rarely, a patient may present with a symptom secondary to pulmonary metastasis, such as hemoptysis.\textsuperscript{88}

Classically, plain film findings of synovial sarcoma consist of a well-circumscribed, deep-seated soft tissue mass. Synovial sarcoma is one of the few primary soft tissue tumors that frequently calcify. Approximately 30% to 50% of cases reveal radiographically detectable calcifications that may have a fine, stippled, or dense appearance (Figure 123-41).\textsuperscript{89} The calcification may be focal or may present throughout most of the tumor.\textsuperscript{90} Periosteal reaction of adjacent bone is elicited in approximately 20% of cases, but bone is rarely invaded by the tumor.

CT is more sensitive than plain radiography in showing calcification or periosteal reaction. MRI is important in delineating the anatomic extent of the tumor and usually shows a large inhomogeneous mass with areas of hemorrhage. The radiographic differential diagnosis includes hemangiomia, lipoma, synovial chondromatosis, soft tissue chondrosarcoma or osteosarcoma, myositis ossificans, aneurysm, and other sarcomas.

The gross pathology of synovial sarcoma mimics a joint in its early stage of development in that it contains cleft-like spaces and glands delineated by large polygonal (epithelioid) cells surrounded by fascicles of spindle cells. The clefts and glands simulate a microscopic “joint space” that is bounded by synovial lining cells and is supported by subsynovial mesenchymal cells. Because epithelial or spindle cells may predominate, synovial sarcoma has been subtyped into biphasic and monophasic spindle cell and epithelial types, although the latter type is very rare.

Synovial sarcoma commonly affects adolescents and young adults. In a series of 121 cases, the age range was 9 to 74 years, with a median age of 34 years; however, it occurs in children with significant frequency.\textsuperscript{83,84}

The term \textit{synovial sarcoma} implies that the tumor originates from the synovium; however, less than 10% of cases are intra-articular or show continuity with a synovial lining.\textsuperscript{83,86}

Approximately 60% to 70% of synovial sarcomas arise in the extremities, especially the lower limb, in the vicinity of large joints, particularly the popliteal areas of the knee and foot.\textsuperscript{87,88} Regions of the thigh, hand, leg, and digits may be affected, and in the distal extremities, tumors are often adjacent to joint capsules or tendon sheaths or both. Tumors also have been reported in the neck, torso, craniofacial region, retroperitoneum, orbit, tongue, mediastinum, soft palate, heart, kidney, lung, pleura, and prostate.

No clinical features specific to synovial sarcoma distinguish it from other sarcomas. The most common complaint is the development of a slowly enlarging, deep-seated, palpable mass that is painful in about 50% of cases.\textsuperscript{84,88} Symptoms may be present for an unusually long time, ranging from months to 25 years with an average of about 6 months to 2.5 years, before medical evaluation is sought.\textsuperscript{84,88} Delay in diagnosis is more frequent with tumors that are located in the deep soft tissues than in tumors that are based in the more superficial and clinically noticeable regions. In some cases involving the knee region, vague mild pain over several months may occur before a mass is appreciated, and if the tumor reaches a large size, limitation of motion finally may be noted. Head and neck lesions produce symptoms related to their specific sites, such as hoarseness and breathing or swallowing difficulties. Rarely, a patient may present with a symptom secondary to pulmonary metastasis, such as hemoptysis.\textsuperscript{88}

Classically, plain film findings of synovial sarcoma consist of a well-circumscribed, deep-seated soft tissue mass. Synovial sarcoma is one of the few primary soft tissue tumors that frequently calcify. Approximately 30% to 50% of cases reveal radiographically detectable calcifications that may have a fine, stippled, or dense appearance (Figure 123-41).\textsuperscript{89} The calcification may be focal or may present throughout most of the tumor.\textsuperscript{90} Periosteal reaction of adjacent bone is elicited in approximately 20% of cases, but bone is rarely invaded by the tumor.

CT is more sensitive than plain radiography in showing calcification or periosteal reaction. MRI is important in delineating the anatomic extent of the tumor and usually shows a large inhomogeneous mass with areas of hemorrhage. The radiographic differential diagnosis includes hemangiomia, lipoma, synovial chondromatosis, soft tissue chondrosarcoma or osteosarcoma, myositis ossificans, aneurysm, and other sarcomas.

The gross pathology of synovial sarcoma mimics a joint in its early stage of development in that it contains cleft-like spaces and glands delineated by large polygonal (epithelioid) cells surrounded by fascicles of spindle cells. The clefts and glands simulate a microscopic “joint space” that is bounded by synovial lining cells and is supported by subsynovial mesenchymal cells. Because epithelial or spindle cells may predominate, synovial sarcoma has been subtyped into biphasic and monophasic spindle cell and epithelial types, although the latter type is very rare.

Synovial sarcoma commonly affects adolescents and young adults. In a series of 121 cases, the age range was 9 to 74 years, with a median age of 34 years; however, it occurs in children with significant frequency.\textsuperscript{83,84,86}

The term \textit{synovial sarcoma} implies that the tumor originates from the synovium; however, less than 10% of cases are intra-articular or show continuity with a synovial lining.\textsuperscript{83,86}

Approximately 60% to 70% of synovial sarcomas arise in the extremities, especially the lower limb, in the vicinity of large joints, particularly the popliteal areas of the knee and foot.\textsuperscript{87,88} Regions of the thigh, hand, leg, and digits may be affected, and in the distal extremities, tumors are often adjacent to joint capsules or tendon sheaths or both. Tumors also have been reported in the neck, torso, craniofacial region, retroperitoneum, orbit, tongue, mediastinum, soft palate, heart, kidney, lung, pleura, and prostate.

No clinical features specific to synovial sarcoma distinguish it from other sarcomas. The most common complaint is the development of a slowly enlarging, deep-seated, palpable mass that is painful in about 50% of cases.\textsuperscript{84,88} Symptoms may be present for an unusually long time, ranging from months to 25 years with an average of about 6 months to 2.5 years, before medical evaluation is sought.\textsuperscript{84,88} Delay in diagnosis is more frequent with tumors that are located in the deep soft tissues than in tumors that are based in the more superficial and clinically noticeable regions. In some cases involving the knee region, vague mild pain over several months may occur before a mass is appreciated, and if the tumor reaches a large size, limitation of motion finally may be noted. Head and neck lesions produce symptoms related to their specific sites, such as hoarseness and breathing or swallowing difficulties. Rarely, a patient may present with a symptom secondary to pulmonary metastasis, such as hemoptysis.\textsuperscript{88}

Classically, plain film findings of synovial sarcoma consist of a well-circumscribed, deep-seated soft tissue mass. Synovial sarcoma is one of the few primary soft tissue tumors that frequently calcify. Approximately 30% to 50% of cases reveal radiographically detectable calcifications that may have a fine, stippled, or dense appearance (Figure 123-41).\textsuperscript{89} The calcification may be focal or may present throughout most of the tumor.\textsuperscript{90} Periosteal reaction of adjacent bone is elicited in approximately 20% of cases, but bone is rarely invaded by the tumor.

CT is more sensitive than plain radiography in showing calcification or periosteal reaction. MRI is important in delineating the anatomic extent of the tumor and usually shows a large inhomogeneous mass with areas of hemorrhage. The radiographic differential diagnosis includes hemangiomia, lipoma, synovial chondromatosis, soft tissue chondrosarcoma or osteosarcoma, myositis ossificans, aneurysm, and other sarcomas.

The gross pathology of synovial sarcoma reveals a well-demarcated, pink-tan, fleshy mass that easily detaches or “shells out” from its tumor bed (Figure 123-40B). The cut surface is usually uniform, gray-yellow, and rubbery. Calcified areas are gritty and hard. In larger tumors, areas of hemorrhage or necrosis or both with cystification and gelatinous breakdown of tissue may be seen. Synovial sarcoma sometimes grows between tendons, muscles, and fascial planes, or wraps around neurovascular bundles.

Synovial sarcoma is subtyped into three patterns on the basis of predominant microscopic findings: monophasic spindle cell, monophasic epithelial, and biphasic variants. Use of such a classification system requires some subjectivity because many of these tumors have a variable histologic picture. A useful differential observation is that marked
Immunohistochemistry has shown that epithelial and spindle cell components frequently stain with antibodies to keratin and epithelial membrane antigen, which usually are associated with epithelial neoplasms.\textsuperscript{91-93} This pattern of reactivity has helped make it possible to separate synovial sarcoma from morphologically similar tumors, such as fibrosarcoma and malignant peripheral nerve sheath tumor.\textsuperscript{93} It has also provided evidence that synovial sarcoma does not arise from or recapitulate the synovium, because normal synovial cells do not stain with these antibodies.\textsuperscript{94}

Cytogenetic studies of synovial sarcomas have detected a consistent translocation t(X;18)(p11.2;q11.2) in almost all cases, regardless of whether the tumor is biphasic or monophasic.\textsuperscript{87,95} This finding provides insight into the genesis of synovial sarcoma, which may be related to dysregulation of transcription and can be used as a diagnostic feature. No relationship between the location of the translocation breakpoint and the prognosis has been discerned.

The prognosis of synovial sarcoma is poor. In one study of 150 patients with nonmetastatic disease, 5-year, 10-year,
and 15-year disease-free survival rates were 59%, 52%, and 52%.\textsuperscript{86} Many factors influence prognosis; tumors that are small (<5 cm), that arise in patients younger than 25 years of age, and that lack poorly differentiated areas have a high rate of cure.\textsuperscript{84} In contrast, large tumors (≥ 5 cm) that arise in patients 25 years old or older and that contain poorly differentiated areas have a dismal outcome.\textsuperscript{84} The impact of histologic subtype has been controversial. Radiographically, the tumors appear as well-circumscribed soft tissue masses. However, in about 25% of cases they produce extrinsic erosion of the adjacent cortical bone, and when this occurs the margin of the excavation is sclerotic (see Figure 123-36).\textsuperscript{31,75}

The natural history of synovial sarcoma consists of local recurrence, which may be repetitive. Most recurrences manifest within 2 years after initial treatment, but intervals longer than 10 years are not exceptional. Ultimately, metastases occur in many patients, and the most common site is the lungs; regional lymph node involvement has been reported in a few patients.\textsuperscript{96} About 10% of patients die within 1 year after diagnosis with metastatic disease; 90% of these patients have massive pulmonary metastases.

Treatment must contend with issues involved with local and systemic therapy. Successful local control usually can be achieved by limb salvage surgery combined with radiation.\textsuperscript{96,97} Because regional lymph nodes may be involved, their status should be evaluated carefully, and they should be treated if enlarged. Systemic treatment consists of various chemotherapy regimens, which have provided questionable benefit, although adjuvant chemotherapy usually is recommended for patients who are at high risk—patients with tumors larger than 5 cm.\textsuperscript{98-101}

**Secondary Malignant Tumors of the Joint**

**Sarcomas**

Primary sarcomas of bone, such as osteosarcoma and chondrosarcoma, infrequently involve a joint because intact articular cartilage usually acts as a barrier to direct tumor extension. However, when it is observed, joint invasion usually occurs via a pathway created by a transarticular fracture, via growth along tendoligamentous structures, or through capsule insertion sites. This circumstance may make it difficult to distinguish on histologic grounds alone some forms of synovial chondromatosis from a low-grade intraosseous chondrosarcoma that has secondarily spread into the joint.

Similarly, primary soft tissue sarcomas gain access into the interior of a joint by growing through the joint capsule in conduits occupied by pre-existing vascular structures or along tendons and ligaments. This complication can make providing adequate therapy challenging because treatment may require en bloc resection of the joint.

**Metastatic Carcinoma**

The synovium, in contrast to other richly vascular tissues, is rarely the site of metastatic carcinoma; this may reflect the fact that only clinical cases in which joint symptoms prevail are reported because at autopsy, joints are not routinely examined. Most carcinomas that metastasize to the synovium originate in the lung, followed by the gastrointestinal tract and breast.\textsuperscript{99-101} Affected patients are usually elderly, and the knee is the most frequently involved joint. In many reported cases, the underlying bone also contains metastatic deposits.

**Malignant Lymphoproliferative Disease**

The various types of malignant lymphoproliferative diseases, including leukemia, lymphoma, and myeloma, can involve the synovium and may produce osteoarticular symptoms.\textsuperscript{102-104} This complication occurs most frequently in leukemia and is seen in both acute and chronic forms.\textsuperscript{104} Joint symptoms have been observed in 12% to 65% of children and in 4% to 13% of adults with leukemia.\textsuperscript{104} Arthritis can develop at any time during the disease course and can be the presenting complaint. Large joints are affected more commonly than small joints, and the arthritis is often pauciarticular, asymmetric, migratory, and severe. Symptoms may result from leukemic infiltration of the synovium or irritation of the neighboring periosteum. When arthritis is the major presenting symptom, it may cause confusion with septic arthritis, rheumatic fever, subacute bacterial endocarditis, or rheumatoid arthritis.

**References**


The references for this chapter can also be found on www.expertconsult.com.
Index
Antinuclear antibodies (ANAs) (Continued)
clinical utility, 592
conditions, 790t
detection
FANA, usage, 792
methods, 790-794
diagnostic characteristics, 791t-792t
diseases, 790t
association, 795-800
enzyme-linked immunosorbent assay, 793
history, 789-790
immunofluorescence, usage, 792-793
impact, 577-578
insufficiency, 790
negative titer, 1373
pattern, 794-795
presence, 790, 795t
list, 795t, 799t-800t
relevance, 790
rheumatologic evaluation, algorithm,
800-801
systemic lupus erythematosus, association,
795-797
testing, clinical utility, 800-801
titers, 795
usage, algorithm, 794f
Antioxidants
impact, 1046
rheumatoid arthritis, risk (relationship),
1049
Antiperinuclear factor (APF), 805
presence, 805
Antiphospholipid antibodies (aPLs), 1938
aPL-related thrombosis/placental injury,
mechanism, 1333f
clinical/obstetric events absence, 548-549
contribution, 1279
presence, 548-549, 1332t, 1938
prevalence, studies, 1938
treatment, recommendations, 1337t
Antiphospholipid antibody-β2GPI dimer
binding, impact, 1333
Antiphospholipid antibody syndrome (APS),
281-282, 336
anticardiolipin ELISA, sensitivity, 1335
antiphospholipid antibodies, relationship,
548-549
antiphospholipid antibody-negative
individuals, clinical events, 1339
antiphospholipid antibody-positive
individuals, ambiguous events, 1338
aPL-mediated thrombosis, contributory
mechanisms, 1333
aPL-positive ELISA tests, prevalence,
1336
asymptomatic antiphospholipid antibodypositive individuals, 1338
catastrophic antiphospholipid syndrome,
1334
cause, 1331-1332
clinical features, 1334
controlled studies, absence, 1337
criteria, 548
definition, 1331
diagnosis/diagnostic tests, 1334-1337
differential diagnosis, 1336-1337
epidemiology, 1331
imaging studies, 1335
laboratory studies, 1334-1335
livedo reticularis, 1334f
long-term outcomes, 1339
manifestations, 1320
obstetric manifestations, 548
open-label phase, 1338
pathogenesis, 1332-1334

Antiphospholipid antibody syndrome (APS)
(Continued)
pathology, 1335-1336
pregnancy, 1319-1320
morbidity, 1334, 1338
prognosis, 1339
renal thrombotic microangiopathy, 1336f
revised Sapporo classification criteria, 1332t
thrombosis, usage, 1337
thrombotic manifestations, 548
treatment, 1337-1339
usage, 1319-1320
vascular occlusion, 1334
Antiphospholipid syndrome-associated
thrombosis, antiplatelet/anticoagulation
therapy (requirement), 1299
Anti-polymyositis-scleroderma (anti-PM-Scl)
antibodies, 798
myositis-SSc (overlap), 798-799
Anti-RA33 antibodies, discovery, 811
Antirheumatic agents, impact, 166-167
Antirheumatic drug therapies, cancer risks
(association), 520-524
Anti-ribonucleic acid polymerases (RNAPs),
797
Antiribosomal P protein (anti-P), 796-797
Antiribosomes, presence, 796-797
Anti-RNA polymerase, 1392
Anti-RNA polymerase III antibodies, diffuse
skin disease (association), 1372
Anti-RNA polymerase III-positive patients,
gastrointestinal disease (absence), 1372
Anti-RNP antibodies, appearance, 1431-1432
Anti-Ro/SSA antibodies, pathogenic role
(absence), 1177
Anti-Ro/SSA autoantibodies, impact, 603-604
Anti-small nuclear ribonucleoproteins
(anti-snRNPs), 796
Anti-Smith (anti-SM), 1274
Antispasmodic agents, usage, 1028-1029
Antispasmodic medications, 1030
efficacy, 1030
Antispastic medications, 1029-1030
Antistreptolysin O (ASO) titers, increase,
1878-1879
Antisynthetatase syndrome, 1415
Anti-TNF monoclonal antibodies, success, 149
Anti-TNF reagents, usage, 397f
Anti-TNF therapy, 227
reports, 823
usage, 1198-1199
Anti-TNF Trial in Rheumatoid Arthritis with
Concomitant Therapy (ATTRACT) trial,
958-959, 1144
Anti-topoisomerase, 1392
Anti-topoisomerase I (Scl-70), 797
antibodies, presence, 1372
positive value, 1384
Anti-tumor necrosis factor (anti-TNF)
agents, 522-523
adverse effects, 986
autoantibodies, development, 965
availability, 958
continuation, 544-545
effectiveness, 1755-1757
impact, 1035
usage, 523
exposure, analysis, 523
therapy, 680
impact, 523
introduction, 1829
treatment, meta-analyses/cohort studies,
522t
Anti-U1-RNP antibodies, levels (elevation),
1443

i7

Anti-U2 snRNP specificities, 798-799
Anti-U3-RNP (fibrillarin) antibodies
diffuse disease, association, 1372
presence, 1387-1388
Anti-U11/U12-RNP antibodies, 798
Aortic aneurysm/rupture, Takayasu’s arteritis
(maternal complication), 547
Aortic disease, 1734
APACHE II scores, 1326
Apatite-associated destructive arthritis,
1586-1588
Aphthae, 1526
Aphthous ulceration, 1604f
Apical ectodermal ridge (AER)
signaling center, 5
specialized epithelial cells, involvement, 5
Apley compression test, 688-689
Apley grind test, 573-574
Apoptosis, 382, 1097-1098
acceleration, 396
assumptions, 1179
Bcl regulation, 388
biochemistry, 385-386
Caenorhabditis elegans paradigm, 383f
defectiveness, 393-394
developments, 382
electron microscopic morphology, 383f
execution, 387-389
gene regulation, 1097-1098
increase, therapeutic interventions,
1097-1098
induction mechanisms, 1097
initiation, 387-389
intracellular inhibitors, 388-389
laboratory detection, 391-393
modulation, 1581
occurrence, 394
osteonecrosis, relationship, 1698
protein-like protein, human inhibitor,
1408-1409
protein modification, 1434
proteins, involvement (modular
components), 385t
regulators, observations, 395
resolution, 161-162
rheumatic disorders, relationship,
393-397
therapeutic manipulation, avenues, 397f
Apoptosome, 387
Apoptotic bodies, components, 1434
Apoptotic cells
clearance, 332
receptor promotion, 140
defective uptake, 394
detection methods, 392f-393f
ingestion, effects, 391
membrane, chemical alterations, 390-391
phagocytosis, phagocytic receptors, 143f
processing, 394
PS exposure, 286-287
recognition receptors, 123
removal, 390
Apoptotic material, uptake, 129
Apoptotic neutrophils, impact, 162
Apoptotic pathways, drugs (impact), 396-397
Appendicular skeleton, development, 1-2
Apprehension test, 690f
APRIL (a proliferation-inducing ligand),
1313
TNF superfamily cytokines, 380
A priori power calculations, 458
A proliferation-inducing ligand (APRIL), 197,
1313
Aprotinin, Kunitz-type inhibitor, 104-105
Aquatic exercise, usage, 536


Arthrocentesis, 778-779
carpal tunnel syndrome, 780-781
clinical evidence, 770-771
complications, 775-776
list, 775t
contraindications, 774-775
list, 775t
corticosteroid administration, 1803
corticosteroid injections, 771-772, 774
de Quervain's tenosynovitis, 780
dorsal wrist tendons, 780
elbows, 779
joints, 779
feet, 783-784
flexor tenosynovitis (trigger fingers), 781
ganglia, 781
glenohumeral joint, 778, 778f
hands, 779-778
hemostat, helpfulness, 776
hip (acetabular) joint, 782
iatrogenic infection, 775
indications, 770-771
list, 771t
interdigital neuroma, 784
interphalangeal joint, 781
intra-articular biologic therapy, 785
ischiofemoral bursitis, 781
joint irritation, 785
knees
lateral approach, 783f
medial approach, 783f
lateral epicondyle, 779
local steroid injection, 773
complications, 775
lumbosacral spine area, 781
medial epicondyle, 779
metacarpophalangeal joint, 781
metatarsophalangeal joints, 784
nonarticular conditions, 773-774
olecranon bursa, 779
pelvic girdle, 781-782
performing, 1776
periarticular knee pain syndromes, 783
postprocedure instructions/care, 777
preparations, 774
procedure, 1772
challenge, 775
radioan joint, 779-780
regional techniques, 777-778
results, 727-728
rotator cuff tendons, 778
shoulder, 778-779
joint, 778
site preparation/technique, 776-777
sterile technique, 776
subacromial bursa, 778
subtalar joint, 784
systemic absorption, occurrence, 775-776
tarsal tunnel syndrome, 784
techniques, 776-777
temporomandibular joint, 777-778
therapeutic injection, 771-774
indications, 771t
thumb carpometacarpal joint, 781
tibiotalar joint, 783
trends, 785
trochanteric pain syndrome (bursitis), 782
injection, 782f
ultrasound-guided arthrocentesis/injection, 785
usage, 782
wrists, 778-781
Autoimmune polyendocrinopathy-candidiasis-ectodermal dysplasia (APECED), 177
T cell-mediated autoimmune destruction, association, 290
Autoimmune polyglandular syndrome and candidiasis (APECED), 127
Autoimmune regulator (AIRE), 127
AIRE-deficient mice, 292
AIRE-mediated transcription, 271-272
gene, discovery, 177
transcription factor, impact, 290
Autoimmune rheumatic diseases
malignancy, 515-520
pathophysiology, 293-294
Autoimmune thyroid disease, 226-227
Autoimmunity, 128-131, 1067-1071
Autoimmune regulator (AIRE), 127
Autoimmune polyendocrinopathy-candidiasis-ectodermal dysplasia (APECED), 1562-1563
16p12 locus, impact, 1563
Autosomal dominant medullary cystic kidney disease (ADMCKD), 1562-1563
Autosomal locus, alleles (distribution), 311f
AUUUA RNA motif, copies, 343
Avascular connective tissue matrices, 1578
Avascular fibrocartilage, 12
Avascular necrosis (AVN) commonness, 1924-1925
diagnosis, 1926
impact, 755
systemic lupus erythematosus (SLE), relationship, 1288
Avian joints, collagen types/keratan sulfate (distribution), 3-4
Avocado and soy unsaponifiables (ASUs), 1750
Axial spondyloarthritis (SpA)
criteria, 1203
classification criteria, 1204t, 1223t
development, 1223
criteria, 1203f
contrast-enhanced Doppler US, 860
criteria, 1203
inflammatory back pain, aspects, 1207t
definitions, 848t
magnetic resonance imaging (MRI), 849-852
term, usage, 1202-1203, 1221-1222
ultrasonography, 860
Axis, upward subluxation, 1208
Axis of rotation, 80
Azathioprine (AZA), 946-948, 1141-1142
Basal lamina, cytoskeleton (linkage), 73
Basal joint stress radiograph, 716f
Babesia microti
Bacillary angiomatosis, 1860
osteomyelitis, 1860
Bacille-Calmette-Guérin-induced arthritis, 1265-1266
cause, 1265
clinical features, 1265-1266
diagnosis, 1265-1266
epidemiology, 1265
outcome, 1266
pathogenesis, 1265
treatment, 1266
Bacille-Calmette-Guérin (BCG) vaccine, 1834
Bacillus coagulans GBI-30 supplements, 1055
Back conditions, 449
Back-knee deformity, 684
Back pain
natural history, 665
nonsurgical interventional therapies, 678
radiation, 666-667
symptom, 675
Balb/c mice, collagen antibody-induced arthritis, 1534
Balb/c mice, collagen antibody-induced arthritis, 1534
Balb/c mice, collagen antibody-induced arthritis, 1534
Balb/c mice, collagen antibody-induced arthritis, 1534
Ball-and-socket joints, 1
formation, 568-569
Balb/c mice, collagen antibody-induced arthritis, 1534
Ball-and-socket joints, 1
formation, 568-569
Balb/c mice, collagen antibody-induced arthritis, 1534
Ball-and-socket joints, 1
formation, 568-569
Balb/c mice, collagen antibody-induced arthritis, 1534
Ball-and-socket joints, 1
formation, 568-569
Behçet’s disease (BD) (Continued)
ophthalmic features, 1527
pathogenesis, 1525-1526
pregnancy, relationship, 547-548
prognosis, 1530
pulmonary abnormalities, 1527
renal manifestations, 1527
revised international criteria, 1528
flowchart, 1528f
synovial membranes, biopsy specimens, 1528
System for Thalidomide Education and Prescribing Safety (STEPS) protocol, 1530
systemic disease, 1530
systemic manifestations, 1527
treatment, 1529-1530
list, 1529t
Behçet’s syndrome, 281, 964-965
asymmetric polyarthritis, 1128
Belimumab
phase III placebo-controlled RCTs (BLISS) evaluation, 1313
usage, 1295-1297
Benchmarking states, 470
Benign bone tumors, phalange location, 718
Benign cutaneous PAN, 610
Benign giant cell synovitis, 1963
Benign synovitis, 1963
Benign vascular tumors, rarity, 1956
Benzbromarone, potency, 1570
BeST (Behandel-Strategieen Treatment Strategies) study, 1146
Beta (probability), 456
Beta-catenin, usage, 5
Beta-carotene, concentration levels, 1049
Beta-cryptoxanthin, dietary intake, 1049
Beta-tryptase
serum levels, elevation, 237
systemic levels, anaphylaxis marker, 236
Beta-adrenergic (β-adrenergic) receptors, 422
Beta; (β1GPI) apolipoprotein H, 1331-1332
Beta; (β) integrin chains, expression, 21
Beta, microglobulin (β2-microglobulin) concentration, impact, 1899
fibril formation, factors, 1891
Beta, (β) integrins, presence, 22
BO9588, usage, 1313-1314
Bias
information bias, 437
minimization, 436
recall bias, 437
selection bias, 437
source, 434
Biceps, long head
dislocation, 654
tenderness, 642f
Biceps brachii, mixed connective tissue disease (presence), 1443f
Biceps tendon
long head, origination, 658
rupture, occurrence, 654
Bicipital tendinitis
early phases, association, 654
rotator cuff tear, association, 654
rupture, 654-655
suspicion, 643
treatment, 654-655
Bicipital tendons, arthroposes, 778-779
Bifidobacterium lactis (comparison), 1055
Biglycan, 37
proteoglycan encoding, 3-4
Carpal tunnel syndrome (CTS) (Continued)
splinting, usage, 710
surgical release, indication, 711
symptoms, 710
treatment, 710
night splint, usage, 710f
viewpoint, 712
Cartilage, marketing, 93
Cartilage
anatomic cartilage, 20
anatomic distribution, 33
anlagen, formation, 5
cell density, decrease, 34
cellular component, 1712
collagen, 35-37
damage, extrinsic pathway, 112
defect, intrinsic repair, 56
degradation, hyaluronic acid (HA) marker, 1633
destruction, 1101-1105
proteases, impact, 112-113
psoriatic arthritis, relationship, 1243-1244
development, 1959
direct analysis, OA patient usage, 48
embryonic lineages, 8
exploit (organ) cultures, 26
extracellular matrix components, 35t-36t
formation, endochondral ossification, 5-8
global gene expression, genomic/proteomic analyses, 40-41
growth plate, 1719-1720
hair hypoplasia, 1723
histschemical analysis, 33-34
immune complexes, embedding, 1083
intermittent compression, 17
lesions, protective effects, 776
matrix-degrading proteases, 48-50
metabolism, quantitative determination, 1640
morphogenesis, molecular signals, 6
nodules, attachment, 1959
oligomeric protein, appearance, 5
organic constituents, 33-34
osteoarthritis, impact, 535
pathology, chondrocyte role, 48-55
proteoglycans, 37
degradation, 110
radiographic changes, 1958
regenerative process, 1626
repair
challenges, 56-57
chondrocyte role, 55-57
structure, 20-26, 1720
skeletal chondromatosis, cellular characteristic, 1959f
template, replacement, 2
vascularization, 16
Wnt signaling, 92
Cartilage-associated protein (CRTAP), 1728-1729
Cartilage-associate articular surface, 831
Cartilage-derived morphogenetic protein 1 (CDMP-1), 2-3
Cartilage destruction
chondrocyte-derived proteases, role, 49f
cytokine balance, 50-54
cytokine regulation, 51t
Cartilage intermediate-layer protein (CILP), 1632
chondrocyte expression, 38
expression, 1580-1581
Cartilage matrix
aggreganases, impact, 1623-1625
anabolic factors, 1622
Cartilage matrix (Continued)
biomechanical changes, 1621-1622
cartilage degradation, 1623
cartilage repair, 1622
catabolic factors, 1623
collagenases, impact, 1624
components, structure-function relationship, 35-38
cysteine proteases, impact, 1623
damage, 16
degradation, 27
chondrocyte protease mediated, 50t
markers, 56
enzyme inhibitors, impact, 1625
metabolic changes, 1622-1623
metabolism, alterations, 1621-1625
metalloproteases, impact, 1623
proteases, classes, 1623
remodeling events, 7
serine proteases, impact, 1623
turnover, markers, 56
Cartilage metabolism
cytokine balance, 51f
cytokine signaling pathways, involvement, 54-55
growth/differentiation (anabolic) factors, roles, 45-48
Cartilage oligomeric matrix protein (COMP), 37, 483, 1632, 1726
diagram, 1726f
expression, 483
illustration, 3f
levels, increase, 483
markers, usage, 482
molecules, accumulation, 1346
mutations, heterozygosity, 1726
origin, 56
Cartilage protein-induced arthritis, 404-406
Cartilage-specific antigens, 1070
autoimmunity, 1069-1070
Cartilage-specific collagens, encoding, 35
Cartilage-specific matrix proteins, 38
Cartilage-specific type II collagen, expression, 41
Cartilage-synovium junction, 21f
Case-control studies, 434
bias, 434
controls, selection, 434
definition, 434
nested case-control studies, 435
recall, variation, 437
weaknesses, 434
CASPAR (Classification Criteria for Psoriatic Arthritis) study, 995
Caspar activation and recruitment domain (CARD), 261-262
CARD, R [purine]-binding, pyrin, lots of
weaknesses, 434
Cell death, 1096-1098
eating, 389-391
initiation, 389f
location, 389-391
program, 385
regulation, 382
regulators, Bel-2 family, 388
removal/response, 389-391
Cell determinant CD4 (CD4)+T cells, 277
Cell determinant CD4 (CD4)+T cells, impact, 378-379
Cell determinant CD8 (CD8)+ T cells, helper cell regulation, 294
Cell determinant CD8 (CD8)+ T cells, thymocytes, 175
Cell determinant CD8 (CD8)+ T cells, thymus, 175
Cell determinant CD8 (CD8)+ T cells, thymus, 175
Connective tissue activating protein III (CTAP-III), 361
Connective tissue disease (CTD), 1522
CTD-associated vasculitis, 1515-1516
direct immunofluorescence, 1516f
cutaneous vasculitis, association, 1516
malignancy, association, 1941t
overlaps, management, 1446-1448
vasculitis, occurrence (rarity), 1515-1516
Connective tissue growth factor (CTGF)
CCN2, 1360
impact, 362
skin fibrosis development suppression, 1396
SSc, association, 1344
Connective tissue mast cells (CTMCs), 233-234
CONSORT Guidelines, 459
Constant region, 192
Constipation, opiates (impact), 1022
Constitutive chemokines, 226-227
Construct validity, 455
assessment, 469
change measurement, 469-470
measurement, 469
Content validity, 455
appraisal, 469
Contrastive perichondrial envelope, interzone (presence), 7
Continuous blastaema, absence, 8
Contractile apparatus, 70-72
components, 64f
Contractile proteins, 68t
Contrast-enhanced T1W FS sequences, 846
Control/confidence, loss, 1034
Control group, sampling, 434
Conventional chondrosarcoma, 1965
Conventional dendritic cells (cDCs), 117-120
decrease, 129-130
increase, 130
LAMP1+ endocytic compartments, 129
subcategorization, 118
Conventional DMARDs, 1140-1141
types, 1140
Conventional radiography, imaging
considerations, 1640
joint issues, 1640-1641
Convertases, regulation, 30f
Coracoclavicular ligament, response, 651
Core binding factor (Cbfa1), 38-39
Core sets
disease-specific instruments, 462-465
indices, 465
longitudinal observational studies, 463t
rheumatologic conditions, 463t
Corneal melt, 621-622
peripheral cornea, tissue (thinness), 622f
Coronal alignment, assessment, 684f
Coronary artery disease (CAD)
risk factors, 1124-1125
SLE, relationship, 1294-1295
Corrective osteotomies, usage, 700-701
Cortical bone, composition, 61
Cortical thinning, 1728
Corticosterone-binding globulin, 896-897
Corticosteroid-induced osteonecrosis, 1698
Corticosteroids
blood pressure, elevation (risk), 549-550
dose levels, control, 546
effectiveness, 1216-1217
intralesional injections, 166
locally injected corticosteroids, negative effects, 776
Crohn’s disease (CD) (Continued)

tumor necrosis factor inhibitors, 1261r
UC, contrast, 1259
Cronbach alpha coefficients, usage, 469
Cross-bridge cycle
actomyosin, 72f
chemomechanical events model, 72-73
Cross-linked class II collagen-containing fibrils, network, 33-34
Cross-presentation, 125
Cross-priming, 391
Cross-product, 303-304
Cross-reactive immune response, 401
Cross-reactivity, example, 292
Cross-sectional hypertrophy, resistance training (impact), 76

Cuff tear arthropathy, 651, 657-658, 1586-1588
therapy, difficulty, 658
Cuff tendinopathy, radiographs, 651
Cultured autologous chondrocytes, transplantation, 56
Cultured RA synovial fibroblasts, disease-specific phenotypes, 224
Curved joint, transverse displacement, 87f
Cushing’s disease
primary Cushing’s disease, cause, 1929
primary musculoskeletal disease, confusion, 1929
Cushing’s syndrome, 1424
Cutaneous DM, skin biopsy, 606
Cutaneous lesions
diagnosis, 601
types, 601
Cutaneous leukocytoclastic angiitis, 1456, 1511-1512
term, CHCC preference, 1456-1457
Cutaneous lupus, variants, 604-605
Cutaneous lupus erythematosus (CLE), 1315
Cutaneous mucinosis, 1438-1439
Cutaneous NLE manifestations, 1776
Cutaneous polyarteritis nodosa (cutaneous PAN), 610, 1503
nodular skin lesions, 1794f
Cutaneous vasculitis, 1182
association, 1516
blood vessels, involvement, 1509f
Cutis laxa, 1736
Cut-repeat homeobox protein (Cux), 3f
Cux-1 (homeobox factor), 2-3
CX/CXC families, 1091
CXCL13
binding, 1091
expression, 1176
CXCLs, 1091
CXCL1 (fractalkine), involvement, 362
Cyclooxygenase monophosphate (cAMP) AA access (blockage), NSAID function (impact), 875
decrease, 413-414
Gs-mediated increase, 874
increase, 107
intracellular concentrations, impact, 158
Cyclic citrullinated peptide (CCP), 806-807
anti-CCP antibodies, clinical relevance, 806-807
studies, 1747
Cyclical RGD peptide, intra-articular administration, 1100
Cyclin D1, convergence, 6
Cyclobenzaprine, 1030
Cyclooxygenase (COX), 871
action, mechanism, 876-877
active site (inhibition), NSAIDs (impact), 876
activity, inhibition, 162
adverse effects, 882-888
biochemistry, 874-876
biology, 872-876
classification, 877-879
COX-2-selective NSAID, impact (clinical trials), 882
cyclooxygenase-independent action mechanisms, 877
dosing, 877-881
impact, 1044-1045
inhibition, 876
isoenzymes, inhibition, 340-341
isomers, relative inhibition, 885
isolation, 872
Cyclooxygenase (COX) (Continued)
  isozyme selectivity, 877
molecular biology, 876
pathway, 161, 342-344
  products, 344-351
pharmacology, 877-881
reactions, occurrence, 874-875
structural biology, 874-876
substrates, 872-873
Cyclooxygenase-1 (COX-1)
balance, 343-344
expression
generalization, 876
  regulation, 342-343
gene (PTGS1), expression, 342-343
inhibition, 882-883
serine 530 (acylation), ASA function
  (crystallographic studies), 876
structural differences, 875
substrate binding channels, 875f
Cyclooxygenase-2 (COX-2)
balance, 343-344
COX-2-specific inhibitors, 1650
discovery, 872
expression, glucocorticoid regulation,
166-167
expression, regulation, 343-344
gene
  genetic deletion, 887
  structure, 876
inducible COX-2, expression, 1630-1631
inhibition, 872, 882-883
inhibitors, 457
knockout mice, inflammation (occurrence),
3 time course
production stimulation, 52
restriction, 1193-1194
selectivity, 877
structural differences, 875
substrate binding channels, 875f
Cyclophosphamide (CYC), 942-945
absorption, 942
action, mechanisms, 942
administration
dosage, 943
route, 943, 1306-1308
bladder toxicities, 944
clinical indications, 943
cyclophosphamide-induced pulmonary
toxicity, occurrence, 944-945
distribution, 942
dosage, 943
drug interactions, 945
elimination, 942
extrarenal disease, 1287-1288
hematologic side effects, 943-944
impact, 944, 1664
IV pulse therapy, 943, 1310t
malignancy, 944
metabolism, 942
flowchart, 942f
pharmacokinetic considerations, 942-943
pharmacology, 942-943, 1306-1308
pregnancy, relationship, 553
pulse intravenous therapy, administration/
monitoring (NIH protocol), 1290t
reproductive side effects, 944
structure, 942
toxicity, 943-945
  minimization strategies, 945
urinalysis, 945
urologic side effects, 944
usage, 522, 543, 944, 1306-1308
Cyclosporin A (CsA), 1311-1312
AZA, comparison, 1311
pharmacology, 1311
Cyclosporine (CSA), 948-951, 1142
absorption, 948
action, mechanism, 948
clinical drug interactions, 951t
clinical indications, 949-950
clinical use, 949t
concomitant administration, 898
cyclosporine-induced hyperuricemia/gout,
1563
distribution, 948
DMARD, 521-522
dosage, 948, 949
drug interactions, 950-951
elimination, 948
gastrointestinal side effects, 950
hypertension, relationship, 950
lactation, 950
malignancy, 950
metabolism, 948
nephrotoxicity, 950
ophthalmic preparation, 1186
pharmacology, 948
pregnancy, 950
structure, 948
toxicity, 950
  minimization, strategies, 950
usage, 606
Cystathionine β-synthase deficiency, 1735
Cysteine proteinases, impact, 1623
Cysteine proteinases, 97-98
cathepsins, 1103
inhibitors, 105
lyosomal enzyme, 49
Cysteine-rich domains (CRDs), 386
cysteine-rich secretory protein 3 (CRISP3),
description, 154
cysteine X cysteine 12 (CXC12),
cysteine proteases, impact, 1623
Cysteine proteinases, 97-98
cathepsins, 1103
inhibitors, 105
lyosomal enzyme, 49
cysteine-rich domains (CRDs), 386
cysteine-rich secretory protein 3 (CRISP3),
description, 154
cysteine X cysteine 12 (CXC12),
cystatin C, 1103
viral triggers, 1174
Cytokine expression
Cytokine-deprivation-induced apoptosis,
166-167
Cytokine expression
CD4+ T cells, usage, 183
modulation, post-translational regulation,
373
NFκB mediation, 55
promotion, 372
regulation, 371-373
Cytokines, 477
acute inflammation, 373
acute phase response, 823
alternative processing pathways, 373
antigens, 1092-1093
BAFF family, 197
schematic, 198f
cell regulation, 372-373
classification, 369
co-localization, 51
cytokine-specific neutralizing antibodies,
usage, 51
Cytokines (Continued)
effect function, 373-380
function, in vitro/in vivo assessment, 369-370
generation, 164-165
activity, 1243
inhibitory cytokines, 52
levels, increase, 823
networks, 51-52
modulation, non-MHC genes (usage),
1196
pleiotropic factors, 50
polymorphisms, 1062
pre-formed cytokines, 236
production, 44, 161
triggering, 264
profile, 1093
roles, 375f, 377f
secretion, 121
signaling, 294, 371f
pathways, 372f
SSc, relationship, 1359-1362
synthesis, 236-237
time course, 1081
Cytokines and chemokines (CCL19-21),
production, 129
Cytolytic T cells (CTLs)
CTLA-4, expression, 1484
CTLA-4 fragment, extracellular portion,
989
CTLA-4 gene, genetic deletion, 183
CTLA-4-lg, recombinant protein, 1314
viral infection recognition, 183
Cytomegalovirus (CMV)
CD28+ T cells, increase, 181
viral triggers, 1174
Cytopenias, supportive treatment (necessity),
1209
Cytoskeleton
basal lamina, linkage, 73
extracellular matrix, connections, 74f
CytoS, protein degradation, 1434f
Cytosolic phospholipase (cPLC), 341-342
Cytosolic PGH-PGE isomerase, 340-341
Cytosolic phospholipase (cPLC), 341-342
Cytosolic phospholipase (cPLC),
341-342
Cytosolic phospholipase (cPLC),
341-342
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
Cytosol, protein degradation, 1434f
C2 homologous complement component deficiency, 334
C3 activation/degradation, 329f
C3/C5 convertases, proteins, 328f
C3 deficiency, 334
C3 receptors, 330t
C4 activation/degradation, 328f
C4/C4a, redaction, 1295
C4 genes, duplication, 334
C4 receptors, 330t
C5a peptide, 1885-1886

D
Dactylitis, 1225f
DAIR (Debridement, antibiotics, and implant retention), 1810-1811
Damage-associated endogenous TLR ligands, 1361
Damaged progression, longitudinal assessment (CT usage), 842
Damage indices, 465
DANCER (Dose-ranging Assessment: Damage indices), 465
Danger-associated molecular patterns
Danger model, 286
Dancers
injuries, distribution, 501
musculoskeletal disorders, 501
Danger model, 286
Danger-associated molecular patterns (DAMPs), 117, 145
protein families, 257, 391
recognition, 257
self-molecules, 284
Dantrolene, 1029
Dasatinib, usage, 1396
DAS28, DANCER changes, 982f
Dasatinib, usage, 1396
Dantrolene, 1029
Dassart, usage, 1396
DAS28, DANCER changes, 982f
Daten analysis, involvement, 459
data-based stages, 467-468
data-based, 467-468
incompleteness, problem, 458-459
daVinci, Leonardo, 79
DBA/1 mice, collagen-induced arthritis, 119
DC natural killer (NK) lectin group receptor-1 (DNGR-1), 119
DC-specific ICAM-3 grabbing nonintegrin (DC-SIGN), 122, 126
role, 122
DC/T cell interaction, DC-SIGN (role), 122
Deactivating stimuli, impact, 148
Death, 465
vascular involvement, 1408
cardiovascular causes, 1134
disease-specific core sets, relationship, 465
ligands, 386-387
Death domain (DD), 386
three-dimensional structure, 386
Death effector domain (DED), 386
Death receptors (DRs), 386-387
inhibition, 388
signal transduction, 386
Debranching enzyme deficiency, 1539
Débridement, antibiotics, and implant retention (DAIR), 1810-1811
Decay-accelerating factor (DAF), 221-222, 1071
CD55, 21
DAF-coated fibrillin-1 microfibrils, usage, 21
protease displacement, 330f
Decision-making process, 467
algorithm, 468f
Declaration of Helsinki (World Medical Association), 457
Decompressive laminectomy, alternative, 680
Decorin, 37
proteoglycan encoding, 3-4
proteoglycans, 12
Dectin
dectin-1, 259
dectin-2, CLR association, 106
receptors, engagement, 284
DEC-205, 122
Dedifferentiated phenotype, 41-42
Deep heat applications, 535-536
Deep mycoses, epidemiology/clinical features, 145
Deep perivascular lymphocytic infiltrate, 1181-1182
Deep tendon stretch reflexes, performing/grading, 635
Defective ribosomal products (DRPs), 125
Defective tolerance, 290-291
Deficiency interleukin-1 (IL-1) receptor antagonist (DIRA), 1613
Deformity, observation, 562
Degenerative arthritis, 560
manifestation, 628
physical activity, impact, 496-497
Degenerative cartilage disease, CPPD crystal deposition disease (association), 1584-1586
Degenerative joint disease, impact, 727
Degenerative rheumatic disorders, apoptosis (acceleration), 396
Degenerative spinal stenosis, 673f
Degenerative spondylolisthesis, development, 671
Degradative enzymes, n-3 fatty acids (effect), 1046
Degranulation, phagocytosis, 158-159
Degrees of freedom, 80
Dehydration (volume depletion), 1542
Dehydroepiandrosterone (DHEA), usage, 1315
Dehydroepiandrosterone (DHEA), usage, 1315
Dehydroepiandrosterone (DHEA), usage, 1315
Delta opioid receptor activation, 1022
identification, 1015
Demargination, 166-167
Demyelinating disease, impact, 622
Demyelinating syndromes, tumor necrosis factor inhibitors, 948
Dendritic cell inhibitory receptor (DCIR), 122
Dendritic cells (DCs), 1077-1078
activation, 121
role, 1046
antigen presentation pathways, 124f
antigen-presenting cells, 117, 198-199, 1277
antigen-specific interactions (maintenance), 1277
immunologic synapses (usage), 270-271
antigen-specific T cells, contact, 272
chemoattractants, release, 130
DC-mediated modulation, 119f
derivation, 136
development, 120-121
differentiation, 1464
division, 117
extracellular vesicle generation, 128
immune responses, 1355-1356
innate immune system cell types, 285
life history/heterogeneity, 135-137
ligands, 121t
maturation, 139f
morphologies, 275f
multiple roles, 130
Dendritic cells (DCs) (Continued)
NK cell-mediated activation, 127
pathogen infection, 123
Peyer’s patch (PP), role, 1255f
positioning, 139f
stimulatory capacity, defect, 131
subsets, 117-120
synovial sublining macrophages, usage, 765-766
TLR expression, 121t
tolerance, 127-128
vitamin D expression, 1046
Denosumab (Prolia), administration, 1670
Degenerative rheumatic disorders, apoptosis (acceleration), 396
Degenerative arthritis, 560
manifestation, 628
physical activity, impact, 496-497
Degenerative cartilage disease, CPPD crystal deposition disease (association), 1584-1586
Degenerative joint disease, impact, 727
Degenerative rheumatic disorders, apoptosis (acceleration), 396
Degenerative spinal stenosis, 673f
Degenerative spondylolisthesis, development, 671
Degradative enzymes, n-3 fatty acids (effect), 1046
Degranulation, phagocytosis, 158-159
Degrees of freedom, 80
Dehydration (volume depletion), 1542
Dehydroepiandrosterone (DHEA), usage, 1315
Dehydroepiandrosterone (DHEA), usage, 1315
Dehydroepiandrosterone (DHEA), usage, 1315
Delta opioid receptor activation, 1022
identification, 1015
Demargination, 166-167
Demyelinating disease, impact, 622
Demyelinating syndromes, tumor necrosis factor inhibitors, 948
Dendritic cell inhibitory receptor (DCIR), 122
Dendritic cells (DCs), 1077-1078
activation, 121
role, 1046
antigen presentation pathways, 124f
antigen-presenting cells, 117, 198-199, 1277
antigen-specific interactions (maintenance), 1277
immunologic synapses (usage), 270-271
antigen-specific T cells, contact, 272
chemoattractants, release, 130
DC-mediated modulation, 119f
derivation, 136
development, 120-121
differentiation, 1464
division, 117
extracellular vesicle generation, 128
immune responses, 1355-1356
innate immune system cell types, 285
life history/heterogeneity, 135-137
ligands, 121t
maturation, 139f
morphologies, 275f
multiple roles, 130

Diffuse antineutrophil cytoplasm antibody, 621
Diffuse anterior scleritis, 621
Dietary components, effects, 1048t
Diet
Diarthrodial joint
Diapedesis, 155
Dialysis
Dial test, 688
Diagnostic methods, absence, 457
Diacerein
Diabetic stiff hand syndrome, 1930
Diabetic muscle infarction, 1931
Diabetic chiroarthropathy, 1930
Diabetes mellitus (DM), 912, 1930-1932
Dexamethasone, fluorine position, 549
Developmental bone growth, 64
Developing joint, innervation (role), 7-8
Destructive rheumatoid arthritis, chronic
Design considerations, 453
Descriptive analysis, 459
Dermatomyositis (DMS), positive results, 921
uric acid, relationship, 1543-1545
DMARDs, interactions, 1054
family, 1104-1105
Dickkopf-1, 64-65
weight-bearing surfaces, 13
structural properties/biochemical components, 8
weight-bearing surfaces, 13
Dipeptide peptidase IV, localization, 10
Dipyridamole, pain/fever relief, 677
Direct chemokine receptor inhibition, 364
Direct costs per case, 499
Direct immunofluorescence (lupus band test), 603f
Disability, 1036-1038
cross-cultural comparisons, 1037
insurance, 1037-1038
unemployment insurance, comparison, 1037
scales, 465
work disability, 465-466
Disadvantage category (Wolfe), 465
Discoid domain receptor 2 (DDR2), integrin (contrast), 44
Discoïd lupus, terminology, 602
Diseases, 1140
combinations, 1145-1146
biologic DMARDs, 1140
azathioprine, 521
alkylating agents, 522
alcohol, 522
Disease activity index for Psoriatic Arthritis (DAPSA), 1240-1241
Disease Activity Score (DAS), 455, 462, 1780
baseline DAS28, 510
basis, 456
DAS28, 465
calculation, 563
changes, improvement/deterioration, 544
remission criteria, 470
DAS28, DANCER changes, 982f
DAS28-CRP, 543-544
mean decrease, 457-458
measurement, 1163-1164
usage, 465
Disease-causing mutations, identification, 27
Disease-modifying antirheumatic drugs (DMARDs), 379
alkylating agents, 522
azathioprine, 521
biologic DMARDs, 1140
combinations, 1145-1146
biologic agents, 1145-1146
EPA suppression, 355
conventional DMARDs, 1140-1141
types, 1140
cyclosporine, 521-522
definition, 1140
diet, interactions, 1054
discontinuation, 1162-1163

Enthesitis-related arthritis (ERA) (Continued) feature, 1762 outcomes/prognosis, 1763 treatment, 1763 Enthesopathy, ultrasonographic pathology definition, 858t ENU-mutagenized mice, lupus models, 282 Environmental genes, interaction, 1063 triggers, 295-296 Environmental factors, contribution, 1194-1196 Enzyme inhibitors, impact, 1625 Enzyme-linked immunosorbent assay (ELISA) Enzyme inhibitors, impact, 1625 Environmental factors, contribution, 433t Enzyme-linked immunosorbent assay (ELISA), usage, 391-393, 793 Enzyme pre-identifications, 1481 Eosinophil-derived neurotoxin (EDN), 171 Eosinophilia-myalgia syndrome (EMS) Eosinophilic fasciitis (EF), 608, 1787 Eosinophilic fasciitis (EF), 608, 1787 diagnosis, 608 example, 1438f fascia, inflammation, 608 groove sign, 1399-1400 scleroderma, relationship, 1399-1400 skin changes, 1438 Eosinophilic fasciitis (EF), 608, 1787 Eosinophilic fasciitis (EF), 608, 1787 diagnosis, 608 example, 1438f fascia, inflammation, 608 groove sign, 1399-1400 scleroderma, relationship, 1399-1400 skin changes, 1438 Eosinophilic fasciitis (EF), 608, 1787 Eosinophilic fasciitis (EF), 608, 1787 diagnosis, 608 example, 1438f fascia, inflammation, 608 groove sign, 1399-1400 scleroderma, relationship, 1399-1400 skin changes, 1438 Eosinophilic fasciitis (EF), 608, 1787 Eosinophilic fasciitis (EF), 608, 1787 diagnosis, 608 example, 1438f fascia, inflammation, 608 groove sign, 1399-1400 scleroderma, relationship, 1399-1400 skin changes, 1438 Eosinophilic fasciitis (EF), 608, 1787 Eosinophilic fasciitis (EF), 608, 1787 diagnosis, 608 example, 1438f fascia, inflammation, 608 groove sign, 1399-1400 scleroderma, relationship, 1399-1400 skin changes, 1438 Eosinophilic fasciitis (EF), 608, 1787 Eosinophilic fasciitis (EF), 608, 1787 diagnosis, 608 example, 1438f fascia, inflammation, 608 groove sign, 1399-1400 scleroderma, relationship, 1399-1400 skin changes, 1438 Eosinophilic fasciitis (EF), 608, 1787 Eosinophilic fasciitis (EF), 608, 1787 diagnosis, 608 example, 1438f fascia, inflammation, 608 groove sign, 1399-1400 scleroderma, relationship, 1399-1400 skin changes, 1438
INDEX
Fibromyalgia (Continued)
diagnostic criteria, 739
differential diagnosis, 739-740
direct-to-patient advertising, 737
drugs, usage, 746
education, 745-746
enkephalins, concentrations (impact), 743
epidemiology, 741
etiopathology, 741-743
exclusion, diagnosis, 1435
exercise, 746
problem, 746
existence, problem, 736
explanation, 743
familial factors, 742
features, 730t
functional brain imaging, 743
functional questionnaire, results, 740
genetic factors, 742
gray matter loss, 744
hippocampus dysfunctioning, proton magnetic resonance spectroscopy (usage), 743-744
historical development, 738
imaging techniques, 743-744
low back pain, 676
management, 744-749
medicalization, 736-738
medicolegal issues, 748-749
microtrauma, 742
muscles, involvement, 742
neurometabolic assessment, proton magnetic resonance spectroscopy (usage), 743-744
1990 ACR criteria classification, 736t
nociceptive information process, deficits, 744
nonmedical/societal issues/concerns, 737f
nonpharmacologic treatments, 747
occurrence, 739
opioids, usage, 748
outcome, 749
Outcome Measures in Rheumatoid Arthritis Clinical Trial, recommendation, 740
pain, 1036
inhibition, decrease, 743
processing, 743
pathophysiology, 741-743
pathogenesis
description, 560
resistance, 748
screening (National Data Bank for Rheumatic Diseases), 745
pharmacologic treatments, 746-747
physical function, 740
physical measures, 740
physical therapy, 748
polarticular scenarios, contrast, 588-591
polysymptomatic distress, assessment, 736f
prevalence, 741
primary neuroendocrine dysregulation, 742
proton magnetic resonance spectroscopy, usage, 743-744
psychosocial factors, 742
psychotropic agents, 746
randomized controlled trial (RCT), 745
multicomponent treatment, meta-analysis, 747
recommendations, 747-748
research
data, 741
studies, 744-749
research questionnaire, 740
self-report measures, 740
Fibromyalgia (Continued)
sensory gating, 744
severity, assessment, 740-741
single photon emission computed tomography (SPECT) studies, 743
sleep disturbance, 742
social construction, 736-738
somatosensory stimulation, 744
spa treatment, 748
stimulus-response function, 743
stress-related neuroendocrine dysfunction, 742
studies, interpretation, 745
symptoms
National Data Bank for Rheumatic Diseases, 738t
severity, 740
syndrome, 733
central nervous system (CNS), involvement, 743
tender point examination, 740-741
limitations, 741
treatment
alternatives, 748
compliance, 744
trials, 744-745
value, 744
types, 739
understanding, EULAR recommendation, 747-748
upper/lower back pain, 738
work impairment, 1038
work status, 740
Fibromyalgia Impact Questionnaire (FIQ), 740
Fibroconnectin, 11
chondrocyte adhesion, 44
degradation, 111
extra domain A (EDA), 1361
fragments, binding, 44
molecules, accumulation, 1346
presence, 807-808
Fibroconnectin-binding proteins, 1773
Fibroconnectin-like immunoglobulin, components, 65
Fibrosing disorders, examples, 1438f
Fibrositis, 1124, 1356-1359
distribution, 1348
effector cells, impact, 1358
impact, 545
increase, disorders (association), 1435t
molecular determinants, 1359
pathogenesis, 1357f
treatment, 1396
Fibrosubintima, 24
Fibroxanthoma, tendon sheath, 1964
Ficat and Arlet stage 3 osteonecrosis (treatment), Charnley/Bicentric hemiarthroplasty (success rate), 1706
Ficolins, 258
ficolin 1, description, 154
15d-PGJ2, endogenous ligand, 1362
15-epi-LXAr, anti-inflammatory SPM, 352
15-hydroxyprostaglandin dehydrogenase (PGDH), PG degradation, 345
Fifth digit, tophus, 1559f
Filopodia, 20
Find-me signals, 389
Fingernail dystrophy, AL amyloidosis, 1893f
Fingers
clubbing, 1260
description
boutronniere deformity, 567
swan neck deformity, 567
scleroderma vasculopathy, histologic appearance, 1346f
swelling, 567
Fingertip, distal pulp infection (felon), 718
Finite spatial rotation, sequence (importance), 81-82
Finland Rheumatoid Arthritis (Fin-RA) trial, 936
FINRACO study, 1144
FIP-1-like 1 (FIP1L1) genes, fusion, 172-173
First extensor compartment, tenosynovitis, 566
First flexion wave, 80-81
Fish oils
adverse effects, 1052-1053
anti-inflammatory doses
impact, 1052
NSAIDs, comparison, 1052t
toxicity, 1052-1053
5-amino-salicylic acid (5-ASA) (mesalamine), 929
5-hydroxy-eicosatetraenoic acid (5-HETE) impact, 347-348
production, 1670
5-hydroxyperoxy-eicosatetraenoic acid (HPETE), impact, 347-348
5-lipoxygenase (5-LO) gene, isolation/characterization, 346
pathway, 347f
5-lipoxygenase-activating protein (FLAP), contact, 346
5’-triphosphate nick end labeling (TUNEL) assay, 393
staining, 395
5,12-dihydroxy-eicosatetraenoic acid, 346
Fixed determinants, modifiable determinants (contrast), 1040
FK506 (Tacrolimus), 948-953
Flagellin, exposure, 1252-1253
Flat bones, 61
Flexion abduction external rotation (FABER) test (Patrick test), 570
usage, 695
Flexion contractures, sclerodactyly, 607f
Flexion McMurray test, 688-689
indication, 689f
Flexor carpi radialis, 711
Flexor carpi ulnaris tendinitis, 711
Flexor hallucis longus tendon, tendinitis, 501
Flexor tenosynovitis (trigger fingers), 1930
arthrocentesis, 781
FLICE-inhibitory protein (FLIP), 388
Flinching polyarthritis, 581-582
Fludarabine, usage, 1308
Fluid-phase inhibitors, 330
Fluorescein isothiocyanate (FITC), 410
Flitumustine, usage, 1773
Fluids
chronic kidney disease, 551
solvent, 81-82
Flotillin, usage, 1377
Flottard, usage, 1378
Fluorescence isothiocyanate (FITC), conjugation, 391
Fluorescent antineural antibody (FANA) development, 789-790
interpretation, 794-795
positivity, cutoff, 795
results, interpretation, 792-793
tests, 792f-793f
Focal adhesion kinase (FAK), 388
Focal adhesion plaques (FAP), 43
involvement, 363
role, 219-220
Focal CNS deficits, 1183
Focal myocardial fibrosis, 1389
Folate (vitamin B9) deficiency, SSZ (impact), 551
definition, 1122
INDEX
Giant cell arteritis (temporal arteritis) (GCA), 1467
Giant cell arteritis (temporal arteritis) (GCA) (Continued)
large arteries, 1467f
involvement, clinical evidence, 1467
liver function test results, 1468
manifestations, 1465-1467
MTX, usage, 921
muscular arteries, inflammation (presence), 1463
negative unilateral biopsy, management, 1470
neurologic problems, occurrence, 1466
ocular complication, 1465-1466
pathogenesis, 1463-1465
model, proposal, 1464f
pathology, 1463-1465
physical findings, 1468t
plasma proteins, nonspecific changes, 1468
polyanglia rheumatica, relationship, 824, 1468
prednisone, continuation, 1471-1472
proximal aorta, involvement, 1464f
respiratory tract symptoms, occurrence, 1467
serum creatine kinase levels, 1468
signs, likelihood ratios, 1469t
skin findings, 610
symptoms, 1465t
likelihood ratios, 1469t
synovial fluid analyses, 1468
TA, comparison, 1477t
temporal artery
color duplex ultrasound examination, 1471f
cross-section, 1463f
implantation experiments, 1464
involvement, 1465f
thoracic aortic aneurysm, increase, 1472
Toll-like receptors (TLRs), involvement, 1464
treatment, 1471-1472
temporal artery biopsies, 1470
visual symptoms, 1465
Gignac, Work Activity Limitations Scale, 1318
Gillworth, Work Instability Scale, 465-466
Glenohumeral disorders, 656
Glenohumeral instability, 659-660
diagnosis, 660
manifestation, 660
pathologic condition, 659
plain radiographs, 660
subluxation, recurrence, 659
treatment, 660
Glenohumeral joint, 639
arthrocentesis, 778, 778f
ostearthrosis, 657
proximal structures, 2
Glennoid labrum tears
anterior-inferior portion, double-contrast
arthrography, 645f
identification, DCAT (usage), 644
Glioma-associated oncogene homolog (Gli), 3f
Gliomerular filtration rate (GFR), allopurinol, 1004-1005
Glomerulonephritis, 1183
ISN/RPS classification, 1274
Glucocorticoid-induced osteoporosis, 1671-1672
prevention/treatment, recommendations, 1671t
Glucocorticoid receptors (GRs), 352-353
Glucocorticoid-responsive elements, binding, 898-899
Glucocorticoids (GCs), 1140
administration, 900
timing, 906
adverse behavioral effects, 912-913
adverse events (AEs), 906, 909-913, 909t
study, 913
alternate-day regimens, 906
anti-inflammatory effect, duration, 903
atherosclerosis, 911
biologic characteristics, 895
biologic half-lives, 897
body weight, 912
cardiovascular adverse effects, 911
cataracts, 911
characteristics, 894-898
chronobiology, 905-906
classification, 894-895
concomitant use, 897-898
dermal adverse effects, 912
DMARD function, 904-905
doses
adaptations, 907-908
combination, 1318
drug interactions, 897-898
effects, negative studies, 905
decline adverse effects, 912
fat redistribution, 912
gastrointestinal adverse effects, 910-911
genetic action, 899f
genomic mechanisms, 898-899
genomic/nongenomic effects, 898
glaucoma, 911-912
hazard, 1834
IBM response, 1426
immunologic adverse effects, 911
immunosuppressive actions, 907
intralesional/intra-articular injections, 908-909
joint-sparing effect, 905
mechanisms, 898-903
mineralocorticoid effects, 911
minor mood disturbances, 913
monitoring, 909-913
myopathy, 910
ocular adverse effects, 911-912
osteopenia, 910
osteoarthritis, 910
patient cessation, 1151
peptic ulcer disease, 910
perioperative care, adaptations, 907-908
pharmacodynamics, 897t
pharmacokinetics/pharmacology, 896-897
pharmacologic doses, 901
plasma half-lives, 897
plasma proteins, affinity, 896-897
pregnancy/lactation, relationship, 898
pulse therapy, 908
radiologic joint damage, 904-905
red type, effects, 900f
sensitivity/resistance, 906-907
side effects, hypothesis, 899
skeletal adverse effects, 910
soluble glucocorticoids, onset, 908
steroid psychosis, 913
stress regimen, adaptations, 907-908
structure, 894-895
tapering scheme, 907t
term, usage, 894
Gout (Continued)
therapeutic consequences, 895
therapeutic ratio, improvement, 906
therapy, 903-905
interplay, 910f
requirement, 1672
signs/symptoms, 903-904
spectrum, 920f
toxicity, 1306
treatment, 903-909
indications, 903
usage, 1324
withdrawal regimens, 907
Gouty arthritis, 1128
Gouty tophus, 568f
tophaceous deposits, 568f
tophaceous gout, 1550
images, 837f
transition, 1545
treatment, 1564-1571
colchicine, usage, 167
collagen, 1050
cancer, 1050
compliance, 1572
options, 1571f
urate-lowering properties, 1007t
ultrasound, 861
diagnostic/prognostication
usage, 861
effects, 862f
usage, 837-838

Gouty arthritis, 725
G protein-coupled FPRL1, binding, 822
G protein-coupled receptor (GPCR), 156
GPCR-related receptors, extracellular
domain, 137f
mediation function, 874
PG actions, 351
G proteins, 156
composition, 156
GPR32 (human orphan receptor), binding, 352
gp39, cartilage-specific antigens, 1070
gp130 IL6 receptor, mutation, 408
Grading of Recommendations Assessment,
Development, and Evaluation (GRADE)
framework, 530
Graft-versus-host disease (GvHD), 92-93,
1947
systemic disease, 282-283
Gram-negative bacteria, 135-136
Gram stain, performing, 757-758
Granulocyte chemotactic protein 2 (GCP-2),
361
Granulocyte-macrophage colony-stimulating
factor (GM-CSF), 351
cytokine production, 44
signaling, 135-136
impact, 246-247
presence, 120, 1090
support, 1090
Granulomatosis with polyangiitis (GPA), 988,
1454-1455, 1486-1489
cardiac involvement, 1489
clinical features, 1486-1487
Granulomatous angiitis of the central nervous system (GACNS), 1519

treatment, 1523

vasculitis, histopathologic findings, 1519f

Granulomatous arteritis, 1461

Granulomatous inflammation, vasculitis (association), 1456f

Granulomatous vasculitides, 610

skin biopsy, usage, 610

Granulomatous angiitis of the central nervous system (GACNS), 1519

treatment, 1523

vasculitis, histopathologic findings, 1519f

Granulomatous arteritis, 1461

Granulomatous inflammation, vasculitis (association), 1456f

Granulomatous vasculitides, 610

skin biopsy, usage, 610

Granulomatous inflammatory, vasculitis (association), 1456f

Graves’ disease, 1461

Graves’ disease (hyperthyroidism), 281-282, 1663, 1928

Hashimoto’s disease, relationship, 1928

Graves’ ophthalmopathy, 1928

Graves’ disease, 1461

Greater trochanters, examination, 1209

Gremlin, upregulation, 48

Groin pain, impact, 570

Ground reaction force (GRF), 83

body rotation, 85f

location, 84f

Group A β-hemolytic streptococci, 1876-1877

Rockefeller University Hospital rheumatic fever patients, 1876f

Group A carbohydrate (GRA-CHO) antibodies, role, 1877

immunogen, 1885

valvular glycoproteins, cross-reaction, 1875

Group A streptococcal carbohydrate epitope, cross-reactivity example, 292

Group A streptococcal membranes, rabbit serum immunization, 1877f

Group A streptococci, 1876-1877

classes, 1876

involvement, 1875-1876

phagocytosis, group A anticyclic carbohydrate antibodies (role), 1877

schematic representation, 1876f

Group for Research and ASsessment of Psoriasis and Psoriatic Arthritis (GRAPPA), 462-464, 1240

psoriatic arthritis treatment strategies evidence, 1244-1245

Growth and differentiation factor (GDF), 3f

chondrocyte response, mediation, 47-48

pathway activation, 47

Growth differentiation factor 5 (GDF-5), 2-3

Growth factors

exposure, prolongation, 395

SSc, relationship, 1359-1362

synthesis, 236-237

Growth hormone (GH)

axis, importance, 92

systemic administration, 92

Growth plate

development, molecular signals (involvement), 6

FGFR3, involvement, 6

formation, 5

models, 42

Growth-promoting peptides, impact, 1683-1682

Growth-related oncogene (GRO)α, 53

Guanosine triphosphatases (GTPases), 145

Guanosine triphosphate (GTP)

conversion, 1536

GTP-binding proteins, 156

Guillain-Barré syndrome, acute motor axonal neuropathy subtype, 292

Gull-wing deformities, 1640-1641

Gut

biology, microbiota, 1251-1254

immune privilege, 1252f

immune system, colonization (effects), 1256f

native bacteria, presence, 1252

permeability, increase, 1259

tolerance induction machinery, 1254f

trophic functions, 1252

Guillain-Barré syndrome, acute motor axonal neuropathy subtype, 292

Guyon’s canal (ulnar nerve entrapment), 711

Hands

Hamate hook fracture, coronal CT scan, 712f

Hamate fracture, 711-712

Halofuginone, usage, 1396

Hallux valgus deformity

preoperative/postoperative anteroposterior radiograph, 704f

Haloferminone, usage, 1396

Hamate fracture, 711-712

Hamate hook fracture, coronal CT scan, 712f

Hand-arm vibration syndrome, 494

Hand-based custom-molded splint, usage, 716f

Hand pain

arthroscopy, 710

diagnostic tests, 709-710

etiologies, 710-719

history, 708

imaging studies, 709

injections/aspirations, 710

magnetic resonance imaging, 709

neurodiagnostic tests, 709-710

arthroscopy, 710

clinical photograph, 702f

trephine/oldoperative anteroposterior radiograph, 704f

Hashimoto’s disease, relationship, 1928

Hashimoto’s thyroiditis, 281-282

Hashimoto’s disease, Graves’ disease (relationship), 1928

Hashimoto’s thyroiditis, 281-282

Hawley-type maxillary bite appliance, 731f

Hematitis influenzae, 1318-1319

protease IV protein, expression, 292

Hair loss, colchicine treatment (impact), 889

Haltex valgus deformity

clinical photograph, 702f

HapMap Project, 309

Harris Hip Score, 693

form, 1700f

usage, 1700

Harrison’s grooves, 1673

Hashimoto’s disease, Graves’ disease (relationship), 1928

Hashimoto’s thyroiditis, 281-282

Hawley-type maxillary bite appliance, 731f

Head size (increase), Paget’s disease (impact), 1414f

Headache

epidemiologic studies, 1483-1484

EULAR/PReS-endorsed consensus criteria, 1790

eyes, impact, 1488

functional polymorphism, allele frequency, 1459

occurrence, 1458

pachymeningitis, 1489f

pathology, 1486

pathophysiology, 1486

remission reduction therapy, 943

renal involvement, 1488

saddle-nose deformity, 1487f

occurrence, 718

strength, assessment, 568

symovial chondromatosis, 1958f

tenosynovitis/synovitis, demonstration, 1837f

Haplotype

definition, 303t

relative risk (determination), family structure (usage), 309f

Health and Nutrition Examination Survey (HANES I), 495

Health Assessment Questionnaire (HAQ), 462, 982, 1152-1153, 1370

Health care costs, study, 448

medical care costs, 442-444

per-person costs, 442

quality, improvement, 437

resources, distribution (inefficiency), 441

service, unit prices (change), 442

services, units, 443t

utilization/disparities, determinants, 1038

Health outcomes assessment

growth, areas, 471-472

instruments, availability, 462-466

content, measurement, 466-467

measurement, 466-467

needs, defining, 466

satisfaction, 466

Health state

description, 464

value, utilities, 464

Health status, 464-465

Health and Nutrition Examination Survey (HANES I), 495

Health Assessment Questionnaire (HAQ), 462, 982, 1152-1153, 1370

Healthcare cost

costs, study, 448

medical care costs, 442-444

per-person costs, 442

quality, improvement, 437

resources, distribution (inefficiency), 441

service, unit prices (change), 442

services, units, 443t

utilization/disparities, determinants, 1038

Healing wound borders, mast cell accumulation, 239

Health generic measures, 464

reading/educational levels, impact, 1038-1039

Health, Aging and Body Composition (Health ABC) Study, 1593-1594

Health and Nutrition Examination Survey (HANES I), 495

Health Assessment Questionnaire (HAQ), 462, 982, 1152-1153, 1370

Health assessment, 530

measurement, 1163-1164

modified HAQ (MHAQ), 530

rheumatoid arthritis (RA-HAQ), 530

disability, 464

Health care costs, study, 448

medical care costs, 442-444

per-person costs, 442

quality, improvement, 437

resources, distribution (inefficiency), 441

service, unit prices (change), 442

services, units, 443t

utilization/disparities, determinants, 1038

Health outcomes assessment

growth, areas, 471-472

instruments, availability, 462-466

content, measurement, 466-467

measurement, 466-467

needs, defining, 466

satisfaction, 466

Health state

description, 464

value, utilities, 464

Health status, 464-465

general health status, 464

symptoms, 464-465

utilities, 464

Hearing loss (progression), Paget’s disease, 1414-1415

Iron deposition, 1909
Hepatitis B virus (HBV) (Continued)

Hemolysis, elevated liver enzymes, low Hgb.

Hemoglobinopathies

Hemochromatosis, 583

Hemiarthroplasty

Hematoxylin and eosin (H&E) histologic staining, 1408f, 1416f

Hematopoietic stem cell transplantation (HSCT), 1320

Hematopoietic stem cells (HSCs), DC production, 99

Hematopoietic monocytic precursor cells, 63

Hematopoietic malignancies, 1325

Hematologic disease, 1318-1319

Hemarthrosis, acute hemarthrosis, 1915-1916

HELLP (hemolysis, elevated liver enzymes, low platelets) syndrome, 1334

Hemarthrosis, 1915

Hemorrhagic cystitis, risk (increase), 522

Hemostasis, 247

Hemostatic response, platelet granule compounds (role), 246t

Hen egg lysozyme (HEL), modification (absence), 125-126

Hereditary hemochromatosis (HHC), 1907

Hereditary glucocorticoid resistance, 906-907

Hereditary hemochromatosis (HHc), 1907

Hereditary periodic fever syndromes, 584, 1129-1130

Hepatitis B virus (HBV) (Continued)
non-hepatitis B virus polyarteritis nodosa, 1503
pulsed cyclophosphamide/methylprednisolone, usage, 1503t
pathogenesis, 1870-1871
polyarteritis nodosa, 1503
prognosis, 1504
reactivation risk, 965

Hepatitis C virus (HCV), 1129, 1871
clinical features, 1871
diagnosis, 1871
epidemiology, 1871
HCV-associated vasculitis, 1500
joint problems, relationship, 1129
mixed cryoglobulinemia, association, 1871
pathogenesis, 1871
treatment, 1871

Hepatitis virus-associated arthritis, 1266
clinical features, 1266
epidemiology, 1266
genetics, 1266
pathogenesis, 1266

Hepatocyte growth factor (HGF), 227

Hepatocytes, confocal imaging, 274f

Hepatitis, 1129

Hemophilia

Hemophilic arthropathy

Hepatitis A virus (HAV), 1129

Hepatitis B virus (HBV), 1129, 1870

Hepatitis C virus (HCV), 1129, 1871

Hepatitis E virus (HEV), 1129

Herpes zoster, lumbar nerve roots, 165-166

Herpetic whitlow, paronychia (comparison), 1432

Herpes simplex, 1432

Herpesvirus, 1432

Heterodimers, formation, 6

Heterogeneous nuclear ribonucleoprotein (hnRNP), 1432

Heterologous CII, 405-406
Human leukocyte antigen (HLA) (Continued)

pandemic, 1851
reactive arthritis, features, 1853
rhabdomyolysis, 1856
seronegative lower extremity peripheral
arthritis, 1852-1853
rheumatic diseases
association/occurrence, 1852
response, 1860-1861
vasculitis, association, 1858-1859

Human leukocyte antigen (HLA) alleles
investigation, 306
molecular definition, precision, 303
naming, 302-303
autoimmune diseases, association, 304
base pairs, 301
class I alpha chain genes, 301
class I molecule, 299-301
function, 1205
HLA-A2, crystal structure, 300
isotypes, functional correlates, 301
structure, comparison, 300
three-dimensional structure, 302
CARD15, association, 1258
class II alleles, RA (association), 305
class II associations, autoimmune diseases,
304-305
class II genes, centromeric situation,
301
class II molecules, 299-301
isotypes, functional correlates, 301
restricted tissue distribution, 301
cohort/case-control studies, contingency
table, 304
study association (detection), case-control
method (alternatives), 309
encoding, 299
HLA-ABC, staining, 1409
HLA-B27, 1205
discovery, 1194
gene, frequency, 592
impact, 1194-1196
misfolding, propensity, 1205
physiology, non-MHC genes
(modification), 1196
presence, 1853
prevalence, 1224
serologic groups, 303
serologic specificity, 304
structures, differentiation, 1195
testing, 1227
transgenic rats, DC stimulatory capacity,
131
HLA-DQ
autoimmune diseases, association, 306
molecules, combinatorial diversity, 306
subregion, 302, 306
HLA-DR
role, 1061-1062
shared epitope (SE) alleles, 813
subregion, 301
substance exposure, interaction, 1063

Human leukocyte antigen (HLA) (Continued)

HLA-DQ alleles, presence, 1130
serologic groups, 303
subset, 187
linkage disequilibrium, 307-309
map, 302
molecules, 299
function, 299
polymorphic characteristic, 302-303
role, 1405
naming conventions, 303
non-HLA genes, 1205-1206
odds ratio, calculation, 303-304
polymorphisms, 1899-1902
population-association studies, 303-304,
306-309
PTPN22 locus, 307
relative risk, estimate, 303-304
rheumatic diseases, association, 303-309,
304
sequence-based nomenclature, 303
signals, association, 1174
SNP variant, 307
terms, glossary, 303
trimalleolar complex, x-ray crystallographic
analyses, 300-301

Humans
ADAM gene family, members, 102
alpha/beta-T cell receptor, three-
dimensional crystal structure (ribbon
diagram), 301
anti-CAM targeting, 363
antichemokine studies, 364
body, pain pathways, 416
chondrocyte cell lines, expression, 43
diaphragmatic joint, structure/function
(relation), 1
FC receptors, expression, 144
genome, ADAM genes, 104
immunoglobulin (Ig) isotypes, properties,
193
interphalangeal joint, sagittal section, 2f
joints, structure (provision), 1
leukocytes, 15-HPETE/15-HETE (addition),
348
mast cells, mediator production, 235
generic organization, 301-302
matrix metalloproteinases, substrates,
100
MHC class II/environmental factors,
408-409
monocyte-derived dendritic cells,
maturational markers, 139
myositis patients, observations, 1410
neutrophils (superoxide generation),
5-HETE/5-HETE (usage), 347-348
PAD homologues, identification, 809
PAD4 (PAD5), discovery, 810
parvovirus B19, 1865
polymorphonuclear leukocyte chemotaxis
(blockage), 349
rheumatoid synovial membrane, 1099
SLE, predisposition, 394
synovium, schematic representation, 9f
T cell development, abnormalities, 178
vertebrae, development, 8
Human T-lymphotropic virus 1 (HTLV-1),
1424, 1871
infection, nodular synovitis (association),
1872
HuMax-CD20/ofatumumab, 986-987
Humeral epicondyles, ossified enthesopathies,
1684
Humeral head, osteonecrosis (impact), 1694
Humerus, medial/lateral epicondyles
(attachment sites), 565
Humoral immune response, 1407
Humoral immunity, 191
host defense, 1816
Hyaline cartilage
cellular components, 33
nodules, 1959
Hyaluronan, 482
glucuronic acid chain, 482
proteoglycan aggregate formation, 37
synthesis, 4-5
Hyaluronate, matrix component, 145
Hyaluronic acid (HA), 27
biologic functions, 27
derivatives, 1651-1652
intra-articular hyaluronic acid preparations,
usage, 774
preparations, testing (randomized trials),
1652
synovial fluid/cartilage component, 27
synthesis, fibroblast-like synovial lining cells
(usage), 15
WOMAC physical function subscale score,
improvement, 1648
Hyaluronidase, absence, 1624-1625
Hydrocodone, 1019
Hydrodynamic lubrication, 86
Hydrogen peroxide (H₂O₂)
generation, 159
production, 159
Hydromorphine, 1019
administration, 1019
Hydroperoxide reactions, occurrence,
874-875
Hydroperoxyeicosatetraenoic acids (HPETEs),
872
Hydroxyapatite (HA) crystal-associated
calcific bursitis, 1587
Hydroxyapatite (HA) crystals
deposition, 61
deposits, 1577-1578
presence, 757
Hydroxychloroquine (HCQ)
absorption, 934
actions, 933-934
anti-inflammatory effects, 1319
antithrombotic properties, 1299-1300
bioavailability, 934
cardiovascular side effects, 935
cessation, 542
dermatologic side effects, 935
discontinuation, 551
distribution, 954
DMARD, 1140
dosing, 934-935
drug contraindications, 935
drug interactions, 935
elimination, 934
gastrointestinal side effects, 935
geriatric patients, 934
half-life, 934
impact, 551
indications, 934
metabolic side effects, 935
methotrexate/sulfasalazine, combination,
1145
neuromuscular side effects, 935
ophthalmologic side effects, 935
pediatric patients, 934-935
pharmacology, 934
pregnancy usage, 551
retinopathy, AAO screening
recommendations, 935
Immune thrombocytopenia, treatment, 1319

Immune system

Immune regulation

Immune reconstitution inflammatory syndrome

Immune recognition, 140-145

Immune pathways, foreign/pathogenic material (recognition), 414-415

Immune recognition

Immune reconstitution inflammatory syndrome (IRIS), 1861
diagnosis, 1861

HAART, continuation, 1861

immunology, understanding (improvement), 1861

Immune regulation

cytokine effects, 380
duction function, 387

Immune responses, eicosanoids (involvement), 353-354

Immune sentinel cells, 215

Immune system

activation, DC role, 1046

apoptosis, 933

bone marrow component, 1696-1697

bone remodeling, 65
cell death, role, 933
effectiveness, 281
glucocorticoid effects, 899-901

NSAID effects, 886-887

self-tolerance, 127

Immune thrombocytopenia, treatment, 1319

Immune thrombocytopenic purpura (ITP), 1279

Immunity, complement system (usage), 332t

Immunizations

risk factors, 1322

SLE, 1322-1324

Immunodysasays, usage, 821

Immune deficiency virus-associated psoriasis, 1854f

Immunodeficient mice, cancer cells (growth), 515

Immunofluorescence, usage, 792-793

Immunoglobulin (lg)

B cells, 191-194

costant region, 192-194
diversity, generation, 194
gene locus, VDJ recombination, 195f

gene generation, gene segments (impact), 194

isotypes, properties, 193t

structure, comparison, 176

superfamily, 359-360

receptors, interaction/attachment, 219

variable region, 194

Immunoglobulin A (IgA), 193-194

Immunoglobulin D (IgD), 194

Immunoglobulin E (IgE), 194, 234-235

Immunoglobulin G (IgG), 193

Immunoglobulin M (IgM), 192-193

Immunoglobulin-like transcript (ILT)4, 127-128

Immunoglobulin superfamily (IGSF), 148

Immunohistologically stained synovial tissue, quantitative data, 763

Immunologic markers, 476

Immunologic synapses, 181-182

formation, 182-182

inhibition, 276

near full arrest, 270-271

Immunomodulatory cells, glucocorticoids (anti-inflammatory effects), 900t

Immunomodulatory drugs, usage, 396-397

Immunoreceptor tyrosine-based activation motif (ITAM), 104, 140

activation, results, 1546

intracellular signaling machinery, 174-175

sequence, 200

Immunoreceptor tyrosine-based inhibitory motifs (ITIMs), 122, 140, 200

Immunosuppression, opioids (impact), 1022

Immunosuppressive agents, 1141-1142
targets, 949f

Immunosuppressive drugs

action, mechanisms, 942t

classes, 941

Immunomodulatory therapy, 1284

Immunotherapay, impact, 1395-1396

Immunosurstrinucine activation motif (ITAM)-based activation motif, 138

Impingement

sign, elicitation, 642-643

arm, forced forward elevation, 650f

stage II impingement, 651

Immunostaining

HLA-B/B8/DR3/DP52/DQ2 haplotype, presence, 1405

Inclusion body myositis (IBM), 1404, 1415-1416, 1836

contrast, 1415-1416

frequency, 1416

glucocorticoid response, 1426

HLA-B8/DR3/DP52/DQ2 haplotype, presence, 1405

malignancy, relationship, 1940

trichome/hematoxylon-eosin staining, 1416f

Incomplete data, problem, 458-459

Indian hedgehog (Ihh)

examples, 3f

expression, restriction, 6

Individual-level data, usage, 432-433

Indoleamine-2,3-dioxygenase (IDO), T cell exposure, 127-128

Induced (i) Tregs, 127-128

Inducible nitric oxide synthase (iNOS)

expression, increase, 262

nitric oxide production, 1630

osteoarthritic cartilage specimen, immunostaining, 1629f

production stimulation, 52

Induction therapy, 1304, 1315-1317

Industrial Injuries Committee of the American Society for Surgery of the Hand, 494-495

Indwelling catheters, requirement, 1918-1919

Infants

dendritic cell (DC) cytokine production, 550-551

gastrointestinal tract, IgG (impact), 553

TNF inhibitor exposure, 553

in utero TNF inhibitor exposure, transfer, 554f

Infected cells, IgG coating, 124

Infection

tolrambucil, impact, 946

cyclophosphamide, impact, 944

lymph node changes, 276

polyarthritis, relationship, 595

risk factors, 1322

SLE, 1322-1324

infectious agents

impact, 1064-1067

tumors, 1345

infectious arthritis, 725

clinical findings, 725

imaging findings, 725

treatment, 725

infectious diseases

impact, 1129

toecenecrosis, association, 1694

Infectious myopathies, 1424

Infiltrating mononuclear cells, coalescence, 1176

Infiltrative diseases, 613

Infused joints, intra-articular injections, 771

Inflamed synovium, cartiluinated peptides, 1068f

Infused tissue, local cell accumulation, 420

Inflammase, 106, 262

activation, 145

inducers, 123

multiprotein complex, 389

term, usage, 1578

Toll-like receptors, 1064-1065

Inflammation

angiogenesis, 361-362

articular, periarticular (contrast), 590

biomarkers, added value, 479

Inflammatory arthritis, intra-articular injections, 771

inflamed joints, 725

Inflamed synovium, cartiluinated peptides, 1068f

inflamed tissue, local cell accumulation, 420

Inflammasome, 106, 262

activation, 145

inducers, 123

multiprotein complex, 389

term, usage, 1578

Toll-like receptors, 1064-1065

Inflammation

angiogenesis, 361-362

articular, periarticular (contrast), 590

biomarkers, added value, 479
Joints (Continued)
movement, 26-27
limitation, tophi (impact), 1560
neutrophils, ingress, 1582-1583
osteoarthritis (OA), 1038
clinical symptoms/signs, 1631
location, 1618
passive constraint component, 88
passive stabilization, ligaments (role), 12
primary sarcomas, 1965-1968
prosthesis, management algorithm, 1811f
reactive arthritis, signs, 1225
resurfacing, 93-94
secondary malignant tumors, 1968
static constraint component, 88
stiffness, MCTD (relationship), 1441-1442
structures, synovial fluid/nutrition, 14-17
surface lesion, repair, 93f
swelling, 562
impact, 560
management, British Society for Rheumatology/British Orthopaedic Association guidelines, 579f
symptoms, prevalence, 1866t
tissues, COMP expression, 483
total joint replacement, 1642
treatment, principles, 1780-1781
28-joint count, 561
Joint space narrowing
anteroposterior pelvis, impact, 696f
impact, 839-840
Joint stability, 86-89
active muscle contraction, importance, 88
capsulo-ligamentous structures, impact, 1071
negative intra-articular pressure, impact, 88f
passive anatomic structures, importance, 87
rotator cuff, impact, 639
JunB/JunD, BMP-induced transcription factor, 1785t
JunB protein, expression, 1242
JunB/JunD, BMP-induced transcription factor, 1785t
JunB protein, expression, 1242
K
Kainate antagonists, 1031
Kainate-type glutamate receptors, 1026
Kallikreins, 99
Kappa opioid receptor agonists, impact, 1020-1022
Kasabach-Merritt syndrome, 1956
Katz Index, 530
Kawai disease (KD), 610, 1787, 1791-1792
classification, 1791
clinical presentation, 1791-1792
definition, 1791
diagnosis, 1791-1792
diagnostic tests, 1792
epidemiology, 1791
etiology, 1795
outcome, 1784-1785
pathogenesis, 1785
prognosis, 1784
outcome, 1784
pathogenicity, 1784
treatment, 1784
Joint-stabilizing osteoporosis, RA feature, 831
Juvenile dermopathy (JDM) (Continued)
Gordon’s papules
demonstration, 1778f
occurrence, 1777
incidence, 1415
laboratory evaluation, 1778
light microscopic features, 1780-1781
lipidolysophy, 1778
management, 1781
medications, step-wise addition, 1782
methotrexate (MTX), usage, 1781
muscle biopsy, 1780
muscle enzymes, serum levels (elevation), 1780
outcome, 1782-1783
pathognomonic heliotrope rash, 1777
presentation, clinical features, 1777t
therapy (response), evaluation (domains/variables), 1780t
treatment, 1781-1782
vasculopathy, 1778
juvenile gout-complicating GSD, 1558
juvenile idiopathic arthritis (JIA), 620
abatacept, 1575
adalimumab, efficacy, 1575
affected sibling pairs (ASPs), 1743
asymptomatic uveitis, ophthalmologic screening frequency (AAP recommendations), 1758t
biologics, 1575
children, rehabilitation, 1767
conditions, 1741-1742
corticosteroid use, 1755
cyclophosphamide, doses, 943
cytokines, 1758t
geneic/environmental factors, 1742-1743
genetics, 1743-1744
heterogeneity, 1744
histologic arthritis, 1743
ILAR classification criteria, 1752-1754, 1753t
imaging, importance, 1766-1767
inflammatory mediators, 1742t
infliximab, trial, 1756-1757
medications, 1756t
mefloquine indications, 921
usage, 1755
muscle biopsy, 1780
muscle enzymes, serum levels (elevation), 1780
muscle biopsy, 1780
Muscle: connective tissue disease, 1446
muscle inflammation, 1781-1782
classification, 1784
muscle protease, 1781-1782
classification, provisional criteria, 1785t
clinical features, 1783-1784
diagnosis, 1784
clinical features, 1783-1784
diagnostic tests, 1784
epidemiology, 1783
etiology, 1785
outcome, 1784-1785
pathogenesis, 1785
prognosis, 1784
treatment, 1784
Jukta-articular osteoporosis, RA feature, 831
Keratoconjunctivitis sicca (KCS), 1170-1171
Keratocan, 37
Keratinocytes
Keratic precipitates, 619f
Keratinocytes
antimicrobial peptide release, 130
degeneration, 1274
Keratocan, 37
Keratoconjunctivitis sicca (KCS), 1170-1171, 1179-1181
diagnostic test, 1184-1185
occurrence, 1278
SS component, 1121
Keratoderm blenorrhagica, 600
example, 1225f
reactive arthritis/HIV infection, inclusion, 1853f
Kidneys
MCTD, presence, 1444
medulla, urate deposit, 1562f
stones, absence, 1667
systemic sclerosis, 1348
Local steroid injections, 773 complications, 775
Loeys-Dietz syndrome, 1736
Löffler's syndrome, 173
Löfgren's syndrome, 1899-1900
Long bones, 61 components, 62f
diarthrodial joints, connection, 84 formation, 2
cartilage anlagen, impact, 3f
Long interspersed nucleotide elements (LINE)-1, 320
Localized synovial hemangioma, phleboliths
Localized scleroderma (LSc), 607, 1398-1399, 1399f
diagnosis, 1785-1786
abdomen, circumscribed/plaque morphea, 1785f
toxicology occurrence, 928 risk, 923
vagal afferents, 420
Liver: derived antimicrobial peptide (hepcidin), 822
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver-derived antimicrobial peptide (hepcidin), 822
Liquid chromatography-mass spectroscopy (LC-MS/MS) assays, 481
Liquid chromatography-tandem mass spectrometry (LC-MS/MS) assays, 481
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Liver X receptors (LXRs), 352-353
Liver function tests (LFTs), 1776
Matrix metalloproteinase (MMP)
Matrix invasion
Matrix components
Matrix calcification, process (avoidance)
Matrilysins
Matrilins
Maternal morbidity (decrease), interventions
Maternal-fetal HLA class II mismatch
Masticatory muscles (TMJ/myofascial pain)
Mast cells (MCs)
Mast cells (MCs), 233-234
origin, 233-234
appearance, 5
functions, differentiation, 234
granule contents, 235
involvement, 237-238, 240
IgG receptors, impact, 235
mediators, 236-237
production, 235f
morphology, 233f
multiple subsets, 233-234
presence, 239f
pathogen defense, 238
sentinel role, 238
synthesized mediators, 236-237
tissue distribution, 232-234
Masticatory muscles (TMJ/myofascial pain), 233-234
activity, blocking, 1104
chondrocyte synthesis/secretion, 48-49
clavage, 479
domain structures, 103f
family, 220-221
groups, division, 1624
impact, 49
induction, mediation, 1102
involvement, 365
levels, elevation, 49
localization, 49
MMP-generated neoepitopes, 482
MMP-1/MMP-3
regulators, 106
MMP-1/MMP-8/MMP-13, 99-101
MMP-3 serum levels, 483
pericellular docking, 110
presence, 153-154
production, 362, 1464
regulation, 1102
psoriatic arthritis, relationship, 1243-1244
substances, 100t
synthesis, 964
factors, 107f
Matrix metalloproteinase (MMP) (Continued)
upregulation, 43-44
usage, 1102-1103
zymogen (proMMPs), 98
zymogens, activation mechanisms, 108-110
Matrix metalloproteinase 1 (MMP-1)
immunoperoxidase staining, 76f
levels, impact, 483
Matrix metalloproteinase 13 (MMP-13)
expression, 1578
MMP-13-mediated ECM remodeling, 53
Runx2 downstream target, 7
Matrix organization, 34f
Matrix synthesis, alterations, 1625
Mature articular cartilage, 13-14
thinness/homogeneity, 34
Mature B cells, signal transduction, 202
Mature joint, organization/physiology, 8-13
Mayaro virus outbreak, 1869
McArdle’s disease, 1424
McMaster Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR), 466
McMurray compression test, 688-689
McMurray test, 574
MD-2 protein, 260
Measurement content, identification, 466-467
Mechanical injury, cartilage response, 1626-1627
Mechanical LBP, 666
clinical spectrum, 669
Mechanics, 79
Mechanic’s hands, 1414f
Mechano-transduction, impact, 1627
Medial ankle pain, 703
Medial epicondyle, arthrocentesis, 779
Medial ankle pain, 703
Medial meniscus, posterior horn tear (indication), flexion McMurray test
sulfated proteoglycan synthesis capacity, 12
fibrocartilage, intra-articular ossicles (embedding), 1954f
Membranous lupus nephropathy, cyclosporin A (usage), 1311-1312
Medication-induced osteoporosis, 1676-1677
Medsger Systemic Sclerosis Severity Scale, 1610
Medium size vessel vasculitis, 1791-1793
Medium size vessel vasculitis, 1791-1793
mediators, 236-237
involvement, 237-238, 240
inflammation, 345
inhibitors, 236
Matrix metalloproteinase 1/MMP-3
localization, 49
levels, elevation, 49
involvement, 365
induction, mediation, 1102
involvement, 365
levels, elevation, 49
localization, 49
MMP-generated neoepitopes, 482
Mitochondrial aldehyde dehydrogenase, 807-808
Mitochondrial membrane permeabilization (MMP)
initiation, 387
loss, 391
Mitochondrial myopathies, 1424
Mitochondrial stress, 387
Mitogen-activated protein kinases (MAPKs),
activity, 16
mediators, 220f
blocking, 1095
ERK, cascade, 180-181
induction/activation, 1022
inflammatory cytokine induction, 106
mediation, 425
pathways
activation, 318
involvement, 351
serine threonine kinase family, 157
signal mediation, 371
therapeutic targets, 1095
Mitral valve prolapse, 1731-1732
Mixed connective tissue disease (MCTD),
789, 1439-1446, 1787
aseptic meningitis, presence, 1445
blood, impact, 1446
blood vessels, involvement, 1445
clinical features, 1440-1446
computed tomography, 1444f
diagnosis, 1440
support, 1279
diagnostic criteria, 1441t
disorder, 1274
erosive arthritis feature, 1274
features, simultaneous presence
(implication), 1440-1441
fever, 1441
gastrointestinal system, involvement,
1444-1445
hand, example, 1442f
heart, involvement, 1443
joint pain/stiffness, 1441-1442
juvenile MCTD, 1446
kidneys, involvement, 1444
lungs, involvement, 1444
magnetic resonance images, 1442f
mucous membranes, appearance, 1442
muscles
appearance, 1442-1443
biopsy, 1443f
nail-fold capillaroscopy, presence, 1437f
nervous system, involvement, 1445
overlap syndromes, 800
pregnancy, 1446
serologic features, 1440
skin, appearance, 1442
symptoms, 1440-1441
vasculopathy, features, 1444f
Mixed cryoglobulinemia, 165-166
confluent purpura, 1514f
presentation, 609
mKatG antibodies (measurement),
recombinant mKatG protein (usage),
1899
Moderate-dose GC, short course (impact),
1285
Modification, effect (measure), 437-438
Modified HAQ (MHAQ), 530
Modified Rodnan skin score
usage, 1370
visual analogue score, combination,
1370
Modified Steinberg staging systems, 1694t
Molecular biomarkers, laboratory testing,
1640
Molecular markers
classification, 476
impact, 478t
Molecular mimicry, 187, 208-209, 1066
cross-reactivity, 208
impact, 295
mechanism, 292
Moment (kinetics), 82-83
arm, 82-83
force, 83f
generation, muscle contraction (usage),
88-89
Mono-ADP ribosylated histone, 319
Monoamine oxidase (MAO) inhibitors, usage,
1019
Monoarthritis
acute calcific periartthritis, 581
calcium phosphate crystal arthritis, 581
cholesterol crystal arthritis, 581
crystal arthritis, 581
evaluation, synovial histopathology (usage),
760
impact, 583
Lyme disease, 582
organisms, impact, 582
pigmented villonodular synovitis, impact,
plant thorn synovitis, 582
reactive arthritis, 581-582
Monoarthropathies, joint diseases (impact),
583
Monoarticular arthritis, candidiasis (impact),
1844
Monoarticular pain, periarticular disorders
(impact), 584-585
Monoocyte activation, 378
Monoocyte chemoattractant protein (MCP), 53
expression, 53
Monoocyte chemoattractant protein 1
(MCP-1), 1410-1411
Monoocyte-derived dendritic cells (moDCs),
120
Monoocyte-derived dendritic cells (moDCs),
120
Monoocyte-derived dendritic cells (moDCs),
120
Monozygotic twins
epiphenomes, differences, 319
relative risk (RR), measurement, 310t
Morning after contraception, 1338
Morning stiffness, 560
inflammatory arthritis sign, 1110
noninflammatory joint diseases, association,
560
Morphea, 606-607
presence, 606-607
Morphine, 1017
codeine conversion, 1019
measurement, 1017
Morphine-3-glucuronide (M3G), 1017
Morphine-6-glucuronide (M6G), 1017
Morquio’s disease (short-trunked dwarfism),
1719
Mortality, asymptomatic coronary heart disease
(impact), 535
Morton’s neuroma, injection, 784f
Mother/newborn, mucosal immunity
(integration), 1257f
Motion, limitation, 560
observation, 562
S-T-L system, 563
Motoneuron diseases, 1423
Motor examination, strength grading, 635t
Mucosal immunity, integration, 1257f
Mucous membranes, appearance, 1287
Mucosal ulcers, 1287
Mucosal T-independent class switch
recognition, 273-275
Mucosal surfaces, antigen detection, 273-275
Mucosal surface (immune exclusion), secretory
properties, 137t
migration/fate, 138
integrins/receptors, usage, 138
distribution, 134, 136f
differentiation, 136f
biology, knowledge (increase), 134
activation phenotype, modulation, 135
immune dysregulation, 135
activation, 318
reduction, 877
Mucosal epithelial cells, 1253f
Mucosal defense, integrity, 1252
Mucosal barrier, anatomy, 1253f
Mucosa-draining lymphoid nodes, induction,
1259
Mucosal lymphoid tissue (MALT), 199
Helicobacter pylori (association), 520
lymphoma, 1184
lymphoproliferative disorder, presence,
1940
small bowel location, 930
Mucosa-draining lymphoid nodes, induction,
199
Mucosal barrier, anatomy, 1253f
Mucosal defense, integrity, 882
Mucosal immunity, integration, 1257f
Mucosal mast cells (MMCs), 233-234
Mucosal permeability/transport, assessment,
1252
Mucosal surface (immune exclusion), secretory
antibodies (provision), 1256f
Mucosal surfaces, antigen detection, 273-275
Mucosal T-independent class switch
recombination (CSR), 206
Mucosal ulcers, 1287
Mucous membranes (xerostomia), occurrence,
1394
Mucus-secreting goblet cells, 1251
Mucocutaneous lymph node syndrome (MRH), 609,
614, 1938
diagnosis, 762-763
rarity, 1132-1133
INDEX

Myeloid APCs, germline-encoded receptors, 140
Myeloid cell antigens (EGF-TM7 family), 137f
Myeloid DCs, 118
Myeloid dendritic cells, 1175-1176
Myeloid inhibitory C-type lectin-like receptor (MICL), 122
Myeloma receptor antibody (MRA), 971
Myelopathy, 629-630
clinical presentation, 629-630
diagnosis, provocative tests (usage), 635
exacerbation, 630
spinal cord compression, 632
Myeloperoxidase (MPO) antibodies, impact, 1504
identification, 1481
MPO-ANCA, detection, 1490-1491
MPO-specific antineutrophilic cytoplasmic antibodies (ANCAs), report, 1390
presence, 153-154
Myeloproliferative diseases, 1937
Myeloproliferative disorders, 1946-1947
Myelopathy, 629-630
Myeloma receptor antibody (MRA), 971
Myeloid inhibitory C-type lectin-like receptor
Myeloid DCs, 118
Myeloid APCs, germline-encoded receptors,
i52 isoform, 68
class II-containing compartment (MIIC), 125
Myelopathy, 629-630
clinical presentation, 629-630
diagnosis, provocative tests (usage), 635
exacerbation, 630
spinal cord compression, 632
Myeloperoxidase (MPO) antibodies, impact, 1504
identification, 1481
MPO-ANCA, detection, 1490-1491
MPO-specific antineutrophilic cytoplasmic antibodies (ANCAs), report, 1390
presence, 153-154
Myeloproliferative diseases, 1937
Myeloproliferative disorders, 1946-1947
Myelopathy, 629-630
Myeloma receptor antibody (MRA), 971
Myeloid inhibitory C-type lectin-like receptor
Myeloid DCs, 118
Myeloid APCs, germline-encoded receptors,
i52 isoform, 68
class II-containing compartment (MIIC), 125
Myelopathy, 629-630
clinical presentation, 629-630
diagnosis, provocative tests (usage), 635
exacerbation, 630
spinal cord compression, 632
Myeloperoxidase (MPO) antibodies, impact, 1504
identification, 1481
MPO-ANCA, detection, 1490-1491
MPO-specific antineutrophilic cytoplasmic antibodies (ANCAs), report, 1390
presence, 153-154
Myeloproliferative diseases, 1937
Myeloproliferative disorders, 1946-1947
Myelopathy, 629-630
Myeloma receptor antibody (MRA), 971
Myeloid inhibitory C-type lectin-like receptor
Myeloid DCs, 118
Myeloid APCs, germline-encoded receptors,
i52 isoform, 68
class II-containing compartment (MIIC), 125

Neural tube defects, folic acid deficiency (impact), 551
Neuroborreliosis, cranial/spinal MRI findings, 1822
Neurodiagnostic tests, 709
Neuroendocrine systemic response, 418-420
Neurogenic claudication (pseudoclaudication), 1078
Neurogenic inflammation, 238
Neuropeptides
Neurotaxins, 1931
Neuropathic arthritis, 1931
Neuropathic pain, 1014-1015
Neuronal enzymes, immunoreactivity, 7-8
Neuronal innervation, changes, 420-422
Neuronal reflexes, role, 414
Neurons, myelination, 62
Neuronal reflexes, role, 414
Neurons, myelination, 62
Neuronal apoptosis inhibitory protein (NAIP), 388-389
Neuronal repulsive factors, production, 420-421
Neuronal muscular junction, scanning electron micrograph, 69f
Neurotechnological systems, development, 424
Neurotechnological systems, development, 424
Neurovascular system, evaluation, 695-696
Neutral hydrolases, impact, 170-171
Neutralizing antibodies, 166-167
Neutrophils
Neutrophilic vascular reaction, representation, 1142
Neutrophilic dermatoses, 166
Neutrophilic dermatoses, 166
Neutrophil extracellular traps (NETs), 160
Newborn/mother, mucosal immunity, 64f
Newton, Isaac
New second Law, 82-84
Newtonian mechanics, reliance, 79
N-linked oligosaccharides, 37
Nicotinamide, 156-157
Nicotinamide adenine dinucleotide (NAD\(^+\)), necessity, 74
Nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, 159
activation, 1547
component, 159
electron transfer, 159
system, assembly, 160f
upregulation, ability, 262
Nicotinamide adenine dinucleotide phosphate oxidase system, reduction, 159-160
Nicotinamide acetylated receptors, junctional
Night splint, usage, 710f
Nilotinib, usage, 1396
NITEGE antibodies, recognition ability, 56
Nitric oxide (NO)
effects, 1630
generation, 159
mediator role, 53
NO-induced breakdown, 396
production, 1097
iNOS, usage, 1630
synthase
activity, inhibition, 877
osteonectosis, relationship, 1699
upregulation, 52
Nitrogen, RA impact, 1097
NK cell activation, 127
N-linked oligosaccharides, 37
NLR family apoptosis inhibitory protein 5 (NAIP5), 126
NLRP3
inflammasome, 1608
activation, 1547f
inflammation, 1546-1547
mutations, effect, 1608-1609
NOD-LRR protein member, 1608
role, 1608
Osteoarthritic articular cartilage, histologic sections, 1621f
Osteoarthritic cartilage
CXC/CC chemokines production, increase, 1629
specimen, immunostaining, 1629f
Osteoarthritic knee, views, 690f
Osteoarthritic cartilage
Osteoarthritic articular cartilage, histologic disease diagnostic/monitoring/prognostication usage, 1639
diagnostic approach, 1639
diagnosis, 723
development degenerative joint disease, 1617
cytokines, impact, 1630-1631
corticosteroids, 1651
computed tomography, 843-844
clinical symptoms/signs, 1631
clinical/radiographic features, 1639f
clinical findings, 723
clinical features, 1637f
clinical/radiographic features, 1639f
clinical symptoms/signs, 1631
clinical trials, extended-release narcotic analogues, 1651
computed tomography, 843-844
diagnostic/prognostication usage, 844
corticosteroids, 1651
costs
estimation, 448-449
studies, 448
cytokines, impact, 1628-1629
degenerative joint disease, 1617
development
retrospective studies, 496
risk factors, 476, 496e
weight-bearing exercise, impact, 499
diaceetrin, synthesis inhibition, 1655-1656
diagnosis, 723
MRI, usage, 837-838
ultrasonography, usage, 863
diagnostic approach, 1639
diagnostic/prognostication usage, 841
diagnostic testing, 1639-1641
disease mechanisms, 1625-1627
progression, inflammatory mediators (role), 1628-1633
Osteoarthritis (OA) (Continued)
disease-modifying drugs, 1655f
disease-modifying therapies, 1654-1656
DISH peripheral joint manifestations, comparison, 1682
eye rheumatoid arthritis, differentiation factors, 1130t
economic impact, underestimation, 448
epidemiologic studies, 495
epidemiology, 1636-1637
estrogen replacement therapy (ERT), 1620
etiologic factors, 1618-1620
examination, studies, 499
exercise impact, 1648
relationship, 496-497
extracellular matrix, changes, 1625
Framingham Study, 1618-1619
F-spondin, impact, 1631
functional assessment, 1641-1642
gender, 1620
gene expression, 1627
generalized osteoarthritis (GOA), 1639
gene therapy, 1656
gene predisposition, 1618-1619
genomic technology, 1619
glenohumeral joint, 657
glucosamine, usage, 1652-1653
hands, joint-specific symptoms/signs, 1638
hips
injections, 772-773
joint-specific symptoms/signs, 1638
hyaluronic acid derivatives, 1651-1652
impact, 1630
hypertrophic changes, 1620-1621
hypocellularity, 1621
imaging, advanced modalities, 1641
impaired/full-lengthing, 335-356
incidence, increase, 497-498
inherited forms, cause, 1619
intervention methods, 1656
intra-articular agents, 1651-1652
joints
location, 1618
malalignment/tumor, 1619-1620
knee
diagnostic/prognostication, 853-856
MRLs, 854f
x-rays, 840f, 854f
laboratory testing, 1640
magnetic resonance imaging (MRI), 853-856
examples, 845f
malalignment, 1638
management, 1646-1656
algorithm, 1647f
initiation, 1646
interactions, 335-356
nonselective NSAIDs, usage, 1650
mechanical injury, cartilage response, 1626-1627
mechanotransduction, 1627
monitoring
MRI, usage, 855-856
ultrasonography, usage, 863
morphologic changes, 1620
mortality, 1642
narcotic analogues, 1650-1651
National Health and Nutrition Examination Survey (NHANES III), 1620
Osteoarthritis (OA) (Continued)
non-narcotic analgesics, usage, 1650
nonpharmacologic interventions, 1646-1649
nonpharmacologic management, 1647t
nonsteroidal anti-inflammatory drugs (NSAIDs), 1650
nurtaceuticals, 1652-1654
last, 1652t
nutrition, relationship, 1050-1051, 1055
OA Biomarkers Network, funding, 1632
OA-derived chondrocytes, 396
obesity, impact, 1618
occupational physical activity, association, 496t
occupation-related rheumatic diseases, 495
onset, subchondral sclerosis (impact), 1620
orthotics, impact, 1648
osteyphye formation, 1621
pain, 657
response, 1650
source, 1646
pathogenesis
biomechanical theories, 1625-1626
events, 51
pathogenic mechanisms, schematic, 1623f
patient assessment, 1646
pentosan polysulfate (Cartrofen), administration, 1655
performance-based assessment, 1641-1642
pharmacologic interventions, 1649-1652
polyarticular osteoarthritis, 1639
polyarticular scenarios, contrast, 588-591
postgenomic technology, 1619
predisposition factors, 1632f
premature chondrocyte senescence, 1625
presentation, 583
prognostication
MRI, usage, 856
ultrasonography, usage, 863
progression
anabolic pathways, dysregulation, 1627
risk factors, 476
proliferative changes, 1620-1621
prostaglandins, impact, 1630-1631
proteasmes, impact, 1629-1630
psychosocial interventions, 1646-1647
radiographic OA, prevalence, 1636
regenerative tissue, transformation, 1627
reparative changes, 1620-1621
running
relationship, 498-499
studies, 498t
secondary OA, etiologies, 1637t
serum nitrate/nitrite, increase, 479
severity (reduction), doxycycline (usage), 1655
skeleton studies, 495
spa therapy, 1649
sports participation, 497t
structure-modifying drugs, 155f
structure-modifying therapies, 1654-1656
subchondral bone marrow lesions, 1642f
subchondral bone sclerosis, 1628
surgical intervention, 1656
susceptibility gene, genome-wide linkage
list, 1652f
symptomatic OA, prevalence, 1636
symptomatc OA, prevalence, 1636
symptoms/signs, 579t, 1637-1638
synovial fluids, high-abundance molecules, 16
synovial tissue, alterations, 1631-1632
systemic agents, 1650-1651
targeting, 838-839
temperature modalities, 1648
Osteoarthritis (OA) (Continued)

- TGF-β, impact, 1630
  - TMJ, involvement, 721
  - topical agents, usage, 1649-1650
  - topical NSAID preparations, usage, 1649-1650
  - tramadol, usage, 1651
  - transdermal fentanyl, usage, 1651
  - treatment, 723

- paradigms, 1646
  - type I collagen concentration, 1621-1622
  - ultrasonography, 862-863

- diagnostic/monitoring/prognostication usage, 863
  - water-sensitive techniques, 854

- osteopenia, 1854
  - age-related bone loss, pathophysiology, 1660-1661
  - OA abnormality, 840
  - paucity, 862-863

- osteoporosis, 1660-1671, 1854
  - absence, 835-836
  - age-related bone loss, pathophysiology, 1661-1664
  - ankylosing spondylitis, relationship, 1209
  - antiresorption, 1670
  - bisphosphonates, usage, 1668-1670
  - calcitonin, usage, 1668
  - calcium, treatment, 1667
  - characterization, 1660
  - clinical signs, 1660-1661
  - COLIA1, 1661-1662
  - denosumab (Prolia), usage, 1670
  - disorder, 396
  - epidemiology, 1660-1661
  - estrogen, usage, 1667-1668

- Osteogenesis imperfecta (OI) (Continued)
  - type VII, 1730
  - variants, skeletal differences (radiographs), 1724f

- Osteopenia, 1854
  - secondary effect, 1929

- Osteopontin
  - sensitivity/specificity (comparison), 1705t

- Osteonecrosis (Continued)
  - musculoskeletal conditions, impact, 1693
  - nitric oxide synthase, relationship, 1699
  - nonsurgical (conservative) approaches, 1707-1708
  - nontraumatic osteonecrosis, 1695-1696
  - osteoimmunology, 1696-1697
  - outcome, 1708-1709
  - oxidative stress, relationship, 1699
  - pathogenesis, 1695-1700
  - genetics/environment, impact, 1699

- Osteomalacia, 1672-1674, 1938-1939
  - causes, 1672t
  - pain, elicitation, 1672-1673
  - phosphate deficiency, association, 1673

- Osteomyelitis, 1859
  - diagnosis/management, PET (usage), 865

- osteoarticular mycotic infections, drug treatment, 1846

- Osteoarticular tuberculosis, diagnosis, 1832

- Osteoblasts, 62
  - function, alterations, 1698
  - migration/differentiation, 7

- Osteoclasts
  - balance, 1697-1698
  - interactions, 64-65

- Osteocalcin
  - levels, circulation, 1675

- Osteochondral defect (OCD), 701

- Osteoclast-associated immunoglobulin-like receptor (OSCAR), 63-64

- Osteoclast function (inhibition), salmon/
  - human calcitonin (impact), 1675

- Osteoclastogenesis-inhibitory factor (OCIF), impact, 396

- Osteoarticular surgeries, 657-658
  - activity, inhibition, 396

- Osteopenia, 1854
  - secondary effect, 1929

- Osteopenia, 1854
  - secondary effect, 1929

- Osteophytes
  - anteroposterior pelvis (impact), 696f
  - formation, 1621, 1627-1628

- OA abnormality, 840

- OA, OA abnormality, 840

- Osteoporosis, 1660-1671, 1854
  - absence, 835-836

- Osteoporosis, 1660-1671, 1854
  - age-related bone loss, pathophysiology, 1661-1664

- Osteoporosis, 1660-1671, 1854
  - ankylosing spondylitis, relationship, 1209

- Antiresorption, 1670
  - bisphosphonates, usage, 1668-1670

- Calcitonin, usage, 1668

- Calcium, treatment, 1667

- Characterization, 1660

- Clinical signs, 1660-1661

- COLIA1, 1661-1662

- Denosumab (Prolia), usage, 1670

- Disorder, 396

- Epidemiology, 1660-1661

- Estrogen, usage, 1667-1668

- Glucocorticoid-induced osteoporosis, 1671-1672
INDEX
Platelets (Continued)
function, 247
inhibition, pharmacologic agents (usage), 250-251
granule compounds, 246
involvement, 248
life span, 246
membrane components, 246
glycoprotein receptors, presence, 245
microparticles, 247, 1082
elevation, clinical disorders (association), 247
numbers, alterations, 249-250
plasma membrane, 245
plugs
anatomy, electron micrograph, 247
formation, 247
rheumatic diseases, relationship, 249-250
role, 248-250
Pleural disease, RA (relationship), 1123-1124
Pleuritis, 1202
Plica syndrome, 573
Pneumocystis jiroveci infection, 1323-1324
Pneumocystis jiroveci pneumonia (PCP), cause, 1357
incidence, 1253-1264
PN-2, Kunitz-type inhibitor, 104-105
POEMS (polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes) syndrome, 608
Polarized immune responses, 1355
Poly ADP ribose polymerase, 389
Polymyxin B, 1396
Polyarticular septic arthritis, 1803
Polyarthritis
causes, 594
complement activation, 336-337
defining, 587-588
development, antagonism, 249
diagnosis, 593
discriminating features, 1127
drug-induced syndromes, 595
elderly, 596
evaluation, synovial histopathology (usage), 760-763
fever, impact, 595
infection, impact, 595
laboratory tests, 591-593
malignancy-related polyarthritis, 595
outcome, 596
pediatric polyarthritis, 595
presentation, acute monoarthritis (impact), 582-583
prognosis, 584
radiography, 593
radiologic studies, 591-593
rash, impact, 595
rheumatoid factor positive, 1747
serologies, 591-592
treatment, 596
Polyarticular arthritis diagnosis, 588-591
epidemiology, 587
fusiform swelling, 1113
societal impact, 587
societal complaints, differential diagnoses, 587
Polyarticular-course juvenile idiopathic arthritis treatment algorithm, 1755
Polyarticular disease, radiographic changes, 593
Polyarticular osteoarthritis, 1639
Polyarticular process, history, 588-590
Polyarticular scenarios, 587
arthritis, arthralgia (contrast), 590
articulart, periarthralgia (contrast), 590
demographics, 588
disease course, 589
examination, assessment (algorithm), 589
extra-articular manifestations, 590-591
frequency, 588-591
history, 588-590
assessment, algorithm, 589
inflammation, presence, 589-590
joint involvement, pattern, 589
physical examination, 590-591
symptom onset, 588-589
Polyarticular septic arthritis, 1803
rarity, 1802-1803
Polyarticular symmetric joint involvement, 534
Polyclonally activated T cells, 1755
Polyclonal antibodies, usage, 328
Polyclonal IgM, 1279
Polyclonal IgG, 1279
Polyethylene glycol (PEG) pegloticase, association, 1009
PEGylation, impact, 963
Polyethylene glycol (PEG) moiety, 958
INDEX
Primary angiitis of the central nervous system (PACNS) (Continued)
reversible cerebral vasoconstriction syndrome (RCVS), 1521
subclassification, attempts, 1519
treatment, 1523
guidance, 1523
Primary angiography-negative childhood CNS vasculitis, 1797f
Primary bone tumors, 1945t
Primary care physician (PCP)
rheumatologist, interaction, 1155-1156
time, 1155f
Primary Cushing's disease, cause, 1929
Primary eosinophilic syndromes, 173
Primary fibromyalgia, 739
Primary fibronectin receptor (α5β1), cartilage degradation, 43-44
Primary gout, 1563-1564
Primary hyperparathyroidism, joint manifestation, 1928
Primary lymphoid tissues, 271-272
Primary lymosomes, fusion, 145
Primary malignant bones
Primary malignant disease, 1177-1179
Primary metastatic brain tumors, 1944-1945
manifestation, 1944-1945
Primary necrotizing vasculitis (skin), 608-610
Primary neuroendocrine dysregulation, 742
Primary ossification center, 1719
Primary osteoarthritis, 1617-1618, 1637
Primary osteoarthropathy, 742
Primary pulmonary hypertension, 1859
Primary sclerosing cholangitis (PSC), development, 1260
Primary Sjögren's disease, autoimmune disease, 281-282
Primary Sjögren's syndrome, 1173-1174
associated diseases, 1184
biologic therapies, 1188
blood levels, 1185
clinical diagnosis, 1185-1186
diagnosis, labial salivary gland biopsy (usage), 1185
differential diagnosis, 1179-1180t, 1186
disease mechanisms, 1188
features, 1176-1177
gastrointestinal symptoms, frequency (increase), 1183
gene susceptibility, 1174
glottal epithelial cell, appearance, 1177-1179
labial salivary gland, histopathology, 1176f
neurologic abnormalities, 1183
NHL complication, 1184
presence, 1182f
recurrent lower extremity purpura, occurrence, 1188
renal disease, clinical significance, 1183
therapy, 1188
treatment, 1186-1188
Primary systemic vasculitides, 1521-1522
blood vessel involvement, 1481
Primary tumors, 1944-1946
occurrence, 693
Primary urate underexcretion, 1541
Primary urate overproduction, 1537-1538
Primary urate underproduction, 1523
PREMIER trial, 1152
results, 1153f
Priority Medicine for Europe and the World (World Health Organization), 476
Raynaud’s phenomenon (RP) (Continued)
extraglandular manifestations, 1169-1170
lupus nonspecific skin lesions, 605
pathogenesis, endothelin-1 (implication), 1377
scleroderma, association, 1376f
severity, 1336
skin involvement, 1377-1380
treatment, 1374-1377
drugs, impact, 1375
Reactive AA amyloidosis, 1607
Reactive arthritis (ReA), 581-582, 600-601, 1251
antibiotic treatment, 1228
response, 1199
bacteria, detection, 1226
biologies, usage, 1227
blood tests, 1226
classification criteria, 1222-1224
clinical features, 1224-1227
cytokines, 1198-1199
definitions, 1221-1222
diagnosis, 600, 1224-1227
differential diagnosis, 1226
disease-modifying drugs, usage, 1227
DMARDs, usage, 1227
ERA categorization, 1762-1763
history, 1224-1225
HLA-B27 testing, 1227
imaging, 1227
impact, 1853f
incidence, 1224
investigations, 1226-1227
joint monoarthritis, 581-582
keratoderma blennorrhagicum, combination, 600f
microbiology, 1226-1227
organisms, association, 1222t
outcome, 1228
pathogenesis, 1197-1199
prognosis, 1228
questions/research, 1229
serology, usage, 1226-1227
signs, 1225
SSZ/sulfapyridine, usage, 931
susceptibility (induction), HLA-B27 (role), 1229
symptoms, control, 1227
synovial fluid, tests, 1226
term, usage, 1221
treatment, 1227-1228
antibiotics, role, 1229
undifferentiated spondyloarthritides, relationship, 1229
usage, 1199
uveitis, association, 619
Reactive hyperemia, impact, 1373
Reactive oxygen, impact, 1097
Reactive oxygen intermediates (ROIs), 387
Reactive oxygen species (ROS) generation, 1350-1352
impact, 1046
oxidative stress, relationship, 1354
production, increase, 1097
release, 249
Reactive thrombocytosis, occurrence, 818
Reading levels
impact, 1038-1039
Internet, impact, 1039
psychosocial factors, 1039
Real outcome, 455
Recall bias, 437
occurrence, 434
Receiving operator curve (ROC) analysis, 1171-1172
Receptor-activated SMAD (R-SMAD), 48
phosphorylation, prevention, 48
Receptor activator of nuclear factor κB (NfκB) ligand (RANKL), 14, 129
cross-talk, 62
expression, 65
inhibitors, 1670
pathway, importance, 1046
production, 1661
Receptor antagonists, soluble receptors
(production), 91
Receptor for advanced glycation end products (RAGE) multiligand RAGE, mediation, 1582
products, 145
signaling, 1582
Receptors BAFF family, 198f
detecting, 226
expression, human articular chondrocytes capability, 53
upregulation, 120-121
usage, 53e, 138
Recession, occurrence, 446
Recombinant human PTH (rhPTH) composition, 1670
effectiveness, evaluation, 1670
Recombinant mKatG protein, usage, 1899
Recombination-Activating Gene 1 (RAG-1), 194
Recombination-Activating Gene 2 (RAG-2), 194
Recombination signal sequences (RSS), sequence, 194
Resident mast cells, 1078
Repertoire selection, 206-207
Repertoire censoring, 197
Regional arthrocentesis techniques, 777-784
Regional blood flow, arteriole regulation, 10
Regional Examination of the Musculoskeletal System (REMS), usage, 577
Regional occupation-related musculoskeletal syndromes, literature, 494t
Regional pain syndromes, ultrasonography, 857f
Registries, 435
Regression toward the mean, 453-454
Regulated on activation normal T cell expressed and secreted (RANTES), 53, 121, 171, 361
blocking, 1091
expression, 1096
usage, 1091
Regulators of G protein signaling (RGS), involvement, 413
Regulatory B cells, clonal expansion, 422
Regulatory T cells (Tregs), 1086-1087
clonal expansion, 422
function, abnormality, 1086
impact, 276
induction, 128
subpopulations, 183
Rehabilitation assessment tools, 529-530
cycle, 529-530
definitions, 531f-532t
descriptions, 531t-532t
goals, 528-529
health professionals, assessment methods, 529
history, 528
principles, 533
problem identification, 529
purposes, 531t-532t
self-management programs, 533
team care models, 528-529
team members, goals, 528-529
Rehabilitation intervention algorithm, 534f
osteoarthritis, 535-536
patient participation, requirement, 536
systemic lupus erythematosus, 535
Rehabilitation management framework, 529
guideline, suggestions, 530
patient education, importance, 533
Reiter’s syndrome, 1221
Relapsing polychondritis (RPC), 584, 612
edema, 1938
incidence, 1938
maternal complication, 1713
impotence, 1938
renal disease, 1713-1714
clinical significance, 1183
renal flares
importance, 1322
treatment, 1317-1318
Renal function changes, 1295
measurement, 1289
Renal impairment, cyclophosphamide (impact), 943
Renal insufficiency, 1542
Takayasu’s arteritis, maternal complication, 547
Renal-limited pauci-immune necrotizing and crescentic glomerulonephritis (RLV), 1481
Renal thrombotic microangiopathy, 1336f
Renal transplantation, 1324-1325
performing, 1391
Renal tubular acidosis (RTA), 1183
Renal urate excretion, plasma uric acid (contrast), 1541f
Renal urate excretion age/gender, 1541-1542
secondary causes, 1541-1543
Renal uric acid resorption, OATs (involvement), 1540
Renotubular phosphate reabsorption, disorders (occurrence), 1673-1674
Repertoire censoring, 197
Repertoire postantigen activation, 205
Repertoire selection, 206-207
tolerance, 206
Repulsin-related proteinases, aggrecan degradation mediator, 49
Resident cells, properties, 136
Resident macrophages, location, 136
Resident mast cells, 1078
Resistance testing, motion (performing), 634
Resistance training, impact, 76
Rheumatic diseases (Continued)
cytokines, roles, 377t
distribution, 1852t
eosinophil, relevance, 172
epigenetic changes, study, 320
epigenetics, relationship, 320-322
etiology, nutrition (impact), 1048-1051
family relationships, 1039
fatigue, conversion, 1035
febuxostat role/indications, 1005-1006
fibroblasts, 223
generic contribution, size (estimation), 310
 genetics, challenges, 312-313
growth factors, relevance, 377t
HLA associations, 304t
human leukocyte antigen (HLA), 303-309
illness, 1034
IL-1 superfamily cytokines, roles, 374t
imaging modalities, technical aspects, 831
joint swelling, impact, 560
leflunomide, indications, 927-928
lymphoproliferative disorders, development (risks), 1940-1944
management
acute phase reactants, usage, 823-825
nutrition, usage, 1051-1055
pain, 1034
pathogenesis
mast cells, impact, 232
platelets, role, 250
pegloticase, role/indications, 1009
phagocytes, overview, 134-135
platelets
numbers, alterations, 249-250
relationship, 249-250
psychosocial adaptation, landmarks, 1035
radiography, 830-841
relevance, 149
response, 1860-1861
rituximab, usage, 987
skin lesions, diagnosis, 599
standardized incidence ratio (SIR), 1942
TNF superfamily cytokines, role, 374t
treatment
nontuberculous mycobacterial infection, emergence, 1838
tuberculosis, emergence, 1833-1834
uricosurics, role/indications, 1007
Rheumatic disorders
apoptosis, relationship, 393-397
musculoskeletal/rheumatic disorders, overuse (association), 500t-501t
tuberculosis, treatment (impact), 1834
Rheumatic fever (RF)
diagnosis, World Health Organization Criteria, 1881t
D8/17 marker, 1878t
heart-reactive antibody titers/laboratory data, 1888t
rarity, 1132
Rheumatic heart disease (RHD)
HLA-DR frequency, 1878
incidence, 1882
World Health Organization criteria, 1881t
Rheumatic pneumonia, presence, 1884
Rheumatoid arthritis (RA), 33, 117, 129-130, 321-322, 1941-1943
abatacept, usage, 989-995
ACR/EULAR classification criteria (2010), 1127t
acute phase response, 823-824
adalimumab, 961-962
adaptive immunity, role, 1059-1061
alcohol consumption, impact, 1049
anakinra, impact, 969
angiogenesis, 1098-1100
angiogenic factors, 361t
angiostatic factors, 361t
animal models
advantages, 400
usage, 1968
ankle pain, 702-703
ankylosing spondylitis, 842-843
anti-CCP antibodies, appearance (kinetics), 807
anti-G6PI antibodies, presence, 811
antikeratin antibodies, presence, 805
antinuclear antibodies (ANAs), impact, 577-578
antiperinuclear factor, presence, 805
anti-TNF treatment, meta-analyses/cohort studies, 522t
Anti-TNF Trial in Rheumatoid Arthritis with Concomitant Therapy (ATTRACT) trial, 943
apoptosis, paucity, 1097
arthrodesis, 705-707
arthroplasty, 705-706
articular cartilage, structure, 110f
atherosclerosis, factors/pathobiologic mechanisms, 1133-1134
autoantibodies
presence, 813-814
production, 1067f
role, 814
specifications, 811-813
autoantigens, 1069f
citrullination, 810-811
citrullinated, 810-811
common autoimmune diseases, 92-93, 281-282
autoimmune diseases, responses, 1070
axial spondyloarthritides, 842-843
baseline CTs-II levels, increase, 481
B cell depletion, 1077
bilateral protrusion acetabuli, 1117f
blood hyaluronan levels, elevation, 482
bone loss, types, 865
bone resorption, 112
cancer, risk (increase), 516
Cancer Registry data, 517
cardiac disease, forms, 1124
cardiovascular comorbidity, 506
cardiovascular mortality, 505
cardiovascular risk factors, impact, 509
cartilage
damage, assessment, 35
ultrasonography, 859f
cartilage destruction, 48, 1101
proteinases, impact, 111-112
case-control studies, meta-analysis, 509
CD28, T cells (increase), 181
cells, accumulation, 1097
cervical spine involvement, 832
x-ray, 833f
cataracts, 125-126
chemokine-directed approaches, 1091
chronic polyarthritis, resemblance, 1129
chronic proliferative synovitis, 111
chronic systemic inflammatory rheumatologic disorder, 534
Secondary osteoarthritis, 1617-1618, 1637
etiologies, 1637
Secondary Sjögren's syndrome, prevalence, 1173
Secondary urate overproduction, hyperuricemia (relationship), 1538-1539
Second-line drugs, usage, 1216-1217
Second messenger, 156
Secreted frizzled related protein 1 (SFRP1), 64-65
Secreted phospholipase (sPLC), 341-342
Secreted-type matrix metalloproteinases (secreted-type MMPs), 99-103
composition, 99
domain structures, 103f
Secreted-type proMMPs, activation, 109f
Secretomat, 94
Secretory antibodies (SigA/SigM) function, provision, 1256f
Secretory component (SC), 1252-1253
Secretory PLA, (sPLA), disulfide cross-linked proteins (presence), 340
Segmented filamentous bacteria (SFB), interactions, 1256f
Selectin-mediated tethering, 269
Selectins
adhesion, 1100
definition, 269t
encoding, 1891
molecular signatures, 157f
Selection bias, 437
Selection processes, net result, 459
Selective estrogen receptor modulators (SERMs), 1668
Selective serotoner receptor inhibitors (SSRIs), 1020
Selenium
concentration levels, 1049
management role, 1053-1054
Selenocysteine, presence, 798
Self-antigens
generation, 1271
pathologic autoantibodies, binding, 292-293
requirements, 291
Self-debris, alteration, 332
Self-efficacy (effective consumer), 465
Self-management programs, 533
Self-MHC molecules, T cell affinity, 127
Self-molecule function, blocking/enhancement, 292-293
Self-nonself discrimination, mechanisms, 286
Self-peptide/MHC complexes, response (homeostatic proliferation), 178-178
Self-reactivity, repression, 288
Self-tolerance, 127
mechanisms, initiation, 271-272
Semiquantitative radiographic assessment, examples, 1641f
Senessence, term (usage), 1625
Sensomotor interventions, 536
Sensory nerve fibers, sprouting, 422-422, 425-426
illustration, 421f
Septic arthritis, impact, 577
Septic arthritis, 658, 838, 1832-1833, 1916
anaerobes, percentage, 1802
antibiotic treatment, optimal duration, 1809
bacterial, 580-581
computed tomography, 843
development, risk factors, 1805f
diagnosis, 757-758
diagnostic/monitoring/prognostication usage, 838
differential diagnosis, 1226
Septic arthritis (Continued)
history, 580
images, 839f
incidence, range, 1801
infection, risk, 1809
knees, 1841
magnetic resonance imaging (MRI), 853
outcome, 1811-1812
problem, factors, 1812t
plain radiographs, 580-581
prognosis, 1811-1812
pseudogout, mimicy ability, 1588
results, 581
symptoms/signs, 579f
synovial fluid analysis, algorithm, 1806f
Thayer-Martin medium, usage, 581
treatment, 1807-1809
ultrasongraphy, 861-862
Seronegative spondyloarthropathies (SpA), 165
Seronegative spondyloarthritis
HLA-B27 gene, frequency, 592t
malignancy, association, 520
Seronegative spondylarthropathies (SpA), 130-131
Serotonin, 246
neurotransmitter, 743
Serotonin-norepinephrine reuptake inhibitors (SNRIs), 1020, 1024-1025
drug interactions, 1026t
pharmacodynamics, 1026t
side effects, 1025
Serentine seven receptors, 156
Serpins, 104
Serum amyloid A (SAA), 362, 822
acute phase response, 822
clinical associations, 822
concentrations, 1576
encoding, 1891
function, 822
levels, increase, 819
Serum amyloid P (SAP) protein, 258
Serum creatinine (SCR), decrease, 1305-1306
Serum CRP levels, correlation, 1274
Serum ferritnin, increase, 1278
Serum muscle enzymes, measurement, 1418
Serum N-propeptide of collagen type II A (PILAM), 481-485
Serum PIIINP levels, increase, 482
Serum prednisolone concentration, rifampin (impact), 898f
Serum procalsitomin (serum PCT) levels, elevation, 1322
Serum protein electrophoresis (SEP), sensitivity, 1893
Serum sickness model, 336
Serum thrombomodulin, levels (elevation), 1443
Serum unconjugated bilirubin, increase, 1278
Serum urate (SU)
concentrations, variation, 1554
increase, organic acids (impact), 1543
supersaturation concentrations, 1050
Services, units, 443f
Seven-transmembrane-domain receptors, 156
ligation, 156
7-amino-actinomycin D (7-AAD), 391
Severe combined immunodeficiency (SCID)
mice
human cartilage, presence, 218
T cells, absence, 234
model, IL-1 function (blocking), 1073
result, 174-175
Severe congenital neutropenia ( SCN), 162
Severely deforming osteogenesis imperfecta (types III/IV), 1729-1730
molecular pathology, 1729-1730
Shared epitope (SE), 305-306
alleles, term (usage), 811
amino acid substitutions, 305f
hypothecy, scrutiny, 305
Shark cartilage, nutritional supplement, 1654
Shank van der Heijde score, 481
Sharp scores, 459
median change, 958-959
modifications, 833
Shank van der Heijde method, 1163
Shedding hypothesis, 1668
Short form, 36 items (SF-36), 464
Short Form Six Dimensions (SF-6D) utility index, 464
Short Form-36 (SF-36) Physical Component Summary (PCS) score, 733-735
Short tau inversion recovery (STIR)
coronal MRI, 697f
sequences, 1210
magnetic resonance, 1211f
technique, 845-846
usage, 1222-1223
Shoulder-hand syndrome, 662, 1931
Shoulder joint arthrocentesis, 728f
examination, 564
skeletal lesions (evaluation), Tc99m methyl diphosphonate (MDP) (usage), 644
stability, negative intra-articular pressure (impact), 88f
Shoulder pain
anesthetics/glucocorticoids, injection, 648
arthroscopy, usage, 646-647
depression, 640t
cervical radiculopathy, 660
Clancy test, 642-643
computed tomography, usage, 645
cross-chest adduction test, 643
diagnosis, 643-648
diagnostic tests, preference, 648
electromyography, 647
extrinsic factors, 660-663
horizontal adduction test, 643
impingement test, elicitation, 642-643
intrinsic factors, 649-660
magnetic resonance imaging ( MRI), usage, 647
musculoskeletal complaints, 639
musculoskeletal/topographic areas, 641f
musculoskeletal/topographic localization, 639-640
Sjögren’s syndrome (SS) (Continued)
corticosteroids, usage, 1188
definitions, 1170-1173
diagnosis, 1184-1186
approach, 1185-1186
diagnostic tests, 1184-1186
Epidemiology Committee of the European
Community study, 1171
European Study Group on Classification
Criteria for Sjögren’s Syndrome, 1171-1172
 genic, 1173-1179
interest, increase, 1170
keratoconjunctivitis sicca, component, 1121
lateral salivary gland biopsy, investigation, 1170
laboratory evaluation, 1185
liver enzyme abnormalities, development, 1183
mouse models, 1175t
mucocutaneous findings, 605
pathogenesis, 1173-1179
peripheral nervous system, involvement, 1183
prognosis, 1188-1189
receiving operator curve (ROC) analysis, 1171-1172
revised international classification criteria, 1172f
salivary gland inflammation, pathogenesis
models, 1178f
Schirmer’s-I test, 1184
serologic markers, characterization, 1170
susceptibility genes, identification, 1174
 treatment, 1186-1188
algorithm, 1187f

Skeletal dysplasias, 1719-1725
biochemical abnormalities, 1725
classification, 1720
clinical evaluation/features, 1720-1723
diagnosis, 1723-1724
embryology, usage, 1719-1720
genetic basis/heterogeneity, 1722
growth hormone, effectiveness, 1725
histomorphology studies, 1724
management, 1724-1725
molecular abnormalities, 1725
mutations, 1726
nomenclature, 1720
radiographs, usage, 1723
testing, 1723-1724
treatment, 1724-1725
Skeletal muscle, signaling/contractile proteins, 68t
Skeletal muscle relaxants, classification, 1028
Skeletal pathology, endogenous PG/NSAIDs (role), 887

Skeletal structures, x-rays, 831

Skeletal tissues, example, 90
Skeletal tuberculosis, occurrence, 1830
Skelton
diabetes mellitus, 1931-1932
OA studies, 495
patterning/architecture, occurrence, 1719
Skin
active skin involvement, 1379
arthritis, 613
atypical infections, 611
biopsy, scleromyxedema, 608
changes, distribution, 1377
characteristic, 1731
dermatomyositis, 1412-1413
diffuse cutaneous systemic sclerosis, 1347f
discoloration, occurrence, 1909

Shoulder pain (Continued)
neoplasms, 662
nerve conduction velocity studies, 647
nerve entrapment syndromes, 661-662
nontraumatic causes, 649
O’Brien’s test, 643
periarticular disorders, 649-652
physical examination: impingement,
suspicion, 643
radiographic assessment, 643-644
reflex sympathetic dystrophy, 662
regional factors, 660-663
scintigraphy, 644
source, determination, 639
Speed’s test, 643
Spurling test, 642
sternoclavicular arthritis, 662
ultrasonography, usage, 645-646

Shoulders
abnormalities, 1115f
anatomy, 639-640
anterolateral aspect, soft tissue swelling, 1115
anteroposterior views, radiographs, 655
arthrocentesis, 778-779
arthrography, usage, 645
arthroscopy, indication/usefulness,
history, 640
hydroxyapatite crystal-associated calcific
bursitis, 1587f
impingement, 649-651
lesions, description (Neer staging system), 650
sign, 650
mechanical instability, 1586-1588
multidetector computed tomography-arthrography, 647f
musculoskeletal disorders, occurrence, 494
osteonecrosis, cause, 657
physical examination, 640-643
components, 640
radiographic examination, 1115
range of motion, recording, 642
rheumatoid arthritis, 1114-1115
rotator cuff tears (identification),
scintigraphy (usage), 644
sagittal section MRI, 656f
superior labral tear, multidetector computed
tomography, 645f
suprascapular neuropathy, pain
(production), 661
T-arthrography, 646f
upper extremity strength testing, 643
Shrinking lung syndrome, occurrence, 1275
Shulman’s syndrome, 1399-1400
Sialography, usage, 1184-1185
Sialylated GalNAc, 1513
Sialylated glycoprotein, binding, 157
Sialyl Lewis(X) (sLeX), 359
Siblings, relative risk (measurement), 310t
Sicca complex, chronic parotitis (association), 1186
Sickle cell disease, joint/bone infections
(complications), 1925
Sickling, occurrence, 1924
Sickness Impact Profile (SIP), 464

Signaling mammalian homologues of
Drosophila mothers against
decapentaplegic (SMAD)
availability, 5-6
cytoskeletal compartmentation, 48
pathways, 1359
impact, 48
redundancy, 38
site, absence, 39
Smad 6/Smad 7, impact, 48
Signaling pathways, 1096
Signaling proteins, 68t
Signaling receptors, defects, 265
Signal molecules, 477-479
Signal recognition particles, 1407
Signal transducers, 1095-1096
Signal transducers and activators of
transcription (STATs), 43-44, 1095
downregulation, 1333
expression, 1095
illustration, 3f
mediation, 819
STAT3
downregulation, 370-371
phosphorylation, 1089
STAT4, 1174
Signaling transduction, 1094-1096
Significant pediatric arthritis and other
rheumatologic conditions (SPARC), 595
Silent nociceptors, 11
Silica exposure, risk (meta-analysis), 1170
Silk Route, 1525
Silence classification, 1727
Simple erosion narrowing score (SENS), 833
Sindbis virus, isolation, 1870
Single nucleotide polymorphisms (SNPs)
ERAP-1, association, 1205-1206
impact, 1062
osteoporosis, association, 1662-1663
single-nucleotide polymorphism-defined
variants, 294
statistical analysis, 1271
Single-outcome variables (replacement),
indices (usage), 455
Single-photon emission computed tomography
(SPECT), 865
gout, image, 866f
usage, 865
Single support, schematic, 85f
Sinus tarsi, tenderness, 701
Sirolimus (rapamycin), 948-953
effects, 1664
Single nucleotide polymorphisms (SNPs)
expression, 1095
downregulation, 1333
Signal transduction, 1094-1096
Significant pediatric arthritis and other
rheumatologic conditions (SPARC), 595
Silent nociceptors, 11
Silica exposure, risk (meta-analysis), 1167
Silk Route, 1525
Silence classification, 1727
Simple erosion narrowing score (SENS), 833
Sindbis virus, isolation, 1870
Single nucleotide polymorphisms (SNPs)
ERAP-1, association, 1205-1206
impact, 1062
osteoporosis, association, 1662-1663
single-nucleotide polymorphism-defined
variants, 294
statistical analysis, 1271
Single-outcome variables (replacement),
indices (usage), 455
Single-photon emission computed tomography
(SPECT), 865
gout, image, 866f
usage, 865
Single support, schematic, 85f
Sinus tarsi, tenderness, 701
Sirolimus (rapamycin), 948-953
effects, 1664
Single nucleotide polymorphisms (SNPs)
expression, 1095
downregulation, 1333
Skin (Continued)
disease, recurrences, 1513
fibrosis development, CTGF suppression, 1396
IBD manifestations, 1260
infections, 611
infiltrative diseases, 613
involvement, 1366
organ-specific pathologic findings, 1346-1347
pathology, type, 1413
primary necrotizing vasculitis, 608-610
rash, 1412-1413
presence, 1777
rheumatoid arthritis, 1121
thickness
changes (measurement), 1370
semi-quantification methods, 1378f
Skin lesions
diagnosis, 599
differential diagnosis, 605, 610
erythema migrans (EM), 1817
granulomatosis, 610
treatment, 601
Skull
involvement, 1675
Paget's disease, 1674f
SLC2A9, transporter, 1050
SLEDAI, improvement, 1314
SLC2A9, transporter, 1050
SLAP lesion, 658
Spinal stenosis, 671-673, 680
Spinal sensitization, 416-418
Spinal muscular atrophy (SMA), 1423
Spinal mobility, 1209
Spinal instability, 670
Spinal accessory nerve injury, 661
Spondylosis, bilateral defects, 672f
Spondylolisthesis, 680
Spondyloarthropathy, 723-725
appearance, 1127-1128
detection, 1242
disease group, 1202
monarthritis, impact, 583
treatment, 679f
treatment goals, 678-679
Spinoarthritis
Amor criteria, 1222t
clinical characteristics, 1203t
European Spondyloarthropathy Study
Group Criteria, 1223t
sacroiliac joints, bone erosion, 842f
corticosteroids, 850f
types, 1203t
Spondyloarthritis Research Consortium of Canada (SPARCC) scoring system, 849-850
Spondyloarthritis, 723-725
appearance, 1127-1128
detection, 1242
disease group, 1202
monarthritis, impact, 583
SsZ/sulfapyridine, indications, 931
Spondyloarthropathy
SsZ/sulfapyridine, indications, 931
Spondylyodiskitis, 674
Spondyloepiphyseal dysplasia congenita, 1723
Spondylolysis, 680
asymptomatic patients, 671
degenerative spondylolysis, 671
radiographic changes, 679t
Spondylosis, bilateral defects, 672f
Spontaneous arthritis, 407-408
generically modified strains, 408
treated patients, 408
observation, 408
Spontaneous dismutase reaction, 159
Spontaneous hemopneumothorax, 1732-1733
Sporothrix schenckii, 1844
Sporotrichosis, 1844-1845
radiographic changes, 1845
Sports-related musculoskeletal disorders, 496-499
Spurling test, 642
Squalen-induced arthritis, 403-404
SRY-related high-mobility group-box protein (Sox), examples, 3f
Sry-type high-mobility group box 9 (Sox9), 38
Sulfate, incorporation, 41
Stable disease activity, 535
Stage II impingement, rehabilitation, 651
Stage zero Kienböck's disease, 713
Stage phase knee flexion, 80-81
Soft tissue (Continued)
examination, 633
fullness, plain x-rays, 1957
hemorrhage, 1916
problems, symptoms/signs, 579f
structures, delineation, 1780
swelling, 1115
tumors, occurrence, 718
Solid organs, antigen detection, 275
Solid-phase immunoassays, usage, 793
Soluble cytokine receptors, 1093
Soluble E-selectin, 363
Soluble frizzled-related protein-1 (SFRP1), expression, 65f
Soluble glucocorticoids, onset, 928
Soluble lectin-type PRRs, role, 258
Somatic hypermutation (SHM)
affinity maturation, 203
mechanism, 203-204
Somatic mutation process, activation, 203-204
Somatosensory stimulation, 744
Sonc hedgehog (Shh), 3f
expression, 5
Southern tick-associated rash illness (STARI), 1817-1818
Spa therapy, 1649
Spodaphilosis
SsZ/sulfapyridine, indications, 931
Sporothrix schenckii, 1844
Sporotrichosis, 1844-1845
radiographic changes, 1845
Sports-related musculoskeletal disorders, 496-499
Spurling test, 642
Squalen-induced arthritis, 403-404
SRY-related high-mobility group-box protein (Sox), examples, 3f
Sry-type high-mobility group box 9 (Sox9), 38
Sulfate, incorporation, 41
Stable disease activity, 535
Stage II impingement, rehabilitation, 651
Stage zero Kienböck's disease, 713
Stage phase knee flexion, 80-81
Stromelysin-1, activity (absence), 1102
Subphenotype analysis, 1273
Submandibular glands, bilateral swelling, 1169
Subintimal vasculature, 24-26
Subject selection, 453-454
Subligamentous tuberculosis, 1831
Sublingual, 20
synovial fibroblasts, interaction, 225f
zone, 222-223
Subluxation, 1109-1110
neuologic symptoms, relationship, 1116
Submandibular glands, bilateral swelling, 1169
Subphenotype analysis, 1273
Substance P
chemotactic/vasodilatory factors, 414
impact, 10
proinflammatory role, 423
proliferative effects, 424
release, 11
wound healing promotion, 422-423
Substrate AA, cell membrane source, 342
Subsynovial compartment (connective tissue), 1145
cartilage (development), 1959
Subtalar joints, arthrootesis, 784
Sudden blindness, 622-623
Sudeck’s atrophy, 662
Suicide inactivation, 343
Sulfasalazine (SSZ), 929-932
Suicide, 931
drug contraindications, 932
drug interactions, 932
effects, 551
eliminatiion, 931
gastrointestinal side effects, 932
gold sodium thiomalate, comparison, 931
half-life, 931
hematologic side effects, 932
bioavailability, 930
cancer, controversy, 930
autoantibodies, comparison, 931
inflammatory arthritis, 662
involvedm, 564
observation, 643
rheumatoid arthritis, 1115
Sternoocular joint, examination, 564
Steroid hormones
cholesterol, building block, 895f
classification, 894
precursor molecule, 894
Steroid-induced myopathy, recognition (difficulty), 1929
Steroids
lipid-altering effects, 1698t
myopathy, 1121
psychosis, glucocorticoids, 913
metabolites, 930f
indications, 931
impact, 10
immunomodulatory properties, 930
impair, 930
drug interactions, 932
effects, 551
elimination, 931
gastrointestinal side effects, 932
gold sodium thiomalate, comparison, 931
half-life, 931
hematologic side effects, 932
hepatic side effects, 932
hydroxychloroquine (HCQ), comparison, 931
immunomodulatory properties, 930
impact, 930
indications, 931
metabolites, 930f
methotrexate/hydroxychloroquine, combination, 1145f
pharmacology, 930-931
pregnancy, relationship, 551

Standardized incidence ratio (SIR), 1325-1326
lymphoma, 516
Standardized mortality ratio (SMR), 505
comparison, 505-506
Stanford Health Assessment Questionnaire (HAQ) disability index, 1641-1642
Staphylococcal arthritis, 401
Staphylococcal enterotoxin B (SEB), impact, 187-188
Staphylococcal surface proteins (SSPs), 1781
Staphylococcus aureus
cell surface proteins, genes, 1803-1804
detection, 580-581
exotoxins, 1774
infection, 674
pathogen, impact, 718
protein A (SpA), 206
schematic diagram, 1803f
twine factors, action (mechanisms), 1315
Staphylococcus epidermidis
detection, 580-581
prosthetic joint infection, 1772
Static constraint, component, 88
Static force equilibrium, 84f
Statics, 83
Statins
effectiveness, 510
myopathy, 1425
usage, 1375-1377
Statistical analysis, 459-460
Statistical interaction, 437-438
Statistical power, 456-457
Statistical test, 459-460
Statistical significance, 459
Statistical test, 459-460
Status epilepticus, 1539
Steinberg test, 1735f
Stem cell factor (SCF), 234
eexpression, 234
occurrence, 234
Stem cells, B cells origin, 979
Steppage gait, 701
Steppage walk, 108
Steppage gait, 701
Stenocostal joint, examination, 564
Sternonocular arthritis, 662
Sternoocular joint, 639
examination, 564
inflammatory arthritis, 662
involved, 564
observation, 643
rheumatoid arthritis, 1115
Sternoocular joint, examination, 564
Steroid hormones
cholesterol, building block, 895f
classification, 894
precursor molecule, 894
Steroid-induced myopathy, recognition (difficulty), 1929
Steroids
lipid-altering effects, 1698t
myopathy, 1121
psychosis, glucocorticoids, 913
treatment, impact, 1409-1410
Steroid withdrawal syndrome, 1929-1930
Stiffness
morning stiffness, 560
patient complaint, 560
test, 87
Still’s disease (juvenile rheumatoid arthritis), 602
diagnosis, criteria, 1131t
Stimulated tissue fibroblasts, NGF production, 420-421
Stimulation assays, hiiKRNAP-A1 induced T cell proliferation (usage), 811
S-T-L (Swelling Tenderness Limitation of motion) system, 563
Stoke AS spondyilitis score (SASSS), 1210
Stomal, edema (involvement), 1383
Ston, Edward, 872
Stratified randomization, 453
post hoc subgroup analysis, contrast, 453
Streptococcus pneumoniae, 1318-1319
Streptococcus salivarius (comparison), 1055
Subcutaneous fibro-fatty nodule, back muscle, 1642f
Subcutaneous heparin, usage, 548
Subcutaneous nodules, 548
Subacute cutaneous lupus (papulosquamous variant), 1276
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute pseudogout, association, 1584
Subacute lupus erythematosus (SCLE), 1272
Subacute cutaneous lupus (SCLE), 1272
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute cutaneous lupus (SCLE), 1272
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute cutaneous lupus (SCLE), 600, 1277-1278
Subacute pseudogout, association, 1584
Subaxial cervical spine, components, 222-223
Subaxial subluxation, frequency, 833
Subaxial involvement, frequency, 833
Subcapsular macrophages, 199
Subchondral bone
collapse, pyrophosphate arthropathy (association), 838
interactions, articular cartilage, 13-14
plate, articular cartilage (bonding), 2f
Subchondral bone (Continued)
sclerosis, 1628
separation, 13-14
turnover, 1632
Subchondral bone marrow lesions (subchondral BMLs), 1642f
Subligamentous tuberculosis, 1831
Subluxation, 1109-1110
neuologic symptoms, relationship, 1116
Submandibular glands, bilateral swelling, 1169
Subphenotype analysis, 1273
Substance P
chemotactic/vasodilatory factors, 414
impact, 10
proinflammatory role, 423
proliferative effects, 424
release, 11
wound healing promotion, 422-423
Substrate AA, cell membrane source, 342
Subsynovial compartment (connective tissue), 1145
cartilage (development), 1959
Subtalar joints, arthrootesis, 784
Sudden blindness, 622-623
Sudeck’s atrophy, 662
Suicide inactivation, 343
Sulfasalazine (SSZ), 929-932
Absorption, 930
action, 929-930
site, 930
active sera, controversy, 930
autoantibodies, comparison, 931
bioavailability, 930
chemical structure, 929
dermatologic side effects, 932
distribution, 931
DMARD, 1140
dosing, 931
D-penicillamine (D-Pen), comparison, 931
drug contraindications, 932
drug interactions, 932
effects, 551
elimination, 931
gastrointestinal side effects, 932
gold sodium thiomalate, comparison, 931
half-life, 931
hematologic side effects, 932
hepatic side effects, 932
hydroxychloroquine (HCQ), comparison, 931
immunomodulatory properties, 930
impair, 930
indications, 931
metabolites, 930f
methotrexate/hydroxychloroquine, combination, 1145f
pharmacology, 930-931
pregnancy, relationship, 551
Synovial fluid (Continued)
healthy appearance, 753
hyaluronic levels, decrease, 482
hydroxyapatite, deposits (presence), 757
immune complexes, 1082-1083
laboratory testing, 1640
leukocytes
characterization, 756
count, 755
Wright stain, usage, 756
lymphocytes, 1082
subsets, 1082
microorganisms (detection), gram stain (usage), 757-758
noncellular portion, analysis, 758
osteocartilaginous, laboratory testing, 1640
polymerase chain reaction (PCR)
sensitivity/specificity, 758
presence, 1
proteins, concentration ratio, 15f
samples, analysis algorithm, 754f
specimen, birefringent material (appearance), 756
T cells, presence, 1082
ultrasonographic pathology definition, 1071-1081
volume/composition, 14-15
wet smear analysis, polarized microscopy (usage), 756-757
Synovial hemangiomas, symptoms, 1956
Synovial hyperplasia, 762
Synovial hypertrophy, ultrasonographic pathology, 1071-1072
Synovial immunohistochemistry, 763-767
sampling/quantitative analysis, considerations, 763
Synovial innate immunity, activation, 1060
Synovial intimal cells, transmission electron microscopy, 21f
Synovial intimal fibroblasts, 21-22
Synovial intimal lining
architecture, 1072
cells, type A/B synoviocytes, 1071-1073
Synovial intimal macrophages (type A cells), transmission electron micrographs, 22f
immunohistochemistry classification, 23f
Synovial joints
classification, 1
development, 2f
example, 2f
lubrication mechanisms, 86
Synovial-like cells, proliferation, 7
Synovial lining, 9-10
inflammation, 1631
layer, cadherin 11 (expression), 764f
resident fibroblast population, 23
Synovial lining cells (SLCs)
adhesion molecule expression, 10
composition, 20
layer, 763-764
mitotic activity, 22-23
origin, 23
proliferation, 22-23
turnover, 22-23
ultrastructure, 20
Synovial lymphocytes, 764-765
Synovial macrophages, 21
lineage-specific molecules, detection, 21
presence, 137
urate crystal phagocytosis, 164-165
Synovial mastocytosis, presence, 240t
Synovial membrane, 9
biopsy specimens, 1528
healthy appearance, 759-760
Synovial mesenchyme, formation, 2f
Synovial microvasculature
porosity, 26
Synovial osteochondromatosis, 1957
Synovial pannus, detection, 858
Synovial pathology, 759-763, 1071-1081
usage, 766-767
Synovial proliferation, development, 1112
Synovial sarcoma, 1963-1968
cytogenetic studies, 1967
gross pathology, 1966
immunohistochemistry, 1967
monophasic spindle cell variant, 1967f
plain film findings, 1966
prognosis, 1967-1968
subtypes, 1966-1967
Synovial stromal markers, geographic/functional significance, 217f
Synovial sublining macrophages, 765-766
Synovial T cell immunoreactivity, 1075
Synovial T cell phenotype, 1074-1075
adhesion molecules, 1074
chemokine receptor, 1074-1075
co-stimulatory molecules, 1074
Synovial tissue
active infection, 1065
alterations, 1631-1632
biologic markers, 484-485
compartment, responsiveness (decrease), 1075
examination, 1073
inflammation, 482, 1069
morphologic types, transmitted light photomicrographs, 24f
sampling, 758
tissue-fluid composition, 9
Synovial tissue samples
antigen retrieval methods, usage, 759
hematoxylin and eosin (H&E) histologic analysis, usage, 759
pathologic analysis, 760
processing, 759
Synovial T lymphocytes, 1073-1076
immunohistologic patterns, 1073-1075
Synovial vasculature, 10
Synoviocytes, 20
Synovitides, 336b-337b
Synovitis, 703f
B cell contribution, 1077
cluster, 485
demonstration, 1837f
detection, EULAR recommendation, 578
everlhat rheumatoid arthritis, 1080
perpetuation, 1093-1094
Synovitis, acne, pustulosis, hyperostosis, osteitis (SAPHO), 1128, 1680
clinical manifestations, 1688
diagnosis, 1688
etiology, 1687-1688
imaging findings, 1688
laboratory investigations, 1688
osteoarticular features, 1688
pain, clinical manifestations, 1688
pathogenesis, 1687-1688
syndrome, 1687-1688
treatment, 1688
Synovium, 8-10
angiosarcoma, 1965f
appearance, 760f
benign vascular tumors, rarity, 1956
cartilage nodules, attachment, 1959
Synovium (Continued)
cell-cell interaction, 225f
deformability, 26
development, 7-8
fatty lesions, 1954-1956
function, 26-29
hyaline cartilage, synovial chondromatosis nodules, 1959f
immunoperoxidase staining, 763f
impact, 29
inflammation
citrullinated peptides, 1068f
juvenile idiopathic arthritis, histologic features, 1743
joint movement, 26-27
leukocyte extravasation process, 359f
lipomatosis, 1954-1955
lubrication, 26-27
microscopic appearance, 222f
nonadherence, 26
porosity, 26
schematic representation, 9f
starvation, feeding, 1098-1100
structure, 20-26
transmission light photomicrographs, 25f
unmyelinated nerve fibers, presence, 11
vascular lesions, 1956-1957
radiographic evaluation, 1956
vascular supply, 24
Synthesized mediators, 236-237
Synthetic tolerogens, usage, 1314-1315
System for Thalidomide Education and Prescribing Safety (STEPS) program, 954, 1530
Systemic absorption, occurrence, 775-776
Systemic agents, usage, 1650-1651
Systemic amyloidoses, clinical features/treatment, 1892-1896
Systemic anaphylaxis, mast cell mediated, 237
Systemic autoimmune diseases, 281-282
Systemic autoinflammatory rheumatic diseases, malignancy development (association), 515
Systemic disease involvement, 577
Systemic GPA, renal involvement, 1488
Systemic inflammation
ESR/CRP markers, 534-536
systemic cytokines, 534-536
Systemic juvenile idiopathic arthritis (SJIA), 1746-1747, 1763-1766
abatecept, usage, 1766
amyloidosis, relationship, 1766
biologics, 1766
characteristics, 1763
clinical manifestations, 1763-1765
diagnostic features, 1763-1765
diagnostic tests, absence, 1764
differential diagnosis, 1764
gene expression profiling, 1747
MAS
considerations, 1764-1765
impact (diagnostic guidelines), 1764t
occurrence, 1763
outcome/prognosis, 1766
rash, appearance, 1764f
thalidomide, usage, 1766
treatment, 1765-1766
algorithm, 1765f
Systemic juvenile rheumatoid arthritis, 281

Systemic lupus erythematosus (SLE) (Continued)
pulmonary arterial hypertension (PAH), relationship, 1293
randomized controlled trials (RCTs), 1285
rehabilitation interventions, 535
renal biopsy, 1289-1292
renal function, measurement, 1289
renal involvement, 1288-1292
considerations, 1288-1289
types, 1289
renal transplantation, 1324-1325
revised American College of Rheumatology Classification Criteria (1982), 1272
ribonucleoproteins, presence, 796
rituximab, usage, 1312-1313
secondary Sjögren’s syndrome, prevalence, 1173
serologic tests, 1299-1300
utility, 1280
severity, variability, 1272
significance, 1269
skin lesions, 1280
SLE Genetics (SLEGEN), studies, 1271
SLE-related disorders, vasculitis (impact), 1516
smoking, presence, 508
specificities, 792f
spontaneous group, 282
standard mortality rate (SMR), 1775
stroke, risk (increase), 1277
studies, 1035
subphenotype analysis, 1271
synthetic tolerogens, usage, 1314-1315
systemic diseases, mimicry, 1280
systemic sclerosis (scleroderma), association, 797-798
tacrolimus, usage, 1312
target organ damage, mechanisms, 1280
T cells, correlation, 1280t
therapeutic drug monitoring, 1305-1315
therapies, 1314-1315
thrombocytopenia, relationship, 1298
transesophageal echocardiographic (TEE) study, 1275
treatment
algorithms, management, 1315-1320
drug usage, 1305-1315
strategy, 1304
type I interferon signature, similarities, 1356
urinalysis, 1289
urinary tract symptoms, 1326-1327
valvular abnormality, 1294
Libman-Sacks endocarditis (relationship), 1275
vasculature, role, 1280
viral infections, 1280, 1324
women, health issues, 1326-1327
Systemic Lupus International Collaborating Clinics-American College of Rheumatology Damage Index (SDI), usage, 1306
Systemic Lupus International Collaborative Clinics (SLICC), SLICC-Damage index, 1775-1776
Systemic necrotizing vasculitis, 547
Systemic rheumatologic disease, 531-552
Systemic sclerosis, 607
Systemic sclerosis (scleroderma) (SSc), 518-519, 789, 1944
animal models, 1349f
antinuclear antibodies, presence, 797t
autoantibodies, pathogenic considerations, 1356
Systemic sclerosis (scleroderma) (SSc) (Continued)
autoimmunity, 1356
B cells, relationship, 1356
bioactive lipids, 1361
candidate genes, 1344-1345
collagen synthesis, regulation, 1357-1358
developmental pathways, 1360
dietary supplements, usage, 1345
drugs, usage, 1345
eye-stage SSc skin biopsies, 1346-1347
environmental exposures, 1345
etiology, 1343-1345
extracellular matrix, relationship, 1357
fibrosis, 1356-1359
overview, 1356-1357
pathogenesis, extracellular cues, 1358t
gastrointestinal tract, 1348
genetic association studies, 1344
genetic risk, family studies, 1343
genome-wide association studies, 1344-1345
hallmarks, 1346
heart, 1348
immune dysregulation, 1354-1356
cellular effectors, 1354-1355
immune susceptibility genes, 1344
infectious agents (viruses), 1345
juvenile systemic sclerosis, 1783
kidneys, 1348
lungs, 1347-1348
fibrosis, 1348
malignancy, development risk (increase), 1944
Medger Systemic Sclerosis Severity Scale, 1371t
microchimerism, 1345
models
bleomycin-induced mouse model, 1350f-1351f
report, 1348
nail-fold capillaroscopy, presence, 1437f
organ-specific pathologic findings, 1346-1348
oxidative stress, 1354
pathogenesis, 1352-1354
overview, 1352
pathology, 1346
pathophysiologic quartet, 1344f
peroxisome proliferator-activated receptor γ, relationship, 1362
platelet-derived growth factor (PDGF), relationship, 1360
radiation, usage, 1345
skin biopsies, systematic survey, 1346
SLE, association, 797-798
specificities, 793f
SSc-like illnesses, drugs (implication), 1345
SSc-like phenotypes, mice (genetic considerations), 1345
T cell activation, evidence, 1355
TH1/TH2 cytokine balance, 1345
Th1/Th2 cytokine balance, alteration, 1355
tissue fibrosis, 1346
transforming growth factor-β, fibrogenic activities, 1359f
UIP, association, 1348
vascular damage, 1353
vascular injury, 1352-1353
vascular pathology, 1346
vascularopathy, 1352
Systemic vasculitides, nomenclature (CHCC review), 1457
Systemic vasculitis
age/gender/ethnicity, impact, 1458
autoimmune disease, 281-282
blood vessels
clinical manifestations, 1455t
site classification, 1454, 1455f
classification, 1453-1457
confusion, 1456-1457
considerations, 1456t
evolution, 1457
historical attempts, 1456
environment, 1459
epidemiology, 1457-1459
genres, relationship, 1458-1459
geography, 1457-1458
granulomatous features, 1455
nomenclature, historical attempts, 1456
Zeek Classification, 1456
Systems biology, 487
viewpoint, 91f
T

TACI receptor, 197
Tacrolimus (FK506), 948-953
action, mechanisms, 951
clinical indications, 951
dosage, 951
pharmacology, 951
structure, 951
toxicity, 951-952
usage, 1312
Takayasu’s arteritis (TA), 547, 1336, 1472-1478, 1793-1795
American College of Rheumatology (ACR) criteria, 1473, 1473t
angiographic progression, 1795f
aorta, disease (demonstration), 1477
bilateral renal artery stenosis, angiogram, 1475f
blood vessel involvement, frequency, 1476t
cardiac involvement, occurrence, 1474
cases, geographic clustering, 1474
cause, 1473-1474
changes, angiogram, 1475f
chest, magnetic resonance imaging, 1475f
classification, 1793-1794
claudication, impact, 1474
clinical features, 1474-1476
list, 1474t
clinical presentation, 1794
definition, 1793-1794
diagnosis, 1476-1477, 1794
algorithm, 1477f
diagnostic tests, 1476-1477, 1794
differential diagnosis, 1477t
epidemiology, 1473, 1794
GCA, comparison, 1477t
imaging studies, 1475-1476
imaging techniques, comparison, 1476t
laboratory findings, 1475
maternal complications, 547
medical therapy, 1477-1478
ocular thromboaortopathy, 1472-1473
outcome, 1478, 1795
pathogenesis, 1473-1474
peripheral neuropathies, absence, 1474
prognosis, 1478
pulseless disease, 1472-1473
relapses, treatment, 1478
surgical therapy, 1478
symptoms/signs, 1474
systemic inflammation, symptoms, 1474
treatment, 547, 1477-1478, 1794-1795
Talonavicular disease, bone edema (T1-MRI), 1239f
Transplant rejection, antimetabolites (usage), 521
Transporter associated with antigen processing (TAP), peptide transportation, 125
Transrepression, 898-899
Transhrocic echocardiogram (TTE) murmur evaluation, 1275
usage, 1339
Transhyrynine (TTR), 822
amyloidosis, association, 1890
Transhyrynine amyloidosis (ATTR), 1890, 1895-1896
Trauma-related osteonecrosis, anatomic considerations, 1695
Traumatic arthritis characterization, 725
TMJ, 725
treatment, 725
Treatment of Early Aggressive RA (TEAR) trial, 936, 1145-1148
clinical findings, 1147-1148
conclusions, 1148-1149
design, 1150
radiographic improvement/progression, 1149f
treatment groups, schematic, 1148f
treatment results, 1148f
Treatments
dollar costs, 465
effectiveness, evidence, 436-437
escalation, schematic, 1139f
Trendelenburg lurch, 694
Trendelenburg test, usage, 569
Trial analysis, 457-460
hypothosis testing, 457-458
limitation, 458
problems, 458-459
selection processes, net result, 459
Trial design, 452-457
considerations, 453
effect, measurement, 455-456
follow-up considerations, 454
outcome variables, choice, 455
randomized clinical trials, 452-455
subject selection, 453-454
Trial of Atorvastatin in the Primary Prevention of Cardiovascular Endpoints in Rheumatoid Arthritis (TRACE-RA), 510
Trial result, interpretation, 456f
Triamcinolone hexacetonide (THA), usage, 1754-1755
Triangular fibrocartilage complex (TFCC), 714
ancillary studies, 715
injury, 714-715
tears, 714
Trichrome staining, 1416f
Tricuspid annular plane systolic excursion (TAPSE), 1388
Tricuspid gradient (TG), usage, 1386-1387
Tricuspid valve insufficiency, 1731-1732
Tricyclic antidepressants (TCAs), 1023-1024
drug interactions, 1026t
pharmacodynamics, 1025t
side effects, 1024
Trigger fingers, 566, 717, 1930
injection technique, 717f
location, 717
Triggering receptor expressed on myeloid cells (TREM) 2, 63-64
Trigger points
palpation, tenderness (presence), 642f
therapy, benefits, 534-535
Trimethoprim/sulfamethoxazole (TMP/SMX), treatment, 1947
Triple-helical collagen
clavage, collagenases (impact), 1624
type II, B-cell epitope identification, 809
Triple helical glycine residue, 1726
Tripterygium wilfordii active ingredient (Celastrol), 365
Trochanteric bursitis, 570
lateral thigh pain, 693
Trochanteric pain syndrome (bursitis) arthrocentesis, 782
injection, 782f
Troponyinemia umplified infection, 582-583
True ankle, hinged joint, 574
Truthful, term (usage), 455
Tryptophanyl-t-RNA-synthetase, expression, 128
Tsk2 mouse, scleroderma animal model, 1350
Tsk1/+ mouse, characterization, 1350
Tsk2 mouse, scleroderma animal model, 1350
T-tubules, cutting, 63f
Tsk2 mouse, characterization, 725
TMJ, 725
treatment, 725
Tuberculosis (TB)
Anti-TNF Trial in Rheumatoid Arthritis with Concomitant Therapy (ATRACT) trial, 958-959
blockers, lateral thigh pain, 693
cardiovascular risk, 965
CD40, transmembrane protein member, 249
immune function, 302
inhibition, 1150-1151
IL-1 inhibition, combination, 970
infliximab, usage, 1261
level, reduction, 477
locus, genetic association, 1744
monitoring, 947
overexpression, 408
pleiotropic cytokine, implication, 1089
pleiotropic molecule, 1606
pregnancy, relationship, 553
presence, 418
randomized trial, 1151
regulation, 343
superfamily, 1089
cytokines, roles, 374t
targeting, 1194
TNF-homology domain, 1089
TNF-inadequate responders, Phase IIIb/IV open-label study, 992
TNF-related apoptosis-inducing ligand (TRAIL), 161
toxicity, 965-968
usage, 90
vaccinations, 965
Tumor necrosis factor (TNF) inhibitors, 50-51
cachectin, terminology, 51
cytokine production, 44
Tumor Necrosis Factor-Alpha-Related Apoptosis-Inducing Ligand (TRAIL) administration, 395
signals, 387
Tumor necrosis factor (TNF) blockade comparison, 993-994
contraindication, 986
success, 983
Tumor necrosis factor convertase (TACE), 1102
inhibitors, 1103
Tumor necrosis factor (TNF) inhibitors, 957-968, 1314
actions, mechanisms, 964
list, 964t
adverse effects, 966t
antigenicity, 966
autoimmune disorders, 968
congestive heart failure, 968
demyelinating syndromes, 968
efficacy, 1392
exposure, 553
infections, 966-967
infusion, 966
injection site reactions, 966
lipid profile, 965
malignancy, 967-968
opportunistic inhibitors, 967
pregnancy, 965
role, 964-965
surveillance data, postmarketing, 967
therapy, use, 965
treatment, 810-811
usage, 546
Tumor necrosis factor receptor (TNFR), 174
family members, characterization, 386
protein superfamily, 386